An XML based framework for merging incomplete and inconsistent statistical information from clinical trials

Jianbing Ma[1], Weiru Liu[1], Anthony Hunter[2], and Weiya Zhang[3]

[1] Computer Science, Queen's University Belfast, Belfast, Co Antrim BT7 1NN, UK
[2] Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK
[3] Academic Rheumatology, Medical & Surgical Sciences,
City Hospital, Nottingham, NG5 1PB, UK

Abstract. Meta-analysis is a vital task for systematically summarizing statistical results from clinical trials that are carried out to compare the effect of one medication (or other treatment) against another. Currently, most meta-analysis activities are done by manually pooling data. This is a very time consuming and expensive task. An automated or even semi-automated tool that can support some of the processes underlying meta-analysis is greatly needed. Furthermore, statistical results from clinical trials are usually represented as sampling distributions (i.e., with the mean value and the SEM). When collecting statistical information from reports on clinical trials, not all reports contain full statistical information (i.e., some do not provide SEMs) whilst traditional meta-analysis excludes trials reports that contain incomplete information, which inevitably ignores many trials that could be valuable. Furthermore, some trials results can be significantly inconsistent with the rest of trials that address the same problem. Therefore, highlighting (resp. removing) such inconsistencies is also very important to reveal (resp. reduce) any potential flaws in some of the trials results. In this paper, we aim to design and develop a framework that tackles the above three issues. We first present an XML-based merging framework that aims to merge statistical information automatically with the potential to add a component to extract clinical trials information automatically. This framework shall consider any valid clinical trial including trials with partial information. We then develop a method to analyze inconsistencies among a collection of clinical trials and if necessary to exclude any trials that are deemed to be illegible. Finally, we use two sets of clinical trials, trials on Type 2 diabetes and on neurocognitive outcomes after off-pump versus on-pump coronary revascularisation, to illustrate our framework.

keywords XML, Meta-analysis, Merging, Incomplete information, Sampling distribution, Semantic heterogeneity, Information Extraction

1 Introduction

Clinical trials are widely used to test new drugs or to compare the effect of different drugs. A clinical trial is a study that compares the effect of one medication (or other treatment) against another [16]. Trial results are a summary of the underlying statistical analysis. A huge number of clinical trials have been carried out in the last few

decades and new trials are constantly being designed and implemented. For example, many clinical trials have been carried out to investigate issues including: the intraocular pressure (IOP) lowering efficacy of drugs, such as travoprost, bimatoprost, timolol, and latanoprost, (e.g., [4, 7, 15, 21, 30, 32, 34, 37, 40]); oral medications for adults with Type-2 diabetes (e.g., [3, 9, 27, 29, 35, 41]); the neurocognitive outcomes after off-pump versus on-pump coronary revascularisation (e.g., [14, 25, 26, 31, 43]).

Given the huge number of clinical trials and the fact that clinical trials reports are time consuming to read and understand, systematic reviews of related trials is needed by medical practioners and other health care professionals to assess drugs/therapies of interest. Meta-analysis is the technique commonly used in clinical trial research to summarize related trial results, that is, to merge multiple sampling distributions into a single distribution.

Meta-analysis is a very important step in the development of evidence-based medicine, and there are various tools supporting this task, such as SAS, STATA, MetaWin, WEasyMa, etc. However, there are still difficulties carrying out meta-analysis with these tools when a large number of clinical trials need to regularly be considered and when new trials are being completed. First, current meta-analysis technique requires input data to be extracted from clinical trial manually. This is a very time-consuming task particularly when the number of related reports is very large. Second, before inputting data into meta-analysis tools, it is necessary to systematically preprocess the semantic heterogeneity of data. This includes, for example, manually checking whether the data is about the same issue, whether the data uses the same unit of measurement and if not some conversion needs to be done, and whether these clinical trials are of the same duration, etc. So there are a number of low-level but important steps of standardizing the format and checking correspondences. Therefore, some kind of automated process that can extract information from clinical trials reports and can verify to some degree that whether some trials are eligible together for meta-analysis would be very useful.

In a clinical trial, patients are divided into treatment groups, with each group receiving one of the drugs under study. Specific outcomes are measured and the differences between the measurements at the start of the trial and at the end are compared for each group. By convention, clinical trials results are described using sampling distributions. When the full details about sampling distributions are available, merging the results from these trials entails systematic use of established techniques from statistics, as done in the current meta-analyses. However, in reality, some trials reported in the literature are statistically incomplete, for instance, the standard error of mean (SEM) can be missing from a sampling distribution. Traditionally, it is difficult to make use of those clinical trials in meta-analysis. In fact, a clinical trial with incomplete information is often abandoned. However, in [28], a prognostic method and an interval method are proposed to deal with meta-analysis with incomplete information. Obviously, these two methods are useful alternatives to the traditional meta-analysis.

When a set of clinical trials on the same issue are collected for meta-analysis, there might be some clinical trials presenting highly conflict statistical results with results from other clinical trials. For these inconsistent trials results, it is very likely that these trials are done on different populations, and hence should be excluded to achieve a better meta-analysis result.

As the popularity of XML in dynamic data exchange increases, a variety of tools to store and retrieve data in/from XML documents have been developed. Since a clinical trial result may be used on different occasions and in different meta-analysis, storing main statistical results of clinical trials in XML documents is an appealing idea.

In this paper, we present an XML based framework for supporting meta-analysis by defining merging rules for combining complete and incomplete clinical trials data, with a longer-term objective to completely automate this process, e.g., to extract clinical trials information and pre-process the semantic heterogeneity automatically.

More specifically, this paper contains the following contributions.

1. We present a general XML based merging framework that extends the fusion rule technique developed in [17] especially for clinical trials data.

2. We show how our framework can deal with clinical trials with incomplete information where current meta-analysis tools cannot.

3. We show how our framework can analyze inconsistent information and remove highly conflict information by excluding a trial with this nature.

4. We provide a brief discussion on semantic heterogeneity in statistical information merging and on automated information extraction highlights other two important aspects that we will develop in order to realize an automated meta-analysis tool.

5. We illustrate our framework with two case studies (Type-2 diabetes and neurocognitive outcomes after off-pump versus on-pump coronary revascularisation) showing the whole process and its efficacy.

