# Conflict Discovery and Analysis for Clinical Trials*

Bonnie MacKellar
St. John's University
Queens, NY USA
mackellb@stjohns.edu

Christina Schweikert
St. John's University
Queens, NY USA
schweikc@stjohns.edu

## ABSTRACT

Today, cancer patients and their caregivers often prefer to share the decision making process with their physicians and may be highly involved in the process of locating and choosing clinical trials for treatment. One issue is that treatments received on one trial may preclude participation in other trials because of eligibility requirements in the study design. We are developing a system to help patients and caregivers locate clinical trials, focusing on pediatric cancer where clinical trial participation is very high. We present a method by which *conflicts* - that is, the possibility that one or more interventions in one trial may cause a patient to not be eligible for another trial - can be determined in a group of clinical trials for Wilm's Tumor. More specifically, a conflict occurs when a drug or treatment mentioned in an intervention of one trial is also present in an eligibility criterion in another trial. We present results based on generating the conflicts in this group of trials, including the types of trials are most likely to cause conflicts. We also look at the specific treatments and drugs that cause conflicts, using the UMLS Metathesaurus concepts. The conflict generating algorithm will be used as part of the clinical trial search system, allowing patients to determine if a given trial will preclude him or her from other trials in the future.

## CCS CONCEPTS

**• Applied computing → Life and medical sciences → Health informatics**

## KEYWORDS

medical informatics; clinical trials; natural language processing; knowledge representation


*DH '17, July 02-05, 2017, London, United Kingdom*

ACM ISBN 978-1-4503-5249-9/17/07…$15.00
http://dx.doi.org/10.1145/3079452.3079494

## 1. INTRODUCTION

Today, cancer patients and their caregivers often prefer to share the decision making process with their physicians [7]. Especially in a relapse situation, patients/caregivers may be choosing from multiple treatments and clinical trials, and planning for a future that involves a sequence of clinical trials. This has given rise to a number of clinical trial search sites that are useful for patients, most importantly, ClinicalTrials.gov [4, 18]. Many of the disease-specific foundations, such as the Liddy Shriver Sarcoma Initiative (http://sarcomahelp.org/clinical-trials.html) or the American Lung Association (http://www.lung.org/our-initiatives/research/clinical-trials/) have pages on their websites offering advice to patients on how to search for clinical trials.

In the pediatric cancer world, most treatment is given as part of a clinical trial [14]. It is common for children with recurrent cancer to participate in more than one trial over time, and caregivers are considered to be part of the decision making team when choosing a trial [5]. A problem is that trials usually have eligibility criteria that prevent patients from participating based on various conditions, including past treatments. This means that participation in one trial may prevent a patient from participating in another trial later, because of treatments in the first trial. For example, the Parent's Handbook, which is maintained by the Children's Neuroblastoma Foundation, notes "*Some trials may prevent you from being treated with another drug/trial down the road; some may have long periods of recovery with low counts and transfusions; and some may have high risk of side effects that will rule out subsequent treatments. Read the trial documents and consent forms carefully, and talk with your doctors about a possible series of treatments that will make the best use of the available treatments while not precluding the possibility of other effective treatments later.*" [2]. This presents a real challenge when searching for clinical trials. To help parents and caregivers navigate this problem, we have developed a method for automated identification of potential treatment conflicts between trials, with the goal of integrating this capability into a clinical trials search tool aimed at cancer patients and their caregivers. Initial descriptions of this system may be found in [8] and [9]. We have used the algorithm developed for this system to generate and analyze conflicts among a set of clinical trials aimed at Wilm's Tumor patients, and present in this paper an analysis of characteristics of trials and interventions that cause conflicts. Understanding the most frequent causes of conflicts is useful in presenting the information to patients/caregivers searching for trials.

