Journal of Biomedical Informatics 42 (2009) 790–800

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Towards identifying intervention arms in randomized controlled trials: Extracting coordinating constructions

Grace Yuet-Chee Chung

Centre for Health Informatics, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 15 May 2008
Available online 4 January 2009

Keywords:
Information extraction
Biomedical text mining
Biomedical natural langauge processing
Medical informatics

ABSTRACT

Background: Large numbers of reports of randomized controlled trials (RCTs) are published each year, and it is becoming increasingly difficult for clinicians practicing evidence-based medicine to find answers to clinical questions. The automatic machine extraction of RCT experimental details, including design methodology and outcomes, could help clinicians and reviewers locate relevant studies more rapidly and easily. *Aim:* This paper investigates how the comparison of interventions is documented in the abstracts of published RCTs. The ultimate goal is to use automated text mining to locate each intervention arm of a trial. This preliminary work aims to identify coordinating constructions, which are prevalent in the expression of intervention comparisons. *Methods* and results: An analysis of the types of constructs that describe the allocation of intervention arms is conducted, revealing that the compared interventions are predominantly embedded in coordinating constructions. A method is developed for identifying the descriptions of the assignment of treatment arms in clinical trials, using a full sentence parser to locate coordinating constructions and a statistical classifier for labeling positive examples. Predicate-argument structures are used along with other linguistic features with a maximum entropy classifier. An *F*-score of 0.78 is obtained for labeling relevant coordinating constructions in an independent test set. *Conclusions:* The intervention arms of a randomized controlled trials can be identified by machine extraction incorporating syntactic features derived from full sentence parsing.

1. Introduction

Clinicians face many challenges in practicing evidence-based medicine (EBM) [49,45], with the task of finding correct answers to their clinical questions impeded by the massive expansion of the biomedical literature. For example, the most reliable primary evidence for the safety and efficacy of treatment interventions are documented in randomized controlled trials (RCTs) [39]. There are over 230,000 RCT citation entries in PubMed. The publication rate is exponentially rising [53,37] with over 12,000 trials published over the last year. Furthermore, many studies have reported that clinicians lack the time and skills to locate and synthesize the best evidence from the volumes of literature [18,25,26,20,29,11].

Resources such as the Cochrane Collaboration [15], Evidence-Based Medicine [27], the ACP Journal Club [1] and BMJ Clinical Evidence [8], help practitioners find up-to-date answers through manually compiled summaries that have been assembled from extensive searches and critical assessments of the RCT literature. Other efforts have been based on improving the quality and consistency of reporting of RCTs to help readers navigate information more easily [9,42,4,6,2].

An alternative to this is the use of information technology that could increase the likelihood for clinicians in finding correct answers to their queries at the point of care. This could be achieved by better search engine design, improved indexing, information retrieval and even natural language processing systems such as automatic summarization. Much work has already been reported in systems that help users locate and navigate clinical information [14,17,24,30,34,38]. However few of these systems have taken advantage of natural language processing techniques to enhance retrieval.

The overall goal of this work is to use automated text mining to extract the methodological details of each RCT. The automatically extracted elements would be useful as meta-data for indexing RCT documents, and could be synthesized into informative summaries that help both clinicians and systematic reviewers attempting to locate specific studies [13].

According to the principles of sound RCT experimental design [39], usually two or more interventions or therapies are randomly assigned to roughly equal sized population subgroups with similar baseline characteristics. This random assignment of therapies under comparison underpins the design of all RCTs, and is a critical

E-mail address: gigigraceyc@gmail.com
URL: http://www.chi.unsw.edu.au

1532-0464/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2008.12.011

piece of information when a clinician/reviewer is assessing the relevance of a clinical study. Ultimately, we aim to automatically extract the exact intervention arms by text mining. The results could be integrated into a complete system for identifying all experimental details (methodology and outcomes) of RCTs. The collated facts could be used in automatic summaries for clinicians practicing EBM or systematic reviewers seeking to appraise studies and conduct meta-analyses.

In this preliminary work, we explore the viability of exactly identifying intervention arms by finding and classifying a linguistic construct prevalent in expressing treatment comparisons: coordinating constructions. A study on the characteristics of sentences within abstracts of RCTs which describe the comparison of interventions is conducted. This study covers a detailed analysis of the lexical and syntactic patterns for descriptions of comparisons. In the second part of this paper, we describe a system and method for the automatic extraction of comparison descriptions based on the integration of linguistic knowledge from a full sentence parser within a statistical classifier.

The structure of this paper is as follows. Section 2.1 will provide some background information. We first introduce clinical trials and key concepts in interpreting and reporting of RCTs and secondly, we briefly introduce coordinating constructions as a linguistic concept. In Section 3 we present some related work on mining of textual content for EBM. Section 4 describes the initial data collection. In Section 5, an analysis of various types of constructs used to express intervention comparisons is presented. In particular, we examine the lexical and syntactic patterns for single intervention sentences. Then the method for identifying interventions and extracting the comparison constructs is described in Section 6. Experimental results and interpretations of them are elucidated in Section 7. Section 8 discusses limitations, and Section 9 concludes our work.

2. Background material

2.1. Randomized controlled trials

The primary evidence for the efficacy of treatments is documented in RCT reports. High quality RCTs are designed to provide specific and statistically robust answers about the impact of a clinical treatment or intervention on factors such as patients' prognosis or quality of life. According to [39], an RCT may be defined as: "A prospective scientific experiment comparing the value of a treatment strategy in an experimental group with an alternative strategy in a control group, in which allocation to experimental or control group is determined by a chance mechanism." Treatment strategies being assessed could be pharmaceutical interventions, devices, surgical procedures, behavioral therapies or social interventions.

A number of key concepts are crucial to the sound design of RCTs. Many of these are set out in the CONSORT statement [42]. This is a comprehensive guideline for conducting and reporting RCTs, developed by an international group of clinical trialists and biomedical experts, comprising a checklist of 22 items and a participant flow diagram. A central facet of RCT design is the randomised allocation or assignment of each subject to an experimental or control group. These are referred in this paper as the intervention or comparison arms of the trial. The arms being compared correspond with each experimental group in the trial. They generally include one or more arms associated with the therapy being assessed, and an arm associated with the control group which could, for instance, be achieved from a placebo treatment. Randomization in which patients have an equal chance of being assigned to any group mitigates against selection bias, or preference to any experimental group by either the clinician or the patient. Careful randomized allocation ensures bias-free results.

The CONSORT statement recommends the reporting of precise details of the interventions intended for each group, including how and when they were administered. This paper specifically addresses the identification and extraction of pharmaceutical interventions being compared.

2.2. Coordinating constructions

Coordinating constructions consist of two or more constituents of coordinated phrases that are linked by coordinating conjunctions (e.g. *and, or, but*) [32] (e.g. "digoxin or placebo.") The coordinating constructions can be identified based on their symmetry. The constituents should occupy the same status and if one constituent is significantly more salient than another, the construction is not coordinated. The constituents or coordinated phrases could be noun phrases, verb phrases or clauses.

Correct handling of coordinating constructions has historically posed large challenges in computational linguistics [19,35,3,43,10,31]. For correct syntactic analysis and extraction of the construction, one needs to identify the correct boundaries of each constituent and particularly the correct attachment of other constructs such as prepositional phrases. These are often problematic because boundaries can be highly ambiguous.