The remainder of this paper is organized as follows. In Section 2, we give a brief introduction to XML, define the XML document structure for representing the information contained in clinical trials reports, and discuss the automatic information extraction and semantic heterogeneity processing. In Section 3, we formally describe the XML-based merging framework including basic definitions and clinical trials oriented restrictions of tags. Section 4 discusses how to manage the possibly incomplete and inconsistent information contained in XML documents to perform a meta-analysis. Section 5 provides two case studies, one is on Type 2 diabetes and the other is on neurocognitive outcomes after off-pump versus on-pump coronary revascularisation. We use these studies to illustrate our framework. Finally, in Section 6, we conclude the paper. In addition, we put the full DTD description of the XML document structure in the Appendix.

2 XML Document

In this section, we introduce some basic concepts of XML as well as the XML document structure we will use in this paper. We will also discuss issues related to semantic heterogeneity and information extraction from clinical trials.

2.1 Introduction to XML

Extensible Markup Language (XML) has become an important part of Semantic Web, due to its simple and flexible format. An XML document is constructed based on a DTD or an XML Schema that specifies how tags in an XML should be arranged. Initially XML was mainly used to store and exchange static data, such as, metadata standards

by Dublin core, but XML is now playing an increasingly important role in the exchange of a wide variety of dynamic data too, data that are retrieved or obtained upon requests. Typical examples of this kind are [11], [39], and [45], where the former constructs an XML document from a collection of multimedia data about a patient and the latter two generate XML documents that store probabilistic query results and predictive models obtained from data mining or intelligent analysis tools respectively.

To facilitate the modelling of various types of data in XML, the need to represent *uncertain* data has emerged too, as in the case happened to traditional databases where numerous approaches were proposed to create and manipulate probabilistic databases (e.g. [2, 12]). Because XML documents are structured, uncertain information associated with data must be naturally assigned, interpreted and structured. Uncertainty can occur at different levels of granularity and uncertainty can be interpreted in different ways, such as in terms of probabilities, probability intervals, reliabilities, or beliefs. Furthermore, an integration result of XML documents having data values with certainty may create an XML document with uncertain data. Therefore, managing uncertain data in XML raises many challenging issues.

2.2 XML document structure

XML based frameworks for representing and managing uncertain and incomplete information were proposed in many papers, e.g., [1, 17–19, 23, 33], etc. In [17], a general XML based framework was proposed to merge XML documents with uncertain information like probabilities, possibilities, and belief functions. In [18], the proposed XML based framework was focused on merging uncertain information that is defined at different levels of granularity of XML textentries. In [19], the framework paid special interests to deal with reliabilities in different XML documents. In [17–19], structured reports with uncertain weather information were studied. The following two reports are examples.

```
⟨report⟩                                    ⟨report⟩
 ⟨source⟩ TV1 ⟨/source⟩                      ⟨source⟩ TV3 ⟨/source⟩
 ...                                        ...
 ⟨temperature⟩                              ⟨temperature⟩
  ⟨probability⟩                              ⟨probability⟩
   ⟨prob value = "0.2"⟩8°C⟨/prob⟩             ⟨prob value = "0.4"⟩8°C⟨/prob⟩
   ⟨prob value = "0.8"⟩12°C⟨/prob⟩            ⟨prob value = "0.6"⟩12°C⟨/prob⟩
  ⟨/probability⟩                             ⟨/probability⟩
 ⟨/temperature⟩                             ⟨/temperature⟩
⟨/report⟩                                   ⟨/report⟩
```

However, to our knowledge, there are no papers focusing on representing and managing possibly incomplete and inconsistent statistical information from clinical trials in XML frameworks. The needs of representing and combining clinical knowledge raised some important and interesting techniques issues. In this paper, we extend the ideas of [17–19] to create an XML based framework to deal with such information. We investigated many clinical trials reports in order to ensure our XML structure would cover a

wide range of examples. That is, to accommodate our special needs of recording clinical trials information, the DTD of the XML documents should be adapted. The full XML document structure (DTD adaption) are given in the Appendix. Here we only give the DTD adaption of the Result element which contains information about clinical trials results (Fig. 1).

Most of the time, clinical trials results are reported in the form of sampling distribu-

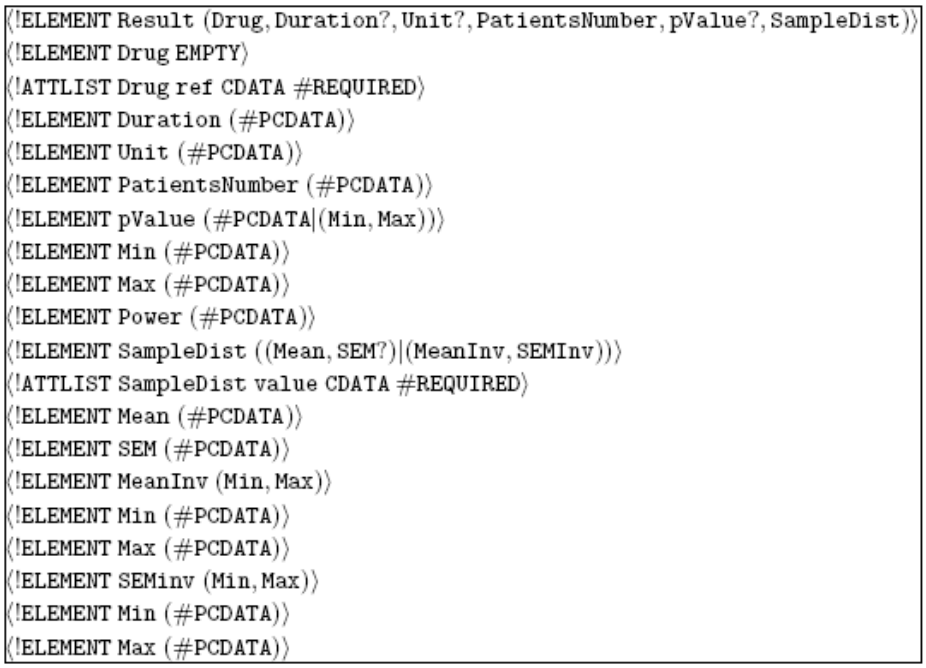

Fig. 1. DTD adaption of the Result element

tions, and sometimes they are given in the form of confidence intervals. If so, we will transform confidence intervals to sampling distributions before putting the data into the XML documents (the transformation process will be introduced in Section 4.2). Sampling distributions are represented in the form of intervals, i.e., MeanInv and SEMInv, only when we use the interval method that will be introduced in Section 4.4 in dealing with trials with incomplete information. The value attribute of the SampleDist element indicates the target of a trial result such as "level of LDL cholesterol", etc. In addition, the Unit element is optional because in some cases, there is no unit child in the Result element, e.g., an odds ratio does not have a unit of measurement.