## 2. TRANSLATION OF ELIGIBILITY CRITERIA

In order to determine whether, by participating in a given trial, a patient could become ineligible for another trial, we need to compare the interventions in one trial against the eligibility criteria (ECs) in the second trial. An example of a EC is the following "*Subjects with prior history of stem cell transplantation must be off immunosuppressive therapy for at least 4 weeks and have no active graft-versus-host disease (GVHD) with the exception of Grade 1 acute at the time of entry onto this trial."*(Trial NCT01222780). For example, a clinical trial that contains stem cell transplant as one of its interventions could *potentially* prevent a patient from entering this trial. Clinical trials are available in XML format from ClinicalTrials.gov. Interventions are listed individually and their UMLS concept identifier can be obtained by matching the intervention terms using MetaMap [1]. Eligibility criteria are more difficult since they are written in free text; thus they must be translated into a structured knowledge representation which captures their meaning. A diagram of the structure of our knowledge representation is included in Figure 1.

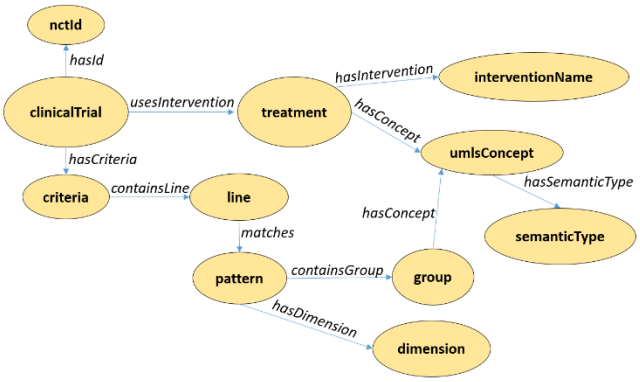


**Figure 1. Knowledge Representation for Clinical Trial Data**

We downloaded a set of 136 trials aimed at Wilm's Tumor patients, a cancer that mainly occurs in children. Two outliers were removed, resulting in a set of 134 trials. As reported in [9], we use a set of pattern rules, first developed by Milian [10, 11, 12] for ECs in breast cancer trials, to parse ECs in the set of Wilm's Tumor trials into a structured representation. There are a total of 163 pattern rules. The XML files were preprocessed to just include the trial interventions and ECs, for ease of processing. Additional pre-processing of the text included separating sentences, as well as bulleted and numbered lists. We used the text mining framework UIMA, which allows a sequence of annotators to be pipelined [3]. In this case, the sequence consisted first of annotation with MetaMap, which marks terms with their UMLS concept, followed by the pattern rules implemented in RUTA. RUTA is a rule based language that matches text to patterns, resulting in annotations [6]. The final component in the pipeline generates the structured representation. The example above, therefore is first annotated with UMLS concept identifiers. For example, the term "stem cell transplantation" is marked with the concept identifier C1504389 (Stem Cell Transplant, semantic type Therapeutic or Preventive Procedure), and "graft-versus-host disease" is marked with the concept identifier C0856825 (Acute GVH Disease, semantic type Disease or Syndrome). The rule-based annotator then matches the EC against the set of patterns; in this case three patterns match. Each pattern contains groups, which allow the most important components in the EC to be marked. It is the text in the groups, along with the UMLS annotations, that will be used in the conflict generation stage. In the last step, the following substructure is generated:

**Clinical Trial** :NCT01222780
**Has criteria**: criteriaList (65 criteria in list)
**containsLine**: *"Subjects with prior history of stem cell transplantation .."*
**Matches**: ecClassifier.R3CurrentPresenceConstraint
**Matches**: ecClassifier.R68CondReq
**Matches**: ecClassifier.R6Dot10Dot3PriorPresenceConstraintTimeIncl
**Group1**: *"history of stem cell transplantation must be off immunosuppressive therapy*"
**hasConcept**: CUI= C0021079 COVERED TEXT = *immunosuppressive therapy* SEMANTIC TYPE = topp
**hasConcept**: UMLS concept = CUI= C1504389 COVERED TEXT = *stem cell transplantation* SEMANTIC TYPE = topp
**Group2**: *:"4 weeks and have no active graft-versus-host disease (GVHD) with the exception of Grade 1 acute at the time of entry onto this trial"*
**hasConcept**: "UMLS concept = CUI= C0856825 COVERED TEXT = *graft-versus-host disease (GVHD)"* SEMANTIC TYPE = dsyn