In biomedical text data, coordinating constructions particularly concerning noun phrases, frequently occur. Much of the community has been concerned with the phenomenon of ellipsis where relevant information is omitted from the surface expressions. For instance, in "10 or 20 mg", the entities "10 mg" and "20 mg" are coordinated and "mg" is omitted from the first part of the expression. The unfolding of coordinating constructions that have various forms of ellipses remains an active area of research [43,10,31].

3. Related work

There is very little prior work on the fine-grained extraction of specific parameters for RCTs. It has been suggested that the PICO framework [48] would be suitable as an information extraction model for clinical studies. PICO is a task-based model for EBM, formulated to assist EBM practitioners to articulate well-formed questions in order to find useful answers in clinical scenarios. The PICO elements are Patient/Population, Intervention, Comparison and Outcome. These have been adopted by some [50,23,44] as a guide to the important key facts to be extracted from the reports of clinical studies. Extracting these elements would help a system come up with answers to typical clinical questions posed by physicians. Demner-Fushman [23] has implemented an intervention extractor that relies heavily on knowledge derived from UMLS semantic categories [40], using Metamap [5]. The interventions for each study are identified, but there was no attempt to distinguish a primary intervention versus a comparison, and there is no attempt to extract the exact intervention arms. Indeed, doubts have been raised about the effectiveness of PICO as a representation for expressing realistic clinical problems and answering clinical queries [36].

Dawes et al. [22] claim that key elements of clinical studies extend beyond PICO and propose a new scheme: Patient–Population-Problem, Exposure-Intervention, Comparison, Outcome, Duration and Results (PECODR). This scheme is a finer grained structure in which duration of exposure is added. While no attempt is made to implement text mining, the study demonstrates that a consistent structure exists in reports of clinical studies.

In the related work of Fiszman et al. [28], a semantic processor was used to extract and interpret comparative constructions which are prevalent in clinical trial reports. Two kinds of statements were examined. The first kind was the comparison of therapies, and the second kind were linguistic constructs that express a comparative relationship between entities. The second kind often appears in outcome statements. By contrast, our work has exploited an alternative linguistic construct, coordinating constructions which commonly occur, and can capture the comparison of treatments.

In general, a growing community of researchers is tackling the problem of text mining of research articles [37] in MEDLINE and in the biomedical domain. The contribution of this work will focus on clinical studies, and exploit the fact that their experimental methodology follows certain guidelines. In particular, RCTs by definition compare at least two treatment options for a particular patient population group. This key piece of information is critical when selecting RCTs for finding clinical answers and also for selecting RCTs as part of the process of meta-analysis and critical appraisal.

The method described in this paper relies heavily on full parsing to extract syntactic information from sentences. Few researchers have successfully employed full parsers [57,46,21], and these have been applied in the domain of documented biological interactions. In contrast, here a full parser used for processing clinical reports and the derived syntactic information is used in a statistical classifier.

4. Data collection and preliminary annotation

In the initial phase of this work, abstracts were compiled from a search on PubMed, and annotation was performed on the collected data.

4.1. Data collection

Because there has been no previous work on this problem, no large annotated corpus of intervention arms in RCTs exists. Thus the first stage involves compiling a corpus for training and development purposes. A broad search was conducted in MEDLINE for randomized controlled trials, published from 1998 to 2006, with the following keywords: asthma, diabetes, breast cancer, prostate cancer, erectile dysfunction, heart failure, cardiovascular, angina. These reflect a typical set of clinical conditions. We relied on the publication type field in PubMed to select RCTs. Fifty RCTs were then randomly selected. From these, the RCTs dealing with only pharmaceutical interventions were manually selected. For this task, the author and a domain expert performed the selection on a subset of 50 abstracts. A Kappa agreement [16] of 0.9 was achieved on this subset and the remaining selection of the subset was undertaken by the author. As a result the initial set contain 206 RCTs, used for initial analyses and subsequently as training set in these experiments.

A second set of abstracts was compiled later to serve as an unseen test set. A search on MEDLINE specifying MESH terms for "Drug Therapy", "Randomized Controlled Trials" and articles published in English in 2007 was conducted. This search yielded 124 abstracts and is used as an independent non-overlapping test set.

4.2. Annotation

The author first annotated the training corpus with the sentence in the abstract that specifies the comparison of the interventions, particularly the allocation or assignment of two or more interventions to the respective population groups, thus clearly indicating the number of arms in the experimental design. An example annotation is:

> <sent>Patients were randomly allocated to allopurinol or placebo.</sent>

These sentences were often found in the "Method" section of an abstract.

Further examination of these sentences was conducted to identify the characteristics of sentences that introduce the assignment, allocation and comparison of the pharmaceutical interventions. We focused on constructs that explicate the actual assignment usually under the "Method" heading rather than implicit references made in the Aims, Results or Conclusions sections. It was determined that frequently, the interventions being compared co-occur within a single phrase (noun, verb or prepositional). In particular, coordinating constructions were prevalent. Results of the analysis are described in Section 5. As a result, coordinating constructions of RCT abstracts that identify the allocation of pharmaceutical interventions were labeled in the corpus. An example of the labeling for a sentence is shown as follows:

> <sent>Patients were randomly allocated to <interv>allopurinol or placebo</interv>.</sent>

5. Analysis of intervention sentences

In this section, a qualitative analysis conducted by the author on the training data is presented. Section 5.1 gives an overview of the characteristics of the way intervention arms are typically described. Section 5.2 focuses on the linguistic patterns found for single intervention sentences. In the detailed analysis, we provide the various categories of coordination found: noun phrase coordination, verb phrase coordination, adjectival coordination, adverbial coordination, prepositional coordination, and sentence coordination. Examples will be given for each type of coordination found. We will also examine cases where coordination was absent.

5.1. Description of interventions

Manual examination of the training data was performed, and a qualitative analysis of the intervention comparison descriptions that have been labeled in the abstracts was conducted.

Based on the annotated data set of pharmaceutical RCTs, it was found that the allocation of intervention arms often involves more than simply comparing two or more types of drugs and that the experimental design involves varying a parameter that pertains to other aspects of drug treatments. These were classified into the following types of comparisons:

- Comparing the administration of two drug therapies for effectiveness and efficacy.
- Comparing a single drug therapy with either a placebo, no drug administration or some other continuation of standard treatment.
- Comparing the dosages of drug therapies, where a different dosage amount represents a different intervention arm or otherwise the drug is administered on a different schedule on each arm:

> "Patients... were randomly assigned to receive PST of paclitaxel doses administered either weekly (for a total of 12 doses of paclitaxel) or once every 3 weeks (four cycles)..."

Each RCT could have multiple arms each with a combination of either/both of the above comparison types. Sentences that describe the allocation to arms are also embedded with descriptions of the route of administration, the frequency of dosage, the dosage amount and the duration of therapy, e.g.:

> "...20 mg atorvastatin or placebo once a day for 45 days."

All the abstracts contain some description explicit or implicit of the interventions that are being studied. For the three of the 206 abstracts, it was found that the assignment of interventions are described across multiple sentences where two patient groups are separately defined and the treatment allocated to each group was described in separate sentences. This type of description was more common for non-pharmaceutical interventions, and seemed

Table 1
Common constructs found in noun phrase coordination embedding the description of intervention allocation. For each category, examples of common cue verbs and nouns are shown in the left hand column, and some sentence examples of each category is given in the right hand column. The coordinating construction considered as the intervention comparison are underlined.