Example 1 *The following is a Result element.*
⟨Result⟩
⟨Drug ref = "l1" /⟩
⟨Duration⟩*3 month*⟨/Duration⟩
⟨Unit⟩*mmol/L*⟨/Unit⟩

```
    ⟨PatientsNum⟩247⟨/PatientsNum⟩
    ⟨SampleDist value = "Intraocular Pressure Reduction"⟩
        ⟨Mean⟩4.1⟨/Mean⟩
        ⟨SEM⟩3.8⟨/SEM⟩
    ⟨/SampleDist⟩
⟨/Result⟩
```

2.3 Extracting clinical trials information to build XML documents

A vital aspect of building up a large collection of XML documents containing clinical trials information is to use an existing *information extraction* tool to extract relevant information.

Information extraction (IE) technology (or synonymously text mining technology) aims to "read" text and pick out the bits of information that are needed. IE systems tend to be developed for focused applications where there is some regularity in the information being presented in the text. For example, in papers on clinical trials, there are some regularities in the information being presented, such a paper is likely to include the patient class of the trial, treatment classes to which the patients were assigned, and the comparative outcomes of treatments. Hence, with an information extraction system for an application, there is the idea of a template that specifies the information that is sought by the system.

A number of viable information extraction systems have been developed [8]. For example, the GATE System provides an implemented architecture for managing textual data storage and exchange, visualization of textual data structures, and plug-in modularity of text processing components [10]. The text processing components includes LaSIE which performs information extraction tasks including named entity recognition, coreference resolution, template element filling, and scenario template filling. Furthermore, a number of natural-language parsers have been developed that can be incorporated in information extraction systems (for a comparison for biological applications see [13]).

Since our main task of the paper is not information extraction, rather it is how to represent and merge such extracted information, below we focus on what information we need to extract from clinical trials.

In our study of clinical trials and in consultation with clinicians, we need to extract the following information from its report, in order to efficiently make use of each clinical trial,

1. The outcomes being measured and compared, including the name of the outcome and its unit.
2. Trial duration including the total length of the study, and any intermediate period intervals, e.g., a 12-month trial report may also provide results of 3 months, 6 months, etc.
3. For each trial group:
   (a) Drug(s) used in that group.
   (b) Number of patients in that group.
   (c) The outcome measurements made for that group, namely:

i. The mean and standard error of the mean at baseline and at each endpoint specified by the testing schedule, or alternatively, the difference of the two.
ii. The p-value, if given.
iii. The confidence interval (CI), if given.

Certainly, there are other items of information that are valuable and useful as well, such as the main conclusion of a trial (e.g., Drug A is more effective than Drug B, or Drug A has severe side effects on patients with condition C etc). In our current merging framework, we have not considered these types of information yet. So although our XML documents will contain such types of information, now we mainly concentrate on statistical information provided in clinical trials and any additional information that is needed when using such statistical information.

2.4 Heterogeneous information management through ontologies

As clinical trials reports come from different sources, semantic heterogeneity occurs frequently. For example, some reports use phrase "Low density lipoprotein cholesterol" while some other reports prefer it by the abbreviation "LDL-cholesterol"; some reports refer to "NF-kappa B" while others may write "p50/p60" as an equivalent term. Not only are different words used for the same meaning, but different reports may also use different units of measurement which are interchangeable. For instance, with regard to a trial duration, 1 year is equivalent to 12 months, 12 weeks is approximately equivalent to 3 months, etc. As another example, LDL cholesterol measurement in diabetes research has two different measurement units mmol/L and mg/L, and so clinicians interested in those reports must manually translate x mmol/l into y mg/L using formula $y = x * 39$, or vice versa.

Therefore, with knowledge and information fusion, semantic heterogeneity becomes a complex and multi-faced topic, and it is central to the merging approach we are discussing here. From the perspective of merging, we consider information to be merged in context. This means we undertake logical reasoning with the information to be merged to determine what it means. For example, in merging two reports on drug trials, we want to use any available information in the reports and background knowledge (e.g., NF-kappa B is equivalent to p50/p60) about the underlying assumptions in the experiments, the stages of the disease, etc, to determine whether merging is appropriate, and if so what kind of aggregation should be used on the constituent parts of this information. For determining whether information in two or more reports are referring to the same issue, we are investigating the use of ontological knowledge, e.g., [38, 42], to assist the selection of clinical trials for possible merging.

The notion of ontology has had a long history in science. Once an ontology incorporates a large number of concepts and relationships, it gives us the ability to standardize the terminology, thereby minimizing ambiguities and facilitating communication. This is particularly important in a distributed environment where one may have numerous users who need to feel confident about the terms and concepts being used. Recourse to an ontology can ameliorate the complexity inherent in content in many applications by providing a common framework for structuring the content. In terms of clinical trials, we need to have an ontology to describe relevant concepts and their relationships used

in each category of clinical trials, such as trials on diabetes etc. Such an ontology for example shall contain information about translation of words with the same meaning in the context, conversion of one measure into another when different trials use different measures, etc. In fact, there are already some known ontologies related to biomedical knowledge, e.g., SNOMED CT, Gene Ontology, etc. SNOMED CT is a clinical terminology (the Systematised Nomenclature of Medicine Clinical Terms) that provides a very large and wide-ranging common computerised language that can be used by all applications in a healthcare system to facilitate communications between healthcare professionals in clear and unambiguous terms. Further important ontological resources for medical science include the Unified Medical Language System, and the framework for sharing of ontologies offered by The OpenBiomedical Ontologies Foundary.

One of our next step research is to build an ontology for clinical trials of selected application domains. This will be done based on SNOMED CT and other related, publicly available ontologies. We will use Protégé, [36], a free, open source ontology editor developed by Stanford University to complete this task.