**Figure 2. Portion of Knowledge Representation for NCT01222786**

## 3. ALGORITHM FOR FINDING CONFLICTS

The motivation for developing a procedure to identify conflicts between clinical trials is to have the capability to inform patients that their participation in a particular trial may affect their eligibility for other trials. Detecting the conflicts among a set of clinical trials for a specific disease also enables us to analyze the kinds of interventions that most commonly lead to a patient being excluded from a trial, as well as the types of trials.

After processing the Wilm's Tumor trials as described in section 2, relevant text, which may represent a type of drug therapy or disease, is stored in capture groups for further analysis. In Figure 2 above, this text is shown in Group1 and Group2. In our implementation, a pattern may have from one to three groups. The text in the groups has been already annotated with UMLS concepts, including Concept Unique Identifiers (CUIs) and semantic types. The interventions have also been similarly annotated. The CUIs are then used to compare intervention text with the text in the capture groups for the ECs. This allows terms in the text that are synonyms for the same concept to be correctly matched. We can examine the following examples, which illustrate this point. In one trial's intervention list, the text

"radiation" is used, while in another trial's eligibility criteria, the text "radiotherapy" is used. Both of these terms map to the CUI C1522449, which enables us to recognize them as a match. Another example are the terms "stem cell transplantation" and "stem cell transplants", both mapping to the CUI C1504389.

Given a list of clinical trials we compute the total number of conflicts that each trial has with the other trials in the list. To do this, we implemented a conflict detection algorithm, which iterates through the concepts of the interventions of one trial, and searches for conflicts with any concepts captured by patterns matching to the eligibility criteria of another trial. When a conflict is found, we store the pattern, concept, and the concept's semantic type for further analysis. After computing the conflict sums for all trials, we sort the trials to see the relative number of conflicts. The algorithms for computing the total number of conflicts for a trial and for finding the number of conflicts for an ordered pair of trials (conflict score), are shown in Figures 3 and 4. If we were performing this analysis for a specific patient, we would only include the subset of trials for which a patient is eligible. The trials would then be sorted according to the number of conflicts and could be filtered to only show trials in a particular phase, such as Phase 3 trials.

```
computeConflictSum(trial_A, trialList)
Initialize conflictSum to 0
For each trial_B in trialList
   If trial_A not equal to trial_B
            conflictSum += computeConflictScore(trial_A, trial_B)
   End if
return conflictSum
```

**Figure 3. Computing the total number of conflicts for a trial**

```
computeConflictScore(trial_A, trial_B)
Initialize conflictScore to 0
For each criterion of trial_B's criterionList
    For each intervention of trial_A's interventionList
        For each concept_A of intervention's conceptList
            For each pattern of criterion's patternList
                For each concept_B of group1's conceptList
                    If concept_A equals concept_B
                        conflictScore++
                    End If
                For each concept_B of group2's conceptList
                    If concept_A equals concept_B
                        conflictScore++
                    End If
                For each concept_B of group3's conceptList
                    If concept_A equals concept_B
                        conflictScore++
                    End If
return conflictScore
```

**Figure 4. Computing a conflict score for an ordered pair of trials**

## 4. RESULTS

From the set of 134 Wilm's Tumor trials, a total of 1056 conflicts were generated. Since a trial can have a number of interventions, it can generate numerous conflicts with other trials. For example, the diagram in Figure 5 shows all the trials that NCT00004157 may conflict with, in other words, the trials that a patient who entered NCT00004157 could possibly be excluded from in the future.