Construct	Examples
1a. Object of verbs in passive voice	
Given, Randomized for, Injected with, Placed on, Supplemented with, Allocated to, Assigned to	"Patients were randomly allocated to allopurinol 300 mg/d or placebo."
1b. Object of verbs in active voice	
Randomized to receive, Randomized to take, Inhaled	"...440 men... were randomized to take placebo or 10 or 20 mg vardenafil."
2. Verb subject	
Administered	"Opioid receptor antagonist naloxone or placebo was administered intravenously..."
3. Auxiliary complements	"Treatments were nortriptyline, fluoxetine, or placebo..."
4. Prepositional phrases attached to nominal forms	
Treatment with, Doses of, Sedation with, Addition of, Injection of, Infusion of	"Nineteen patients were randomized to blinded infusion of glutamate or saline."
	"Subjects were randomized to 13 weeks of 4 mg t.i.d. repaglinide or metformin..."
5. Verbless fragments	
List after ":"	"Placebo or study medications were administered as follows: immediate-release glipizide 30 min before breakfast and 30 min before supper, glipizide gastrointestinal therapeutic system (GITS) 30 min before breakfast, or nateglinide 120 mg 10 min before breakfast..."
Parenthetical remark	"Patients were randomly allocated to treatment with talinolol (100, 200 or 300 mg once daily) or placebo."

to be an anomaly for drug therapies. Hence these three abstracts were discarded from the set in the mean time.

It should be noted that inferences to the intervention arms can be drawn throughout an abstract, in the title, or in statements of experimental aims or in reporting of results, as exemplified below.

> "We used a novel radioimmunoassay to evaluate the effect of nateglinide on plasma concentrations of glycated insulin and glucose tolerance..."

> "Plasma glucose and glycated insulin responses were reduced... following nateglinide compared with placebo."

But these were incomplete in precisely and fully describing the allocation of intervention arms.

5.2. Lexical and syntactic patterns for single intervention sentences

The lexical and syntactic patterns for the 203 abstracts that contain single sentences describing intervention comparisons were examined. Upon examination of the set of single intervention sentences, it was found that the annotated phrases of comparison descriptions occur predominantly with coordinating constructions; these constructions appear to be the natural way to present the arms of comparison.

In the following, some broad classes of syntactic/lexical patterns for the constructs that capture the descriptions are provided. Noun phrase coordinations by far outnumber the other constructs although there are small numbers of other kinds of coordinating constructions.

5.2.1. Noun phrase coordination

The most common structure is noun phrase coordination encapsulating the drug interventions on each side of the coordinating conjunction. One hundred and seventy (83%) of the 203 abstracts contain a noun phrase coordination indicating the drug interventions. The most common constructs under which the coordination occur are described below and summarized in Table 1.

(1) Direct/Indirect Objects. The coordinating constructions are in object position relative to some typical verbs listed in Table 1. Both passive and active voice examples are given.

(2) Verb Subjects. The coordinating constructions occur in the subject position.

(3) Auxiliary Complements. The constructions occur as complements to auxiliary verbs.

(4) Prepositional Phrases. Simple noun phrase coordinations also commonly occur in prepositional phrases attached to nominal forms. The nouns may refer to the method of administration or some other aspect of the drug therapy such as the duration of therapy.

There are also many occurrences of complex compound noun phrase coordination with potentially ambiguous prepositional phrase attachments embedding critical information such as the actual pharmaceutical intervention, e.g.:

"196 women with HER-2-overexpressing MBC were randomly assigned to six cycles of either trastuzumab 4 mg/kg loading dose plus 2 mg/kg weekly thereafter with paclitaxel 175 mg/m^2 every 3 weeks (TP), or trastuzumab 4 mg/kg loading dose plus 2 mg/kg weekly thereafter with paclitaxel 175 mg/m^2 and carboplatin area under the time–concentration curve = 6 every 3 weeks (TPC) followed by weekly trastuzumab alone."

Many more examples contain multiple nested coordinating constructions which would make correct bracketing thereby eliciting the intervention arms a very challenging task, e.g.:

> "...[a target controlled infusion of [propofol] and [sufentanil]] or [remifentanil infusion]."

(5) Verbless fragments. In some examples, 16 of 203 abstracts (7.9%), the coordinating noun phrase denoting the element that is varied at each treatment arm, does not appear within a clause but after a colon or semi-colon as a list, or within a parenthetical remark. As in the parenthetical remark example given in Table 1, while the parameter being varied for each arm is given in the coordinating phrase, the actual drug intervention is in the surrounding context.

In the following example "[[10 or 20] mg vardenafil]," one observes the phenomenon of ellipsis in which some words ("mg vardenafil") are deleted from the surface realization, and the true meaning actually implies that the two arms are "10 mg vardenafil" and "20 mg vardenafil". Clearly, for each coordinating construction, it is necessary to properly identify the true constituent constructs and resolve the elliptical expressions.

This proper "unfolding" of coordinating constructions, is needed to identify the arms of a clinical trial, and could require further pragmatic and contextual information for correct resolution. In general, handling of coordination ellipsis has only been recently addressed in the computational linguistics community [43,10,31] and is beyond the scope of this work in this preliminary study.

5.2.2. Verb phrase coordination

Verb phrase coordination, found in 3% of the set, occurs where the method of administration may vary from one arm to another, e.g.:

> "Patients with CHF were then randomized to maintain standard treatment, double the ACE inhibitor dose or add an angiotensin II antagonist."

Another example shows an ellipsis where one coordinating conjunct includes a negation:

> "Patients... were randomized to [[receive] or [not receive] rosiglitazone] for 6 months."

The unfolded constituents would be "receive rosiglitazone" or "not receive rosiglitazone".

5.2.3. Adjectival coordination

Adjectival coordination, found in 1% of the set, occurs when describing variation in dosage or other parameter. The drug therapy would follow the adjectival coordination, e.g.:

> "Asthmatic patients... were randomly allocated to use either a [[short-acting] or [long-acting] beta2-agonist]"

or

> "Adolescents... were randomly assigned to groups with either [[lower] or [higher] than 9% glycosylated hemoglobin (HbA1c)]..."

In both the above cases, the coordinating adjectives alone do not describe the intervening treatment which can be located within the noun phrase that subsumes the coordinating adjectives.

5.2.4. Adverbial coordination

In some cases, the parameter of variation is described by an adverbial coordination, in an adverbial phrase following the drug therapy. This is found in 1% of the set, e.g.:

> "...either weekly or once every 3 weeks..."

5.2.5. Prepositional coordination

Prepositional coordination, found in 2.5% of the set, occurs where entire prepositional phrases are coordinated, e.g.:

> "...adults were supplemented for 6 months with 1000 μg/day of Cr (as Cr yeast) or with a placebo."

> Alternatively words may be deleted in the case of ellipsis, e.g.:

> "...therapy... with or without metformin."

In the above, the intervening treatment is found within the prepositional phrase that subsumes the coordinating prepositions.

5.2.6. Sentence coordination

In sentence coordination occurring in 3% of the set, each intervention arm is described explicitly on each side of the coordinating conjunction, e.g.:

> "34 patients received once-daily diltiazem and 33 patients received amlodipine."

5.2.7. Examples with the absence of coordination

In 13 of the 203 abstracts (6.4%), instances are found where coordinating constructions are absent from the abstracts and interventions are described in alternate ways.

(1) Comparisons. In some abstracts, the most explicit way to describe the intervention, appearing either in the "Objectives" or "Method" sentence, is the mention of the drug therapies of comparison as follows:

> "Efficacy and safety of toremifene... was compared with tamoxifen... in a group of postmenopausal women with advanced breast cancer, without previous systemic therapy for advanced breast cancer."