3 XML-Based Merging Frameworks

In this section, we introduce an XML-based merging framework. The framework follows the idea of merging uncertain information in structured reports in [17]. First, we present a general definition of the XML based merging framework, for which we define a selection function to select a set of "compatible" trials (in terms of XML documents) and a merging rule to combine the selected trials to get a new XML document. Furthermore, we impose some clinical trials oriented constraints on the general framework.

3.1 Basics of XML-based merging framework

We use XML to represent clinical trials reports. For convenience, we will call them XML reports from now on.

Following [17], we define an XML report formally as follows.

Definition 1 *(XML report) If ψ is a tagname (i.e., an element name), and ϕ is textentry, then $\langle\psi\rangle\phi\langle/\psi\rangle$ is an XML report. If ψ is a tagname, ϕ is textentry, θ is an attribute name and κ is an attribute value, then $\langle\psi\ \theta = \kappa\rangle\phi\langle/\psi\rangle$ is an XML report. If ψ is a tagname, and $\phi_1, \ldots, \phi_n$ are XML reports, then $\langle\psi\rangle\phi_1 \ldots \phi_n\langle/\psi\rangle$ is an XML report.*

This definition for an XML report is very general (similar to Def. 1 in [17] for structured news report). In practice, we would use DTD defined in the Appendix to adapt this definition. For example, we may restrict the root element of an XML report to be a Trial element. Furthermore, if there is a DTD element as $\langle$!ELEMENT A (S)$\rangle$ where A is an element name and S is a set of children names, then for an element named A in the XML report, if B is a child element of A, we restrict that $B \in S$. Since these kinds of application oriented adaptions are not the main topic of this paper and in fact are fully implied in the DTD definitions, here we will not consider these issues further. However, in this paper, we will impose some constraints on XML reports in Section 3.2, to support the handling of uncertainty.

For convenience, hereafter we use $\mathcal{L}$ to denote the set of all XML reports.

To define a general merging framework, we first define a mergeable relation.

Definition 2 *A mergeable relation* R *is a reflexive, symmetrical and transitive relation on* $\mathcal{L} \times \mathcal{L}$.

This definition for a mergeable relation is also very general. In real applications, specific criteria should be introduced to instantiate the relation. In following sections, clinical trials oriented criteria will be given to adapt R.

With a mergeable relation R, if $\alpha_1, \alpha_2 \in \mathcal{L}$ are two XML reports, then α_1 and α_2 are said mergeable iff we have $\mathsf{R}(\alpha_1, \alpha_2)$.

Before performing the merging of XML reports, the XML based merging framework should first select the mergeable XML reports.

Definition 3 *(Selection function) A selection function* S *is a mapping from a set of XML reports to its mergeable subset such that if* A *is a set of XML reports, then* $S(A) \subseteq A$ *and* $\forall \alpha_1, \alpha_2 \in S(A)$*, we have* $\mathsf{R}(\alpha_1, \alpha_2)$.

This definition for a selection function will be instantiated when the mergeable relation R is practically adapted.

Once we have a set of mergeable XML reports, we need to combine them into a new XML document.

Definition 4 *A merging rule is a total function* M *associating a set of mergeable XML reports to an XML report such that if* $\alpha_1, .., \alpha_n \in \mathcal{L}$ *and* $\forall 1 \leq i, j \leq n, \mathsf{R}(\alpha_i, \alpha_j)$*, then* $M(\alpha_1, \ldots, \alpha_n) \in \mathcal{L}$.

Generally, a set of mergeable XML reports $\alpha_1, .., \alpha_n$ in Def. 4 are always from the result of a selection function S.

To summarize, an XML based merging framework is a pair (S, M) that applies to sets of XML reports where S is a selection function and M is a merging rule.

3.2 Representing statistical information in XML frameworks

In this section, we want to introduce some constraints on clinical trials. These constraints are focused on representing and managing statistical information in clinical trials reports.

Definition 5 *The set of* **key statistical tagnames** *for this paper are* `PatientsNum` *and* `SampleDist`. *The set of* **subsidiary statistical tagnames** *for this paper are* `Mean`, `SEM`, `MeanInv` *and* `SEMInv`. *The set of* **auxiliary statistical tagnames** *are* `Drug`, `Duration` *and* `Unit`.

Now we define the representation of the SampleDist element which contains the most important statistical information.

Definition 6 *The XML report* $\langle \texttt{SampleDist} \rangle \sigma_1 \ldots \sigma_n \langle / \texttt{SampleDist} \rangle$ *is a* **valid** *SampleDist element iff one of the following conditions is satisfied.*

1. $n = 1$ *and* σ_1 *is of the form* $\langle$Mean$\rangle\phi\langle/$Mean$\rangle$ *where* ϕ *is a textentry.*
2. $n = 2$ *and* σ_1 *is of the form* $\langle$Mean$\rangle\phi_1\langle/$Mean$\rangle$, σ_2 *is of the form* $\langle$SEM$\rangle\phi_2\langle/$SEM$\rangle$ *where* ϕ_1, ϕ_2 *are two textentries.*
3. $n = 2$ *and* σ_1 *is of the form* $\langle$MeanInv$\rangle\psi_1\langle/$MeanInv$\rangle$, σ_2 *is of the form* $\langle$SEMInv$\rangle\psi_2\langle/$SEMInv$\rangle$ *where* ψ_1, ψ_2 *are of the form* $\langle$Min$\rangle\phi_1\langle/$Min$\rangle\langle$Max$\rangle\phi_2\langle/$Max$\rangle$ *such that* ϕ_1, ϕ_2 *are two textentries.*

All textentries ϕ_i in the above definition can only be numerical values.

Example 2 *The following is a valid SampleDist element.*
⟨SampleDist value = "intraocular pressure reduction"⟩
⟨Mean⟩*4.1*⟨/Mean⟩
⟨SEM⟩*3.8*⟨/SEM⟩
⟨/SampleDist⟩

Definition 7 *An XML report* $\langle$Result$\rangle\sigma_1 \ldots \sigma_n\langle/$Result$\rangle$*is a* **valid** *Result element iff*

1. σ_i*s are different auxiliary statistical tagnames or key statistical tagnames.*
2. *All key statistical tagnames exist in* σ_i*s in which the SampleDist tag is valid.*

In this paper, the main task is to merge sampling distributions contained in multiple XML documents when they refer to the same issue. Therefore, we define the following constraint for merging two valid Result elements.