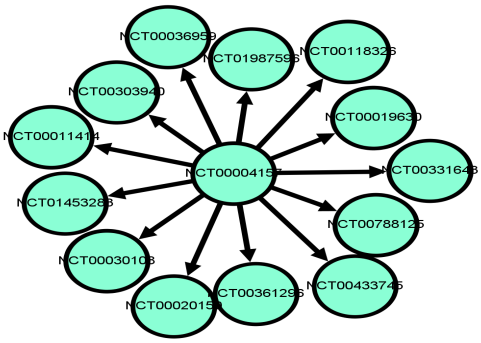


**Figure 5. Conflicts for trial NCT00004157**

It is useful to visualize the entire set of trials in terms of a conflict graph. The following conflict graph shows the trials as nodes, colored by trial phase and sized by number of conflicts with other trials (weighted outdegree).

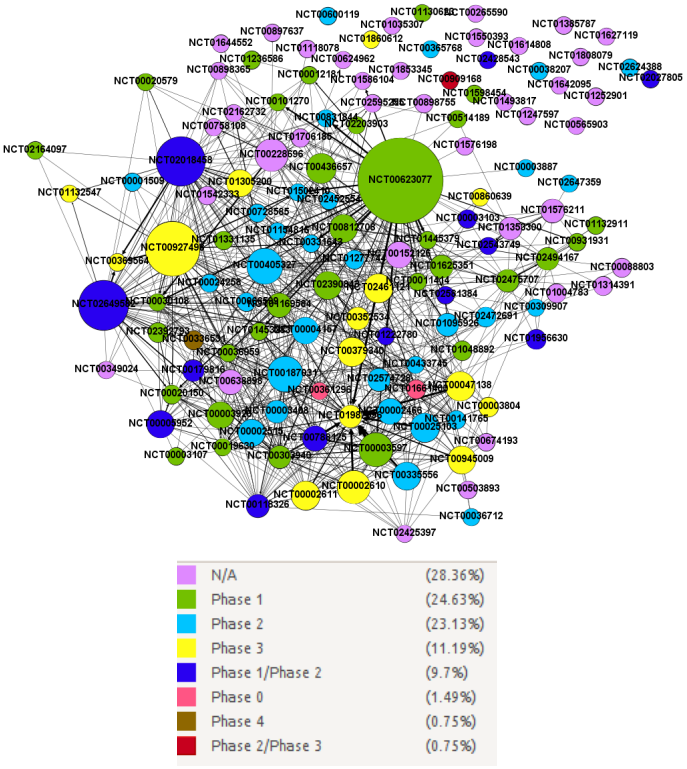


**Figure 6. Visualization of all conflicts in the Wilm's Tumor clinical trial set**

We can see that the disconnected nodes - those with no conflicts at all – are most likely to be trials with no phase. Many of these are observational trials. The trials that generate the most conflicts, which are the largest nodes, are the Phase1, and Phase1/2 trials. Trials that are strictly Phase2 trials seem to generate fewer conflicts in this dataset, which may be useful for patients/caregivers looking for trials that are not as likely to preclude other trials.

The interventions in each clinical trial are listed by type: drug, biological, or genetic. Clinical trials that include drug interventions are most likely to generate conflicts, followed by trials that include biological interventions. Trials that include genetic interventions are the least likely to generate conflicts.

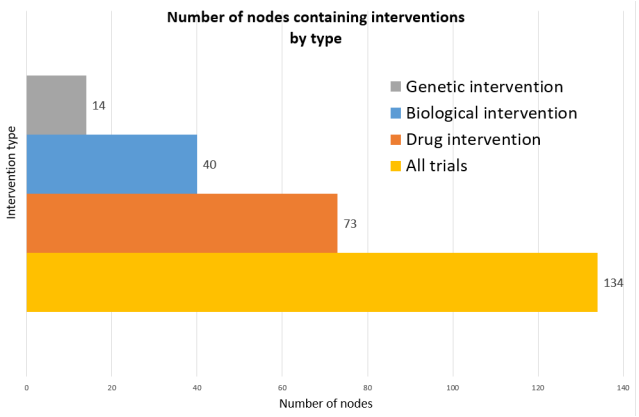


**Figure 7. Number of nodes containing interventions by type**

The top 10 UMLS Metathesaurus concepts that generate the most conflicts are shown in Figure 8.