Many abstracts contain statements of this nature as an assertion of the purpose of the study, with or without an explicit explanation of the actual intervention arms which must be inferred.

(2) Versus. Versus is a cue word that signifies the comparison of two treatments in a trial but this does not explicitly describe assignment, e.g.:

> "...study compared the effect of montelukast versus placebo for 4 weeks..."

(3) Other instances. In certain cases, the abstract contains no references to assignment to each treatment. In the example below, one could assume a two-armed trial between sildenafil and placebo although it can only be certain from reading the full text:

> "...a randomized, 12-week, double-blind, placebo-controlled, flexible-dose escalation study of sildenafil."

In general, it is found from the corpus that coordinating constructions can be a good cue as it is a construct that embeds the interventions being compared within a randomized control trial. The construct is commonly used to express the comparison of two or more pharmaceutical substances or for expressing alternate ways or dosages for administration of a substance. Thus, in this first study, we set out to explore the viability of extracting the intervention comparison simply by finding the coordinating constructions in the abstract and classifying them as a relevant intervention comparison or not. With the advent of full syntactic parsers, it may now be possible to elicit the coordinating construction, identify the segment boundaries and resolve sub-components of the construct, which in this domain would be the intervention arms of the RCT.

It is well-known that complex syntactic structures pose great challenges for full sentence deep syntactic analyses. As exemplified above, there are some main confounding factors associated with the coordinating constructions in the RCT abstracts. First, much quantitative information such as varying dosages for each drug is embedded within the noun phrases. Second, while drug therapies primarily occur in coordinating constructions, many do occur in complex compound phrases, particularly where complex combinations of drugs are used on the different arms. Many also have multiple prepositional phrase attachments which are notoriously hard for parsers.

As in the examples illustrated above, the phenomenon of ellipses is prevalent for each type of coordination and a method would need to be adopted to unfold the coordinated structure for extracting each arm. We do not attempt to unfold the coordinates in this preliminary work.

While some abstracts do not use coordination to describe the interventions, those structures often do not explicitly describe the allocation of treatment arms. Since the goal is to identify the structures that indicate how each arm has been assigned as precisely as possible, the focus here is on the extraction of relevant coordinating constructions as a first pilot experiment.

6. Method for identifying the interventions: extracting coordinating constructions

We have developed a method for the identification of interventions of comparison by labeling the coordinating conjunctions that capture the comparison within the abstracts. Fig. 1 depicts the approach. To simplify the approach, it is assumed that the structure occurs only in "Method" sentences.

To begin with, abstracts are segmented into sentences via a sentence splitter and the sentences in the abstract that could be construed as "Method" sentences are identified by a "Method" sentence labeler, described in Section 6.1. From all the "Method" tagged sentences, a sentence normalizer is used to produce canonicalized forms for certain expressions in the sentences. This is described in Section 6.2. Following this, a full syntactic parser is used to identify all the coordinating constructions, described in Section 6.3. The coordinating constructions are then labeled as relevant or irrelevant by a statistical classifier trained with word-based and linguistic features derived from the parser. This is described in Section 6.4. In the final stage, the relevant coordinating constructions are ones indicating the RCT intervention arms. These need to be unfolded into the component arms of the RCT experimental design and is not addressed here.

6.1. Identifying method sentences

For the data sets, abstracts are either structured in which sentences appear under distinct, labeled sections or unstructured whereby these section subheadings are absent. Structured abstracts, mandated by some journals and recommended by CONSORT, are intended for improved information retrieval and readability [2,33]. As previously reported [13], the names for subheadings in structured abstracts vary widely and many terms are synonymous, e.g. *Patients* and *Participants*. Our first step here is to identify those sentences that refer to methodology in RCT design. Initially, every abstract is processed by a publicly available state-of-the-art sentence splitter [52] that is optimized on biomedical documents. We adopt a different approach for structured and unstructured abstracts.

For structured abstracts, a mapping is constructed to map all related subheadings to the "Method" subheading. Table 2 lists some of the related subheadings. Those sentences that fall under the "Method" equivalent subheading are extracted for further processing.

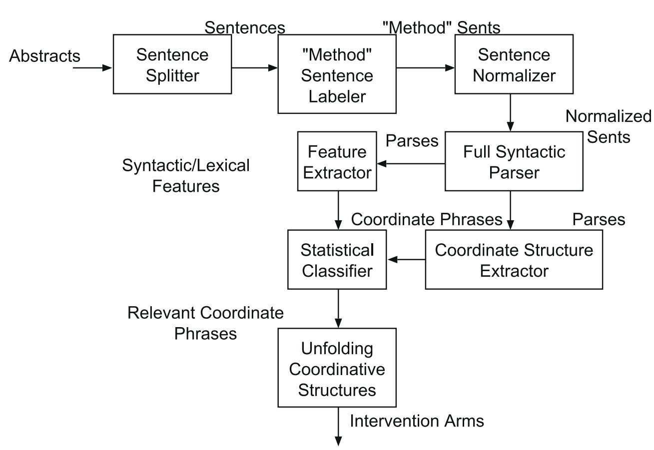

Fig. 1. System architecture for the extraction of coordinating constructions relevant to the assignment of intervention arms in randomized controlled trials.

Table 2
Examples of subheadings in structured abstracts that are mapped to the "Method" subheading. Combinations of each of these subheadings are also common.

Mapped heading names		
Participants	Intervention	Statistical methods
Patients	Procedures	Materials and methods
Subjects	Approach	Measurements
Population	Experimental design	Primary outcome parameters
Setting	Study design	Main outcome measures
Study setting	Design	Endpoints

For unstructured abstracts, an approach previously developed in [13] is used. Essentially, the approach uses statistical classification to label sentences of an abstract to one of four classes: AIM, METHOD, RESULTS, CONCLUSION. We provide a brief summary of the classification approach below.

The classifier is trained on a large corpus of structured abstracts mined from PubMed. The subheadings of the training set of structured abstracts are mapped to one of four classes: AIM, METHOD, RESULTS, CONCLUSION. Prior to classification, every sentence undergoes a normalization stage in which numbers, currencies, dates/times, measurements, statistical and mathematical symbols are reduced to canonical forms.

The classification method chosen is conditional random fields (CRF) [41,51] which model the discourse topics of an abstract as a sequential machine. The classifier is a 4-class linear chain CRF. These are undirected graphical models that are discriminatively trained to maximize conditional probability of a set of output variables given a set of input variables. The input to the classifier is a feature set which includes a simple bag-of-words with the canonical forms in place and part-of-speech tags derived from the GENIA tagger [56].

For experiments in this paper, the training corpus was derived from a random subset of all randomized controlled trial abstracts from 1998 to 2006. It numbers 13,605 abstracts, with 156,622 sentences. Under 15-fold cross-validation, the error rate of the 4-way classifier is 6.48%. The *F*-score for the "Method" section is 0.92. This classification approach was used to automatically extract "Method"-related sentences from the unstructured abstracts. The extracted sentences, from both structured and unstructured abstracts, will be further processed in subsequent experiments for extracting intervention arms, described hereon.