Definition 8 *Given two valid Result elements*

⟨Result⟩	⟨Result⟩
⟨Drug ref = id1/⟩	⟨Drug ref = id2/⟩
⟨Duration⟩ x1 ⟨/Duration⟩	⟨Duration⟩ x2 ⟨/Duration⟩
⟨Unit⟩ y1 ⟨/Unit⟩	⟨Unit⟩ y2 ⟨/Unit⟩
⟨PatientsNum⟩ z1 ⟨/PatientsNum⟩	⟨PatientsNum⟩ z2 ⟨/PatientsNum⟩
⟨SampleDist ref = purpose1⟩	⟨SampleDist ref = purpose2⟩
...	...
⟨/SampleDist⟩	⟨/SampleDist⟩
⟨/Result⟩	⟨/Result⟩,

they are said mergeable iff we have

$$id1 = id2, x1 \simeq x2, y1 = y2 \; and \; purpose1 = purpose2.$$

That is, two clinical trials results can be merged iff they refer to the same drug, have approximately the same duration, use the same unit of measurement and for the same clinical purpose. This definition is a clinical specific restriction before using the merging rule in Def. 4. This restriction can be carried out with the assistance of ontologies tailored for such an application as discussed in Section 2.4. In addition, this definition is a clinical trial oriented instantiation of Def. 2, hence we can use it to select mergeable XML reports.

More specific merging rules for statistical information are introduced in the next section.

4 Managing Statistical Information in XML Documents

In this section, we first recall some basic concepts of statistical information and then discuss how to model such information in XML documents. We define an instantiated selection function (in terms of an algorithm) to exclude inconsistent information and provide two instantiated merging rules to deal with meta-analysis with incomplete information.

4.1 Preliminaries

In statistics, a normal distribution associated with a random variable is denoted as $X \sim N(\mu, \sigma^2)$. For the convenience of further calculations in the rest of the paper, we use notation $X \sim N(\mu, \sigma)$ instead of $X \sim N(\mu, \sigma^2)$ for a normal distribution of variable X.

In statistics, random samples of individuals are often used as the representatives of the entire group of individuals (often denoted as a population) to estimate the values of some parameters of the population. The mean of variable X of the samples, when the sample size is reasonably large, follows a normal distribution. This distribution is typically referred to as a sampling distribution.

We use $X \sim N(\mu, SEM)$ to denote a sampling distribution with mean value μ and standard error of mean SEM.

Conventionally, let $X_i \sim N(\mu_i, SEM_i)$, $1 \leq i \leq k$ and $\omega_i = \frac{1}{SEM_i^2}$, then the meta-analysis result $X \sim N(\mu, SEM)$ is as follows.

$$\mu = \frac{\sum_{i=1}^{k} \mu_i * \omega_i}{\sum_{i=1}^{k} \omega_i}, \omega = \sum_{i=1}^{k} \omega_i. \tag{1}$$

4.2 Obtaining Sampling Distributions from Clinical Trials

In this subsection, we show how we get sampling distributions from clinical trials. Clinical trials results are obtained from three different categories.

- Category I: A sampling distribution can be identified when both μ and SEM are given.
- Category II: A sampling distribution can be identified when only μ is given.
- Category III: A sampling distribution can be constructed when a confidence interval is given.

After looking through a large collection of papers of clinical trials on IOP reductions and on comparing drugs for type-2 diabetes, we believe that the above three categories cover a significant proportion of statistical information (e.g., [4, 7, 15, 21, 30, 32, 34, 37, 40], etc).

For each category of statistical information, we interpret it in terms of a sampling distribution and then put the distribution into the corresponding XML document. We use X to denote the sample mean implied in the context of each clinical report.

For the first category, a sampling distribution is explicitly give, for example, $X \sim N(9.3, 2.9)$ gives

⟨SampleDist value = "LDL − C"⟩
⟨Mean⟩9.3⟨/Mean⟩
⟨SEM⟩2.9⟨/SEM⟩
⟨/SampleDist⟩

For the second category, a sampling distribution can be defined with a missing SEM, for instance, $X \sim N(5.9, SEM)$, so we have an XML segment as

⟨SampleDist value = "LDL − C"⟩
⟨Mean⟩5.9⟨/Mean⟩
⟨/SampleDist⟩

For the third category of information, a confidence interval $[a, b]$ is given. It is then possible to convert this confidence interval into a sampling distribution as follows

$$\mu = \frac{a+b}{2}, \qquad SEM = \frac{b-a}{2k}.$$

As a convention, the presented analysis of clinical trials results usually use the 95% confidence interval. In this case, we have $k = 1.96$. However, if a given confidence interval is not the usual 95% confidence interval (say, it uses the p-confidence interval), it is possible to use the standardization of the normal distribution as $P(Z \in [-k, k]) = p$. Then value k can be found by looking up the standard normal distribution table. Therefore, in this case, we get an XML representation in the same way as for Category one.

To summarize, from our investigation, we can get sampling distributions (some with missing SEMs) from all the three types of information.

4.3 Inconsistency analysis among trials

In this subsection, we aim to investigate how to analyze potential inconsistencies among trials. Based on this analysis, we are able to remove some highly conflicting trials from given trials with full statistical information. We only want to identify and remove those trials which may have been conducted from a different population. In fact, clinicians believe that only this type of inconsistency should result in a trial(s) being excluded from a meta-analysis. Since if a trial(s) is from a different population, then it should not be considered together with other trials.

We take the assumption that the same/similar population shall have the same/similar standard deviation. That is, for a given k trials with full statistical information, we want to measure whether each $\sigma_i = SEM_i\sqrt{n_i}$, $(1 \leq i \leq k)$ is marginally equivalent. In another word, these $\sigma_i = SEM_i\sqrt{n_i}$ values shall all be in a reasonably tight interval and this is the principle of our method to identify an inconsistent trial(s). Note that both SEM_i and n_i can be extracted from XML documents (i.e., elements SEM and PatientsNum).

Assume that we have a set of values $\sigma_1, \ldots, \sigma_k$ from k trials. Without loss of generality, we can also assume that the list is already sorted, i.e., $\sigma_1 \leq \ldots \leq \sigma_k$.