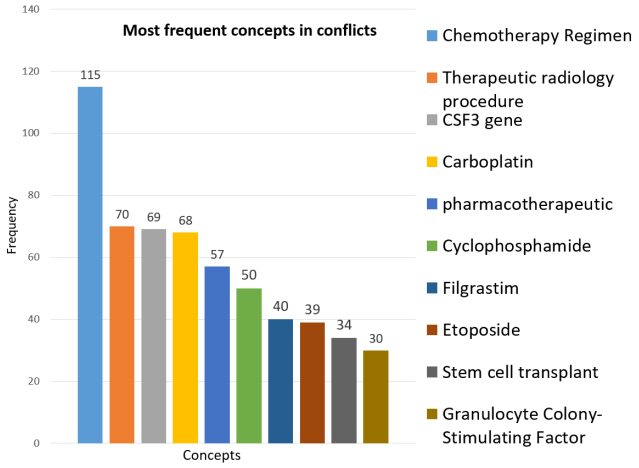


**Figure 8. UMLS Metathesaurus concepts involved in the most conflicts**

As seen in this chart, chemotherapy treatment dominates the most frequent concepts involved in conflicts: carboplatin, cyclophosphamide, etoposide, and "chemotherapy regimen". Stem cell transplant, which is the 9th most common concept, is in essence also a chemotherapy regimen. In many cases, the issue expressed by the EC has to do with the immunosuppressive effects of chemotherapy, as can be seen in the example EC in section 2. In some cases, the issue is not clear from the EC, as in "At least 3 months since therapy with etoposide, carboplatin, or ifosfamide" (NCT00003557). Note that in both cases, the exclusion is time based, so a patient who was treated with carboplatin on a prior trial is not excluded from NCT00003557 permanently. In the next phase of our project, we plan to improve parsing to highlight time based exclusions when the information is presented to patients. Concerns about immunosuppression can also be seen in three of the other top conflict generating concepts: CSF3 gene, filgrastim, and "granulocyte colony stimulating factor", which in UMLS are all synonyms for filgrastim, which is often used to stimulate neutrophil production after high dose chemotherapy.

The second most common concept included in conflicts is "Therapeutic radiology procedure" (C1522449). An example of an EC containing this term is "No prior extensive radiotherapy" (e.g. craniospinal irradiation, total body irradiation or radiotherapy to more than half of the pelvis) from NCT00030108. Finally, the term "pharmacotherapeutic" (C0013217) is very general and matches with any form of treatment including drugs, chemicals and antibodies. In sum, the types of treatment most likely to prevent patients from entering new trials are chemotherapies or other immunosuppressive therapies, and radiation therapy. Many of the ECs are time based. These categories can be used to sort the conflicts presented to patients and to indicate which trials they are not permanently excluded from.

## 5. CONCLUSIONS

Patients and their caregivers increasingly participate as partners with their doctors when making treatment decisions. Patient/caregivers often search for clinical trials on sites designed for patients. For patients with serious diseases that may require advanced treatment over a long period of time, understanding how participation in one trial may affect ability to later participate in other trials is important. We have developed a method to identify the potential for these treatment conflicts, based on comparing interventions in one trial against ECs from other trials parsed into a structured representation. This methodology has been applied to a set of Wilm's Tumor trials. We analyze results according to types of trials and treatments most likely to preclude participation in subsequent trials; this information can be used in a clinical trial search tool to better highlight and sort results for a user. In future research, we plan to improve our parsing methods to identify ECs that are time-limited, expand the framework to clinical trials from other domains, and integrate this method into a patient-focused clinical trial search tool.

## 6. REFERENCES


[1] A.R. Aronson and F. Lang, "An overview of MetaMap: Historical Perspective and Recent Advances", Journal of the American Medical Informatics Association, 17(3), 229-236, 2010.

[2] Children's Neuroblastoma Cancer Foundation, "CNCF Parent Handbook", http://www.cncfhope.org/CNCF_Neuroblastoma_Parent_Handbook, accessed on Feb. 19, 2017.