6.2. Normalization and parsing

The sentences are preprocessed via a normalization script which uses regular expressions to replace complex numerical or mathematical notation into a canonical form. The purpose is to reduce the complexity of these expressions, characteristic of clinical studies reporting so that fewer errors are made by the syntactic parser. The following entities are normalized:

- Numerical Values. All integers (cardinals and ordinals) and real numbers are reduced to respective canonical forms. All integers and real numbers are mapped to symbols INT and REAL.
- Measurements. All entities that represent measurements of dosage, volumes, capacities, weights, their ratios (e.g.*kg, cm, μmol/l, l/min, mg/m^2/d*) and their ranges (e.g. *5–10 mg*) are normalized via a regular expressions script. For instance, *5 mg* is mapped to MEASUREVALUE and *5–10 mg* is mapped to MEASURERANGE The most prevalent of these refer to dosage amounts and their frequency administered for interventions. Some types of abbreviations are specific to the domain of clinical abstracts and drug administration, e.g.: *t.i.d., b.i.d.*

  "rosiglitazone (4 mg b.i.d.)"

- Statistical and other mathematical notation. Example entities are those that denote confidence intervals e.g. *0.08% (95% CI:*

−0.48%, 0.64%), statistical significance *p*-values, risk ratios, and population counts e.g. $n = 100$. However, most statistical expressions rarely appear in "Method" type sentences. For instance, expressions such as $p < 0.05$ are mapped to PVALUE and $n = 100$ is mapped to POPN.

- A closed set of patient-related words are mapped to a single word "Patients" word. The closed set includes common words such as: *men, women, subjects, participants, adults, adolescents, children, outpatients, inpatients*...
- Time and Date Expressions. The script replaces expressions such as `January 1995` with a canonical form DATE.
- Monetary Expressions. For instance *USD $100* is replaced with a canonical form CURRENCY.

6.3. Extracting coordinating constructions from parse trees

The output of the normalizer are "Method" sentences which are input to a full parser. Enju [54], a head-driven phrase structure grammar parser (HPSG) is used. As a full syntactic parser, it recognizes verb subject and object information, encapsulating long distance syntactic relations. Additionally, linguistic rules transform syntactic structure into a set of predicate-argument structures (PAS) which are normalized forms representing syntactic relations. PASs are intended to collapse multiple surface realizations into a canonical form capturing some semantic relationships, including but not limited to, verb predicates and their respective arguments. Each PAS is essentially a tuple consisting of a predicate with one or more argument constituents. Fig. 2 depicts an example parse tree along with the associated PAS. Trained on the GENIA treebank [55], Enju has been used for information extraction tasks in the biomedical domain applied to the extraction of protein–protein interactions [57].

The output of the Enju parser is formatted in XML. We extract from the parse tree, all the coordinating phrases associated with each sentence. In many cases, coordinating phrases are nested within others, especially for lists with three or more elements. For instance, "pioglitazone, atorvastatin or both" yields two coordinating phrases, one involving conjunction "," and conjuncts "pioglitazone" and "atorvastatin", and the second with conjunction "or", linking "pioglitazone, atorvastatin" and "both." In the hierarchical parse structure, the coordinating construction linked by "or" subsumes the one linked by ",". Each coordinating phrase is extracted individually from the parse tree.

6.4. Feature extraction and classification

A Maximum Entropy (MaxEnt) classifier is selected for this task. MaxEnt is a popular and competitive classification technique for natural language processing tasks, particularly for text classification, language modeling [12], text segmentation [7], and part-of-speech tagging [47]. It is based on assigning a class to observation features by computing probabilities from exponential functions of weighted feature sets; the maximum entropy principle estimates uniform distributions in the absence of knowledge from labeled data. This classification technique is believed to handle overlapping features well. The Mallet [41] toolkit is used in the experiments. The following describes features used in the experiments.

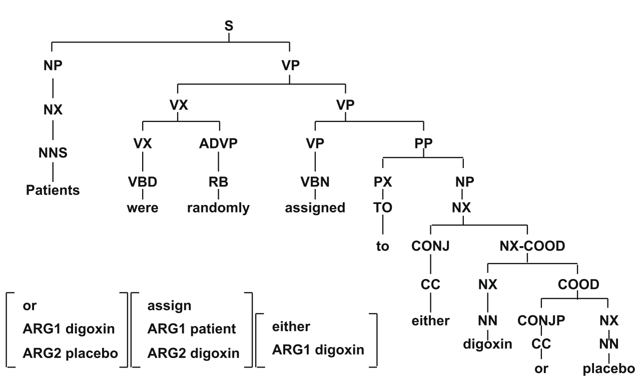

Fig. 2. Example of a parse tree and the associated predicate-argument structures for the coordinating construction and the structures connected with it.

6.4.1. Syntactic features

The type of coordinating conjunction is considered. Examples included are: *-comma-*, *plus*, *but*, *or*, *and*, *whereas*, *nor*, *than*, *and-slash-or.*

The type of coordinating phrase is considered. These are a closed set: *VP, PP, ADVP, NP, ADJP*.

The head (terminal) constituent of the left and the right coordinating phrases are included. We hypothesize that the head words are key content words that would be informative for the classifier to determine relevence. However, to avoid sparse data issues in a relatively small training set, some of these words are collapsed to broad semantic classes. The classes are generalized mappings derived from semantic types provided by UMLS [40]. The MetaMap [5] tool is used to derive concepts from the UMLS Metathesaurus. Surface forms are mapped to three broad classes that fall under Semantic Groups defined in UMLS. Only three classes are chosen as these are most likely to inform the classifier of relevance to a pharmaceutical intervention. These are: (1) Pharmaceutical substance (*Pharmacologic Substance, Clinical Drug, Inorganic Chemical, Organic Chemical, Antibiotic*), (2) Conditions and Disorders (*Pathologic Function, Disease or Syndrome, Mental or Behavioral Dysfunction, Neoplastic Process*), and (3) Medical Procedures (*Therapeutic or Preventive Procedure*).

6.4.2. Related predicate-argument structures

Fig. 2 illustrates all predicate-argument structures related to an example intervention sentence: *Patients were randomly assigned to either digoxin or placebo.* Every coordinating phrase is associated with a coordinate PAS, which resembles a tuple in which the predicate is the conjunction and the arguments are the head terminal constituents of the coordinating conjuncts. For instance as in Fig. 2, "digoxin or placebo" yields a simple PAS representation: *or*(*ARG1* = *digoxin*, *ARG2* = *placebo*). In addition, "digoxin" also appears in two other PASs. These are *assign*(*ARG1* = *patient*, *ARG2* = *digoxin*) and *either*(*ARG1* = *digoxin*) as shown in Fig. 2. These two PASs are deemed to be linked to the coordinate PAS as they share common arguments.

The PASs that are linked or connected to the coordinate PAS are included as features. That is, whenever the arguments of the coordinate PAS are featured in another PAS in the sentence (as a predicate or one of the arguments), the PAS is included as one of the features. The actual arguments that are shared are replaced by a placeholder. Thus, in the above example the structures *assign*(*ARG1* = *patient*, *ARG2* = *X*) and *either*(*ARG1* = *X*) are used as symbolic features in the classifier.

This is inspired by work in [57] in using connected chains of PASs to elicit protein–protein interactions from texts. The "connected" PASs serve to capture the verb relations which most commonly occur with the coordinating phrases.

As the PAS itself is a triple, it can be considered as a feature that resembles a trigram. In the above, *assign*(*ARG1* = *patient*, *ARG2* = *X*) is essentially a trigram (assign, patient, X). Sparse data problems are likely to arise for this small training data set. Thus, we also experiment with using only the arguments of the related PASs, omitting the predicate itself altogether, which would be analogous with a bigram type or word pair feature. Thus we would only use the simplified forms for connected or linked PAS structures (*ARG1* = *patient*, *ARG2* = *X*) and (*ARG1* = *X*).