First, we find the median md of the list, namely, if k is odd, then md is $\sigma_{\frac{k+1}{2}}$, else md is $\frac{\sigma_{\frac{k}{2}}+\sigma_{\frac{k}{2}+1}}{2}$. We choose the median value instead of the mean of the list because inconsistent σ_i(s) may affect the mean value too much while the median value will be more stable. For example, if a given list has values $\{19, 27, 40, 400\}$, then md is 33.5, but the mean is 121.5. Obviously, 33.5 is closer to most σ_is than 121.5 is, hence 33.5 can be used to identify the inconsistent trial(s) while 121.5 can not.

Second, we check each σ_i against md to see to what extent it diverges from md. For this purpose, we should set a threshold t and generate an interval $\mathsf{MDT} = [md/t, md * t]$. If $\sigma_i \in \mathsf{MDT}$, then trial i is consistent with most of the other trials and should be kept, otherwise, it should be identified as an inconsistent trial. Note that the σ_is always vary, so the interval MDT should not be too narrow, otherwise it will reject too many σ_is even if some are acceptable. In the other way round, MDT should not be too broad, otherwise some highly conflicting σ_is will be included. After looking through a large amount of trials results, at moment we think $t = 4$ is an applicable threshold.

Formally, we define the algorithm as follows.

Algorithm IncRemover
Begin
Input: k Trials with $(SEM_1, n_1), \ldots, (SEM_k, n_k)$ and t.
For $i = 1$ to k, let $\sigma_i = SEM_i\sqrt{n_i}$;
Sort $\sigma_1, \ldots, \sigma_k$ to $\sigma_1^s, \ldots, \sigma_k^s$ in ascending order;
If k is an odd value, let $md = \sigma^s_{\frac{k+1}{2}}$, else let $md = \frac{\sigma^s_{\frac{k}{2}}+\sigma^s_{\frac{k}{2}+1}}{2}$;
For $i = 1$ to k, if $md/t \leq \frac{\sigma_i}{md} \leq md * t$ then keep trial i, otherwise remove trial i.
Output: All remaining trials.
End

This algorithm is in fact an instantiated selection function (Def. 3) for which $\mathsf{R}(\mathsf{trial_i}, \mathsf{trial_j})$ iff $\sigma_i, \sigma_j \in \mathsf{MDT}$.

Proposition 1 *Given k trials, let σ_i be the standard deviation of trial i, and define a selection function S for which $\mathsf{R}(\mathsf{trial_i}, \mathsf{trial_j})$ iff $\sigma_i, \sigma_j \in \mathsf{MDT}$, then the output of algorithm IncRemover is the same as the selected subset by S.*

The proof is straightforward and omitted.

4.4 The Prognostic Method and Interval Method

In this subsection, we introduce the methods proposed in [28] to simulate meta-analysis when some trials results do not have complete information. It should be noted that these methods are also applicable to the *between group difference* of two drugs/therapies about two groups [28]. Below we present the two methods in terms of merging rules based on Def. 4 in the XML framework.

Assume there are $k + l$ trials altogether where k trials are with full information, i.e.,

$$(\mu_1, SEM_1, n_1), \ldots, (\mu_k, SEM_k, n_k)$$

and l trials with partial information, i.e.,

$$(\mu_{k+1}, n_{k+1}), \ldots, (\mu_{k+l}, n_{k+l}).$$

The task of meta-analysis is to get the merging result of those $k + l$ trials.

The prognostic method [28] uses the following equation to predict the missing SEM_j value for trial j ($k < j \leq k + l$) with sample size n_j, given that for k trials, each of which has the SEM_i value and its sample size n_i.

$$SEM_j = \frac{\sum_{i=1}^{k} SEM_i \sqrt{n_i}}{k\sqrt{n_j}} \tag{2}$$

When all SEM_j, $k < j \leq k + l$, are calculated, it is able to use the standard meta-analysis method, i.e., Equation 1, to merge all the $k + l$ trials.

This prognostic method can be defined as an instantiation of XML merging rule as follows.

Definition 9 *Given the following $k + l$ mergeable Result elements such that for $1 \leq i \leq k$, the SampleDist element in the ith Result element has both Mean and SEM sub tags, and for $k + l \leq j \leq k + l$, the SampleDist element in the jth Result element has only the Mean sub tag,*

⟨Result⟩	⟨Result⟩
...	...
⟨PatientsNum⟩ z_i ⟨/PatientsNum⟩	⟨PatientsNum⟩ z_j ⟨/PatientsNum⟩
⟨SampleDist ref = purpose⟩	⟨SampleDist ref = purpose⟩
⟨Mean⟩ μ_i ⟨/Mean⟩	⟨Mean⟩ μ_j ⟨/Mean⟩
⟨SEM⟩ SEM_i ⟨/SEM⟩	
⟨/SampleDist⟩	⟨/SampleDist⟩
⟨/Result⟩	⟨/Result⟩
for $1 \leq i \leq k$	for $k < j \leq k + l$

the meta-analysis result by the prognostic method is

⟨Result⟩
- ...
- ⟨PatientsNum⟩ $\sum_{i=1}^{k+l} z_i$ ⟨/PatientsNum⟩
- ⟨SampleDist ref = purpose⟩
  - ⟨Mean⟩ μ ⟨/Mean⟩
  - ⟨SEM⟩ SEM ⟨/SEM⟩
- ⟨/SampleDist⟩

⟨/Result⟩

where μ, SEM are obtained from Equation 1 for which SEM_j are obtained from Equation 2, $k < j \leq k + l$.

In contrast, instead of estimating a single value for each missing SEM as done in the prognostic method, the interval method [28] estimates a reliable interval for each missing SEM.

Let

$$\mu_{k+l}^{1} = \frac{\sum_{i=1}^{k} \mu_i \omega_i + \sum_{i=k+1}^{k+l} n_i \mu_i / \sigma_i^2}{\sum_{i=1}^{k} \omega_i + \sum_{i=k+1}^{k+l} n_i / \sigma_i^2}$$

and

$$\mu_{k+l}^2 = \frac{\sum_{i=1}^{k} \mu_i \omega_i + \sum_{i=k+1}^{k+l} n_i \mu_i / \sigma_i'^2}{\sum_{i=1}^{k} \omega_i + \sum_{i=k+1}^{k+l} n_i / \sigma_i'^2}.$$

$\forall i, k+1 \leq i \leq k+l$, we let

$$\sigma_i = \sigma_{min}, \sigma_i' = \sigma_{max}, \text{if } \mu_i \leq \mu^k,$$

and

$$\sigma_i = \sigma_{max}, \sigma_i' = \sigma_{min}, \text{if } \mu_i > \mu^k,$$

then the interval method gives the following result.