[3] D. Ferrucci and A. Lally, "UIMA: an architectural approach to unstructured information processing in the corporate research environment," *Natural Language Engineering,* 10(3-4):327–348, 2004.

[4] D. Grady, "How to Find Clinical Trials for Experimental Cancer Treatments", New York Times, December 23, 2016.

[5] K.E. Holm, J.M. Patterson, J.G. Gurney, "Results from a Qualitative Study of Parental Involvement and Family-Centered Care in the Diagnostic and Treatment Phases of Childhood Cancer", Journal of Pediatric Oncology Nursing, 20(6), 301-313, 2003.

[6] P. Kluegl, M. Toepfer, P.D. Beck, G. Fette, F. Puppe, "UIMA Ruta: Rapid Development of Rule-based Information Extraction Applications," *Natural Language Engineering*, 22(1):1-40, 2016.

[7] J.W. Mack, J. Wolfe, E.F. Cook, H.E. Grier, P.D. Cleary, J.C. Week, "Parents Roles in Decision Making for Children With Cancer in the First Year of Cancer Treatment", Journal Of Clinical Oncology, 29(15), 2085-2090, 2011.

[8] B. MacKellar and C. Schweikert, "Patterns for Conflict Identification in Clinical Trial Eligibility Criteria," *Proceedings of the IEEE 18th International Conference one-Health Networking, Applications and Services (Healthcom)*, 568-573, 2016.

[9] B. MacKellar, C. Schweikert, & S.A. Chun, "Patient-centered Clinical Trials Decision Support using Linked Open Data," *International Journal of Software Science and Computational Intelligence*, 6(3), 31-49, July-September 2014.

[10] K. Milian, A. Bucur, and F. van Harmelen. Building a library of eligibility criteria to support design of clinical trials. In *EKAW, Lecture Notes in Computer Science*, pages 327-336. Springer, 2012.

[11] K. Milian, R. Hoekstra, A. Bucur, A. Ten Teije, F. van Harmelen, J. Paulissen, Enhancing reuse of structured eligibility criteria and supporting their relaxation. *Journal of Biomedical Informatics,* 56 (2015) 205-219, 2015.

[12] K. Milian, R. Hoekstra, A. Ten Teije, F. van Harmelen, "Patterns of Clinical Trial Eligibility Criteria", *Proceedings of the AIME'11 Workshop on Knowledge Representation for Healthcare (KR4HC11), Lecture Notes AI*, 2011.

[13] G.K. Savova et al., Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5), 507-513, 2010.

[14] R.A. Schoot, C. H. van Ommen, H. N. Caron, W. J. E. Tissing, M. D. van de Wetering, and SKION Aristocaths supportive care group the Netherlands, "Accrual in supportive care trials in pediatric oncology, a challenge!," *Support. Care Cancer*, 20(12), 3149–53, Dec. 2012.

[15] S. W. Tu et al., "A practical method for transforming free-text eligibility criteria into computable criteria." *Journal of Biomedical Informatics*, vol. 44, no. 2, pp. 239–50, Apr. 2011.

[16] U.S. National Library of Medicine, *Unified Medical Language System*, retrieved Jan 1, 2017, http://www.nlm.nih.gov/research/umls.

[17] C. Weng et al., "EliXR: An Approach to Eligibility Criteria Extraction and Representation." *Journal of the American Medical Informatics Association: JAMIA* 18. Suppl 1 (2011): i116–i124. PMC. Web. 14 July 2015.

[18] D.A. Zarin, R.J. Williams, A.M. Bergeris, H. Dobbins, N.C. Ide, R.F. Loane, T. Tse, "ClinicalTrials.gov and Related Projects: Improving Access to Information about Clinical Trials: A Report to the Board of Scientific Counselors", Lister Hill National Center for Biomedical Communications, US National Library of Medicine, April 2013, Technical Report TR-2013-001.


**COMPETING INTERESTS:** The authors have declared that no competing interests exist.