Through experimentation, only a closed set of classes of PASs were used. These were PASs where the predicates were: verbs, prepositions, coordinates, conjunctions, complements, and auxiliaries.

6.4.3. Word-based features

In contrast with the syntactic information afforded by the parser, word-based features were also explored. The words within the coordinating constructions that were in addition to the syntactic heads described above were used as a bag-of-words features.

7. Experiments

For the training data used in the binary classification experiments, each coordinating construction extracted by the parser in the training set is manually checked by the author, and matched with the annotated reference phrase to determine whether the construction is the one identified in the reference labels. If there is a match, the coordinating construction is labeled as a positive example; otherwise, it is labeled as a negative example. For labeled noun phrases that contain a list of three or more constituents, there would be more than one coordinating phrase. For example, "repaglinide, metformin or placebo" is a single phrase consisting of three constituents, and coordinating conjunctions "," and "or." Both conjunctions would be labeled positive.

Furthermore, the boundaries of the parser-extracted phrase could contain deletion or insertion errors or a prepositional phrase (PP) could be erroneously attached. Errors in bracketing associated with PP attachments are common, and discussed later in Section 7.2.1. If the associated conjunction matches one that is in the reference phrase, then the coordinated phrase is labeled as positive, in spite of the bracketing error. For instance, "metformin or placebo twice daily" may be extracted in which "metformin" and "placebo twice daily" are the two entities incorrectly identified.

The test set was not manually annotated for the phrases/sentences containing intervention arms. After extraction of coordinating constructions by the parser, each test set phrase is manually tagged by the author as relevant or irrelevant.

The implemented system is validated in classification experiments using 10-fold cross-validation on the training set. In addition, results are reported where models are trained on the training set and evaluated on the independent test.

7.1. Results

Table 3 details the characteristics of the training and test sets. Normalized sentences (98.4%) parse successfully in Enju without failure. This compares with 95.3% success rate in Enju parsing for sentences that have not been normalized. Parse failures are often due to long sentences that exceed the limit set for number of paths.

Table 3
Characteristics of training and test sets in classification experiments.

	Train set	Test set
Number of abstracts	203	124
Number of unstructured abstracts	72 (35%)	30 (24%)
Total number of sents	2176	1332
Total number of "Method" sents	687	420
Number of coordinating constructions extracted from parse trees	1022	563
Number of abstracts with intervention arms as coordinating constructions	190	n/a
Number of correct coordinating constructions extracted from parse trees	177	139
Number of abstracts where parser found correct coordinating constructions	148	95

Table 4
Experimental results on 10-fold cross-validation on training set. For each system the precision (*P*), recall (*R*) and *F*-score for the recognition of relevant coordinating constructions are reported. Accuracy is reported for the overall classification task. Explanation of Systems 1–6 are given in the text.

Features	*P*	*R*	*F*-score	Accuracy %
[S1] conjunction	0.65	0.71	0.68	89
[S2] S1 + coordination type	0.69	0.66	0.68	89
[S3] S2 + coordinating heads	0.85	0.75	0.80	93
[S4] S3 + related PAS triples	0.88	0.78	0.83	94
[S5] S4 + related PAS arg pairs	0.88	0.81	0.84	95
[S6] S3 + words in phrase	0.85	0.75	0.80	93

The experimental framework evaluated the binary classification of coordinating constructions as positive or negative examples of phrases capturing intervention arms. Results are presented here for multiple systems which differ from one another in the input features to the classifier. System 1 describes one that uses a single classification feature, the name of the conjunction (e.g. *and*, *or*, -*comma*-, *etc*). System 2 uses the coordination phrase type (e.g. *VP, NP, etc.*) and the conjunction. The feature vector for System 3 includes the coordination phrase, the conjunction as well as the head word of the coordinating phrase. System 4 is similar to System 3 but also uses the PAS as triples in the classifier features as described in Section 6.4.2. System 5 uses a composite feature vector including the conjunction, the coordination phrase type, the head word of the phrase, the PAS triple, and argument pairs from other related PAS. System 6 augments System 3 with word level information from the coordinating phrases in the feature vector. This system does not use PAS information at all.

Table 4 tabulates the Precision, Recall and *F*-score for finding the "true" coordinating construction, and the classification accuracy for the cross-validated training set. Table 5 reports the same for the independent test set. Precision, Recall, and *F*-score are computed as follows:

$$P = \frac{\sum(TP)}{\sum(TP + FP)}$$
$$R = \frac{\sum(TP)}{\sum(TP + FN)}$$
$$F = \frac{2PR}{P + R}$$

where *P* represents precision, *R* represents recall, *TP* is a true positive, *TN* is true negative, and *FP* is false negative. The classification accuracy is the percentage of coordinating constructions, correctly labeled as positive or negative.

From the cross-validated results on the training set, a baseline *F*-score of 0.68 can be achieved using the conjunction alone as a feature. The lexical information from the coordinating head words raise performance to *F*-score of 0.80. In System 4, the related PASs are used as triples, and in System 5, the arguments of the related PAS are used as word pairs, used in addition to the triples. System 5 gave the best performance with *F*-score of 0.84. In comparison, using additional lexical information from the coordinating constructions themselves do not enhance performance in the experiments.

Table 5
Experimental results evaluated on independent test set. Systems 1, 2, 3 and 5 as described in the text were compared.

Features	*P*	*R*	*F*-score	Accuracy %
[S1] conjunction	0.76	0.59	0.66	85
[S2] S1 + coordination type	0.78	0.58	0.67	86
[S3] S2 + coordinating heads	0.86	0.64	0.74	89
[S5] S3 + related PAS with triples and arg pairs	0.89	0.65	0.76	90

Similar trends are exhibited by classifier performance on the independent test set with a best *F*-score of 0.76 when the PAS information is used.

In the test set, every coordinating construction in a sentence is labeled by the classifier independent of one another, even though in a phrase consisting of more than two listed elements, the coordinating phrases will be linked with two or more conjunctions, and one coordinating phrase will subsume the other in the hierarchical parse structure.

A post-processing algorithm that picks the best coordinating construction per sentence is implemented. A simple criterion is used where for every sentence, if there are multiple coordinated phrases found, the longest candidate phrase that has been positively labeled is selected. The algorithm is processed on the test set hypotheses of System 5. This is evaluated against a set of reference answers where only the longest coordinating construction is labeled as positive and the other coordinating phrases subsumed are labeled negative. This yields new results values: $Precision = 0.83$, $Recall = 0.73$ and $F\text{-}score = 0.78$ for the extraction of relevant coordinating constructions, and overall labeling accuracy of 93%.

7.2. Discussion and error analysis

The proposed technique of extracting the entire coordinating phrase exploits added linguistic context and not only detects the pharmaceutical substance involved but also the exact wording which explicates the actual nature of comparison. In the following example:

> "...patients... were randomized to receive olmesartan medoxomil monotherapy (40 mg once daily, $n = 302$) or olmesartan medoxomil (20 mg once daily)/hydrochlorothiazide (12.5 mg once daily) combination therapy..."

the system extracted "olmesartan medoxomil monotherapy or olmesartan medoxomil/hydrochlorothiazide combination therapy" from which we can infer the intervention arm from each side of the coordinating construction as "olmesartan medoxomil monotherapy" and "olmesartan medoxomil/hydrochlorothiazide combination therapy."