Let $X_i \sim N(\mu_i, SEM_i)$, $1 \leq i \leq k+l$, denote the ith sampling distribution with sample size n_i such that SEM_i is assumed missing when $i > k$, then the merged result $N(\mu, SEM)$ applying the interval method to these $k+l$ trials is

$$\mu \in [\mu_{k+l}^1, \mu_{k+l}^2], SEM^2 \in [\frac{1}{\sum_{i=1}^{k} \omega_i + \sum_{i=k+1}^{k+l} n_i / \sigma_{min}^2}, \frac{1}{\sum_{i=1}^{k} \omega_i + \sum_{i=k+1}^{k+l} n_i / \sigma_{max}^2}]. \quad (3)$$

The interval method is represented as an instantiation of XML merging rule as follows.

Definition 10 *Given the following $k+l$ mergeable Result elements such that for $1 \leq i \leq k$, the SampleDist element in the ith Result element has both Mean and SEM sub tags, and for $k+l \leq j \leq k+l$, the SampleDist element in the jth Result element has only the Mean sub tag,*

⟨Result⟩	⟨Result⟩
...	...
⟨PatientsNum⟩ z_i ⟨/PatientsNum⟩	⟨PatientsNum⟩ z_j ⟨/PatientsNum⟩
⟨SampleDist ref = purpose⟩	⟨SampleDist ref = purpose⟩
⟨Mean⟩ μ_i ⟨/Mean⟩	⟨Mean⟩ μ_j ⟨/Mean⟩
⟨SEM⟩ SEM_i ⟨/SEM⟩	
⟨/SampleDist⟩	⟨/SampleDist⟩
⟨/Result⟩	⟨/Result⟩
for $1 \leq i \leq k$	for $k < j \leq k+l$

the meta-analysis result by the interval method is

⟨Result⟩
...
⟨PatientsNum⟩ $\sum_{i=1}^{k+l} z_i$ ⟨/PatientsNum⟩
⟨SampleDist ref = purpose⟩
⟨MeanInv⟩ μ ⟨/MeanInv⟩
⟨SEMInv⟩ SEM ⟨/SEMInv⟩
⟨/SampleDist⟩
⟨/Result⟩

where μ and SEM are described by Equation 3.

Recall that an XML merging framework is represented by a pair of a selection function and a merging rule. Until now, with Def. 8 and the Algorithm in the last subsection as two selection functions S_1 and S_2, respectively, i.e., S_1 is used to select mergeable trials and S_2 is used to select consistent trials, and with Def. 9 and Def. 10 as two merging rules M_1 and M_2, alternatively, we have created two instantiated XML merging frameworks $(S_2 \circ S_1, M_1)$ and $(S_2 \circ S_1, M_2)$ where $S_2 \circ S_1$ is the compound function of S_1 and S_2 which means first S_1 is used to select and then S_2 is used to select from the result of S_1.

5 Case Studies

5.1 A Case Study of Diabetes Medications

In this subsection, we use the data from oral diabetes medication for adults with Type-2 diabetes as our first case study.

Many research papers and reports have been published to show the effectiveness of various oral medications for Type-2 diabetes (e.g., [27, 9, 41, 35, 29], etc). Clinicians and patients need a thorough comparison of these oral medications with respect to different aspects of Type-2 diabetes. Meta-analysis is the most frequently used technique for this purpose. It systematically reviews and compares each pair of drugs or therapies from different perspectives. For oral medication of Type-2 diabetes, meta-analysis [3] compares each pair of drugs on systolic blood pressure (SBP for short), diastolic blood pressure (DBP), low density lipoprotein cholesterol (LDL-C) and high density lipoprotein cholesterol (HDL-C), etc.

In this section, we create the XML documents for clinical trials reports and then merge the information contained in such XML documents. Here the meta-analysis is on the between group differences on the effectiveness of pairs of drugs for LDL-C.

Example 3 *(Triazolidinedione versus second generation Sulfonylureas) Low density lipoprotein effect is studied by many papers. They compare LDL-C between different trial groups. For example, to compare Thiazolidinedione and second generation Sulfonylureas, we get five clinical trials reports, i.e., [27, 9, 41, 35, 29].*

Due to the limitation of space, we only provide a simplified XML document for [41].

⟨Trial⟩
  ⟨Source⟩
    ⟨URL⟩*http://www3.interscience.wiley.com/cgi-bin/fulltext/118782480/PDFSTART*
    ⟨/URL⟩
    ⟨Title⟩*Sustained effects of pioglitazone vs. glibenclamide on insulin sensitivity, glycaemic control, and lipid profiles in patients with Type 2 diabetes*
    ⟨/Title⟩
    ⟨Author⟩*M. H. Tan, D. Johns, J. Strand, et al*⟨/Author⟩
  ⟨/Source⟩
  ⟨Objective⟩
    ⟨Drug id = "P1"⟩
      ⟨Name⟩*Pioglitazone*⟨/Name⟩
      ⟨DrugCategory id = "Tria"⟩*Triazolidinedione*⟨/DrugCategory⟩

⟨/Drug⟩
⟨Drug id = "G1"⟩
⟨Name⟩*Glibenclamide*⟨/Name⟩
⟨DrugCategory id = "sgs"⟩*second generation Sulfonylureas*
⟨/DrugCategory⟩
⟨/Drug⟩
⟨Aim⟩*To compare effects of different oral hypoglycemic drugs as first-line therapy on lipoprotein subfractions in type 2 diabetes*
⟨/Aim⟩
⟨PatientsType⟩*type 2 Diabetes*⟨/PatientsType⟩
⟨/Objective⟩
⟨MainOutcome⟩
...
⟨Result⟩
⟨Drug ref = "P1vsG1"/⟩
⟨Duration⟩*52 weeks*⟨/Duration⟩
⟨Unit⟩*mg/dl*⟨/Unit⟩
⟨PatientsNum⟩*100*⟨/PatientsNum⟩
⟨SampleDistvalue = "P1vsG1"⟩
⟨Mean⟩*6.63*⟨/Mean⟩
⟨/SampleDist⟩
⟨/Result⟩
...
⟨/MainOutcome⟩
⟨Conclusion⟩
⟨CompareEfficacy⟩
⟨Drug ref = "P1"/⟩
⟨Degree⟩*more sustained*⟨/Degree⟩
⟨Drug ref = "G1"⟩
⟨Duration⟩*52 weeks*⟨/Duration⟩
⟨/CompareEfficacy⟩
⟨/Conclusion⟩
⟨/Trial⟩

Note that in [41], we have mean change of drug P1 as 0.14 and of G1 as -0.03 in the unit of mmol/L. After using ontologies to relate mmol/L and mg/dl, we changed the unit of measurement to mg/dl, and obtained the between group difference of P1 and G1 as $6.63 = (0.14 + 0.03) * 39$ *with the unit of mg/dl.*

The sampling distributions (in mg/dL) from these five trials are as follows.