From Table 3, coordinating constructions are very common for "Method" sentences. In the training set, 94% (190 of 203) of the abstracts contain representations of the intervention arms as coordinating conjunctions. The number of correct coordinating constructions found from parse trees is counted. There are 177 (17%) relevant constructions in the training data set of 1022. Thus the data set is quite unbalanced. These 177 constructs come from 148 abstracts, 73% of the entire training set of abstracts. Thus at best the intervention arms could be elicited from 73% of the abstracts in the data set, and ideally this would improve to 94% if the parser can extract coordinating constructions perfectly.

The linguistic information afforded by the predicate-argument structures contributed to substantial performance improvements on the extraction of relevant coordinating constructions. The most common PASs connected with the relevant coordinating phrase are: *receive(patient, X), either(X), to(randomize, X), to(assign, X), administer(null, X)* where X represents the left head coordinating conjunct of the relevant phrase. These capture the verb relations in which the conjuncts occur.

The authors of Enju report 88.0% precision and 87.2% recall on predicate-argument structures on GENIA data and also report particular errors related with coordinations in [57]. Here the Enju parser is used "as is", trained on GENIA but it is likely that there is substantial difference between the GENIA training corpus and the corpus of RCT abstracts. The parse errors and classification errors are explored in greater detail below.

7.2.1. Parse errors

In addition to missed coordinated phrases, the parser also made many bracketing and segmentation errors arising from ambiguity due to the full parsing of constituents such as prepositional phrase attachments. In these experiments, the classification is considered as correct when the coordinating conjunction is the correct one even when there is a boundary identification error for the start and end of the left and right coordinate. Errors in segment identification of the coordinating phrase would ultimately inhibit the ability to extract each intervention allocated to each arm of the trial. Examples of typical errors are:

(1) Bracketing/PP Attachment and Segmentation Errors. These are common problems for parsers with handling coordinating conjunctions where the boundaries of the conjuncts are very hard to define, particularly when other constituents nearby attach to either one of the conjuncts or to another phrase in the vicinity. In the example below:

> "Subjects were initially randomized to either metformin once daily or troglitazone once daily."

the detected coordinating phrase turns out to be "once daily or troglitzone." For a more complex example:

> "All enrolled patients... were randomized to receive an intravenous bolus of 0.2 mg/kg of ketamine, followed by a 2-h ketamine infusion at 0.5 mg/kg per hour or an equal-volume regimen with normal-saline placebo."

the parse yields the coordinating phrase "hour or an equal-value regimen" due to the confounding elements presented by each of the prepositional phrase attachments.

(2) Novel conjunction expressions. Coordinating phrases that are expressed as: "placebo/40 mg of atorvastatin" are not detected by the parser. A similar phenomenon occurs for the example: "1 mg terbutaline (or placebo)"

(3) Ellipsis. The deletion of the auxiliary verb "were" seems to have engendered a parsing error for:

> "1532 were assigned to valsartan and 1502 assigned to placebo."

where the erroneous coordinating construction "valsartan and 1502" is produced.

7.2.2. Classification errors

Sources of classification errors are:

(1) Boundary errors from the parser confounded the classifier. In the instance below, the phrase is labeled as the "true" construction for intervention but the head constituents for the coordination are "allocate" and "combine".

> "allocated to three treatment arms consisting of losartan..., ramipril... and combined... for 24 weeks."

The true phrase is "losartan, ramipril and combined'.'

(2) The semantic mapping failed to normalize some drug treatments. For instance, in "SR or RR" and "HFA-BDP and CFC-BDP", the abbreviated references to treatment names were not mapped to the correct semantic type.

(3) Other ambiguous statements generate errors in the classifier. In the example below:

> "Adults ($N = 150$) with perennial AR received FP-SAL and placebo nasal spray during the run-in period."

The cue contextual phrase of "run-in period" indicates that "FP-SAL and placebo" are not part of the intervention being compared and studied in the trial.

(4) Lack of training data for adjectival and prepositional constructs has led to errors for constructs such as: "with or without". More contextual information would be useful particularly for these where the informative content lies outside of the coordinating phrase. Similarly false negatives occur with instances of sentence coordination.

8. Limitations

In this study examining the description of interventions in abstracts of clinical studies, the primary focus is on coordinating constructions in single sentences that explicate the allocation of treatment arms. These, as seen above, are prevalent in the texts that were studied. Other constructs are also present to indicate the interventions used, in that, the allocation of treatment arms may not be encapsulated in coordinating constructions, and could occur in multiple sentences. A robust system for extracting interventions among other parameters of clinical study reports could account for all the potentially informative content.

This study is limited to the use of pharmaceutical interventions. Non-pharmaceutical interventions tend to vary widely in the nature of their descriptions, because these can concern educational, financial and organizational provider or patient-oriented practices and procedures. Therefore the style and detail in which the interventions are presented in the abstracts vary significantly. Their extraction should be addressed in another study.

To build more robust models for information extraction, a larger corpus is likely to yield improved results. To the knowledge of the author, this is a first attempt to extract intervention arms, and no other annotated corpora exists in this area. Hence, the experimental results are limited by the small size of the manually annotated corpus. Furthermore the annotation was conducted by a single individual (the author) and not corrected by a second person, thus could be subject to biasing.

9. Conclusions and future work

This paper has described some preliminary work in developing information extraction techniques for locating experimental details of RCTs. It has been shown here that coordinating constructions are used frequently in statements that describe the comparison, assignment or allocation of two or more pharmaceutical interventions. This syntactic construction can be identified from a full sentence parser. In the experiments presented above, a system for automatically extracting and labeling coordinating constructions that describe intervention arms has been implemented and evaluated. The system has demonstrated promising performance considering the small training set used in this first study.

One future aspect of this work to focus on would be improving the parsing and detection of the boundaries coordinating constructions. This appears to be the major source of error.

The unfolding of the coordinating construction is as yet beyond the scope of this work. This represents the final stage of the system that identifies precisely each intervention arm of the RCT. It will require the correct identification of the boundaries of each coordinating conjunct, and also the recognition of key elements that have been deleted due to elliptical constructions. It is seen that ellipses are common. The resolving of the complex intertwining of multiple juxtaposed and nested coordinating constructions will be a challenging task for future work.

Acknowledgments

This project is funded by the Australian Research Council, Grant No. DP0666600. The author wishes to thank Enrico Coiera for valuable feedback on this research project, Sarah Dennis for her help in selecting RCTs with pharmaceutical interventions.

References

[1] ACP Journal Club. Available from: http://www.acpjp.org.
[2] Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. Ann Intern Med 1987;106:595–604.
[3] Agarwal R, Boggess L. A simple but useful approach to conjunct identification. In: Proceedings, 30th annual meeting of the association for computational linguistics, Newark, DE; 1992. p. 15–21.
[4] Altman DG. Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. BMJ 2005;330(7499):1056–7.
[5] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: Annual symposium of AMIA; 2001. p. 17–21.
[6] Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: can we do better? Ann Surg 2006;244(5):663–7.
[7] Beeferman D, Berger A, Lafferty J. Statistical models for text segmentation. Mach Learn 1999;34(1–3):177–210.
[8] Clinical evidence. BMJ Publishing Group. Available from: http://www.clinicalevidence.com
[9] Boots I, Van Dalen E, Kuin L, Offringa M, Caron H, Kremer L. Methodological quality of randomised controlled trials in paediatric oncology published in 2001 and 2002 [abstract]. In: 12th Cochrane colloquium: bridging the gaps, Ottawa, Canada; 2004.
[10] Buyko E, Tomanek K, Hahn U. Resolution of coordination ellipses in biological named entities using conditional random fields. In: PACLING 2007 proceedings of the 10th conference of the Pacific association for computational linguistics, Melbourne, Australia, September 19–21, 2007. Melbourne: Pacific Association for Computational Linguistics; 2007. p. 163–6.
[11] Chambliss ML, Conley J. Answering clinical questions. J Fam Pract 1996;43:140–4.
[12] Chen S, Rosenfeld RA. Gaussian Prio for smoothing maximum entropy models, Technical report CMU-CS-99-108, Carnegie Melon University; 1999.
[13] Chung G, Coiera E. A study of structured clinical abstracts and the semantic classification of sentences. In: Proceedings of BioNLP'07, June, Prague, Czech Republic; 2007.
[14] Cimino JJ, Li J. Sharing infobuttons to resolve clinicians' information needs. In: AMIA annual symposium proceedings; 2003. p. 815.
[15] The Cochrane Collaboration. Available from: http://www.cochrane.org.
[16] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.
[17] Coiera E, Walther M, Nguyen K, Lovell NH. An architecture for knowledge-based and federated search of online clinical evidence. J Med Internet Res 2005;7(5).
[18] Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985;103:596–9.
[19] Dahl V, McCord MC. Treating coordination in logic grammars. Am J Comput Linguist 1983;9:69–91.
[20] D'Alessandro DM, Kreiter CD, Peterson MW. An evaluation of information seeking behaviors of general pediatricians. Pediatrics 2004;113:64–9.
[21] Daraselia N, Yuryev A, Egorov S, Novichkova S, Nikitin A, Mazo T. Extracting human protein interactions from MEDLINE using a full-sentence parser. Bioinformatics 2004;20(5):604–11.
[22] Dawes M, Pluye P, Shea L, Grad R, Greenberg A, Nie JY. The identification of clinically important elements within medical journal abstracts: Patient–Population-Problem, Exposure-Intervention, Comparison, Outcome, Duration and Results (PECODR). Inform Prim Care 2007;15(1):9–16.
[23] Demner-Fushman D, Lin J. Answering clinical questions with knowledge-based and statistical techniques. In: Computational linguistics, vol. 33, no. 1. MIT Press; 2006. p. 63–103.
[24] Ebell MH, Barry HC. InfoRetriever: rapid access to evidence-based information on a handheld computer. MD Comput 1998;15:289–97.
[25] Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermerr JJ. Obstacles to answering doctors questions about patient care with evidence: qualitative study. BMJ 2002;324:710–3.
[26] Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217.
[27] Evidence Based Medicine. Available from: http://ebm.bmjjournals.com.
[28] Fiszman M, Demner-Fushman D, Lang FM, Goetz P, Rindflesch T. Interpreting comparative constructions in biomedical text. In: Proceedings of the BioNLP workshop, association for computational linguistics; 2007. p. 137–44.
[29] Fozi K, Teng CL, Krishnan R, Shajahan Y. A study of clinical questions in primary care. Med J Malaysia 2000;55:486–92.
[30] Geissbuhler A, Miller RA. Clinical application of the UMLS in a computerized order entry and decision-support system. In: AMIA annual symposium proceedings; 1998. p. 320–4.
[31] Goldberg M. An unsupervised model for statistically determining coordinate phrase attachment. In: Proceedings of ACL'99, 37th annual meeting of the association for computational linguistics, College Park, Maryland; 1999. p. 610–61.

[32] Haspelmath M. Coordination. In: Shopen T, editor. Language typology and syntactic description, vol. II: complex constructions. 2nd ed. Cambridge: Cambridge University Press; 2007. p. 1–51.

[33] Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Ann Intern Med 1990;113(1):69–76.

[34] Hersh W, Hickham DB, Haynes RB, McKibbon KA. Evaluation of SAPHIRE: an automated approach to indexing and retrieving medical literature. In: Proceedings of the annual symposium on computer applications in medical care; 1991. p. 808–12.

[35] Hirschman L. Conjunction in meta-restriction grammar. J Logic Program 1986;3:299–328.

[36] Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. In: Annual symposium of AMIA; 2006. p. 359–63.

[37] Hunter L, Bretonnel CK. Biomedical language processing: what's beyond PubMed? Mol Cell 2006;21:589–94.

[38] Ide NC, Loane RF, Demner-Fushman D. Essie: a concept based search engine for structured biomedical text. J Am Med Inform Assoc 2007;14(3):253–63.

[39] Keech A, Val J. Gebski and Rhana Pike Interpreting and Reporting Clinical Trials. A guide to the consort statement and the principles of randomised controlled trials. Australasian Medical Publishing; 2007.

[40] Lindberg DA, Humphreys BL, Mc Cray AT. The unified medical language system. Methods Inf Med 1993;32(4):281–91.

[41] McCallum A. MALLET: a machine learning for language toolkit; 2002. Available from: http://mallet.cs.umass.edu.

[42] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357(9263):1191–4.

[43] Nakov P, Hearst M. Using the web as an implicit training set: application to structural ambiguity resolution. In: Proceedings of human language technology conference and conference on empirical methods in natural language processing (HLT/EMNLP), Vancouver, Canada; 2005. p. 835–42.

[44] Niu Y, Hirst G. Analysis of semantic classes in medical text for question answering. In: Workshop on question answering in restricted domains, Barcelona; 2005.

[45] Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. JAMA 1993;270(17):2093–5.

[46] Pyysalo S et al., Analysis of link grammar on biomedical dependency corpus targeted at protein–protein interactions. In: Proceedings of international joint workshop on natural language processing in biomedicine and its applications (JNLPBA); 2004; p. 15–21.

[47] Ratnaparkhi A. A maximum entropy model for part-of-speech tagging. In: Proceedings of the empirical methods in natural language processing; 1996.

[48] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123(3):A12–3.

[49] Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone; 2000.

[50] Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16.

[51] Sutton C, McCallum A. An introduction to conditional random fields for relational learning. In: Getoor L, Taskar B, editors. Introduction to statistical relational learning. 2006: MIT Press; 2006.

[52] Tomanek K, Wermter J, Hahn U. A reappraisal of sentence and token splitting for life science documents. In: MEDINFO 2007—proceedings of the 12th world congress on medical informatics; 2007.

[53] Tsay MY, Ma YY. Bibliometric analysis of the literature of randomized controlled trials. J Med Libr Assoc 2005;93(4):450–8.

[54] Tsujii Laboratory. Enju—a practical HPSG parser; 2005. Available from: http://www-tsujii.is.s.u-tokyo.ac.jp/enju/.

[55] Tsujii Laboratory. GENIA project; 2005. Available from: http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/.

[56] Tsuruoka Y, Tateishi Y, Kim J-D, Ohta T, McNaught J, Ananiadou S, et al. Developing a robust part-of-speech tagger for biomedical text. In: Advances in informatics—10th Panhellenic conference on informatics, LNCS 3746; 2005. p. 382–92.

[57] Yakushiji A, Miyao Y, Ohta T, Tateisi J, Tsujii Y. Automatic construction of predicate-argument structure patterns for biomedical information extraction. In: Proceedings of the 2006 conference on empirical methods on natural language processing; 2006.