[27]: $X_{LT} \sim N(10.5, 14.44)$ *with* $n = 20$.

[9]: $X_{CM} \sim N(11.31, 1.59)$ *with* $n = 620$.

[41]: $X_{TJ} \sim N(6.63, SEM_{TJ})$ *with* $n = 100$.

[35]: $X_{PM} \sim N(5, 6.04)$ *with* $n = 86$.

[29]: $X_{MC} \sim N(14.6, SEM_{MC})$ *with* $n = 315$.

Here n is the size of samples (number of patients) in each group of a trial, and SEM_{TJ} and SEM_{MC} stand for the missing values (SEM value) from their respective trials data.

There are two missing SEM values. When applying the prognostic method, we get the difference between groups in LDL-C as

$$X_P \sim N[11.35, 1.32].$$

Alternatively, if we use the interval method, we get

$$X_I \sim N([10.91, 11.88], [1.20, 1.38]).$$

In [3], meta-analysis with known SEM_{TJ} and SEM_{MC} gives $X_{BW} \sim N(10.4, 1.61)$ from these five trials. X_P is reasonably close to X_{BW}.

The XML output by the prognostic method (Definition 9) is as follows.

```
...
⟨Result⟩
   ...
   ⟨PatientsNum⟩ 1141 ⟨/PatientsNum⟩
   ⟨SampleDist ref = "Triazolidinedione vs second generation
                                          Sulfonylureas"⟩
      ⟨Mean⟩ 11.35 ⟨/Mean⟩
      ⟨SEM⟩ 1.31 ⟨/SEM⟩
   ⟨/SampleDist⟩
⟨/Result⟩
...
```

The XML output by the interval method (Definition 10) is as follows.

```
...
⟨Result⟩
   ...
   ⟨PatientsNum⟩ 1141 ⟨/PatientsNum⟩
   ⟨SampleDist ref = "Triazolidinedione vs second generation
                                          Sulfonylureas"⟩
      ⟨MeanInv⟩
         ⟨Min⟩ 10.91 ⟨/Min⟩
         ⟨Max⟩ 11.88 ⟨/Max⟩
      ⟨/MeanInv⟩
      ⟨SEMInv⟩
         ⟨Min⟩ 1.20 ⟨/Min⟩
         ⟨Max⟩ 1.38 ⟨/Max⟩
      ⟨/SEMInv⟩
   ⟨/SampleDist⟩
⟨/Result⟩
...
```

5.2 A Case Study on neurocognitive outcomes

In this subsection, we use data of neurocognitive outcomes after off-pump versus on-pump coronary revascularisation as our second case study.

Off-pump (beating heart) coronary artery bypass grafting (CABG) is very popular as it is considered having numerous theoretical benefits including lower incidence of stroke and neurocognitive dysfunction. Therefore, considerable attentions have been devoted to this area (e.g., [14, 25, 26, 43], etc). We focus on a meta-analysis paper [31] on this topic which undertook quantitative systematic reviews to assess whether there were significant differences in neurocognitive outcomes in patients after undergoing off-pump versus on-pump CABG.

As [31] provided a set of trials with full statistical information, i.e., both the mean and the SEM values, in order to apply methods mentioned in last section, we deleted an SEM value from a trial selected randomly from a set of trials, and applied the prognostic and interval methods to predict the missing value. We then applied the meta-analysis method to merge the trial with the predicted SEM value together with the rest of trials in the group to see how close this new result is to the original meta-analysis result.

Furthermore, as the traditional method for trials with incomplete information always abandons trials with incomplete information, we also compared our methods with this traditional method.

In the following example, we create the XML documents for clinical trials reports and then merge the information contained in such XML documents.

Example 4 *(neurocognitive outcomes after off-pump versus on-pump coronary revascularisation) Neurocognitive outcomes after off-pump or on-pump coronary revascularisation is studied by many papers. Here, we get four clinical trials reports, i.e., [14, 25, 26, 43], to survey whether there were significant differences in neurocognitive outcomes in patients after undergoing off-pump versus on-pump CABG.*

Due to the limitation of space, we only provide a simplified XML document for [14].

⟨Trial⟩
  ⟨Source⟩
    ⟨URL⟩*http://ats.ctsnetjournals.org/cgi/content/abstract/81/6/2105*⟨/URL⟩
    ⟨Title⟩*Neurocognitive Outcomes in Off-Pump Versus On-Pump Bypass Surgery: A Randomized Controlled Trial*
    ⟨/Title⟩
    ⟨Author⟩*Ernest C S, Worcester M U, Tatoulis J, Elliott P C, Murphy B M, Higgins R O, LeGrande M R, Goble A J*
    ⟨/Author⟩
  ⟨/Source⟩
  ...
  ⟨MainOutcome⟩
    ...
    ⟨Result⟩
      ⟨Drug ref = "off−pump vs on−pump coronary revascularisation"/⟩
      ⟨PatientsNum⟩*47*⟨/PatientsNum⟩
      ⟨Duration⟩*47*⟨/Duration⟩

⟨SampleDistvalue = "off − pump vs on − pump coronary revascularisation"⟩
⟨Mean⟩*-0.34*⟨/Mean⟩
⟨SEM⟩*2.49*⟨/SEM⟩
⟨/SampleDist⟩
⟨/Result⟩
...
⟨/MainOutcome⟩
...
⟨/Trial⟩

The sampling distributions (no unit) from these four trials are as follows.
[14]: $X_{EW} \sim N(-0.34, 2.49)$ *with* $n = 47$.
[25]: $X_{LL} \sim N(1.00, 3.71)$ *with* $n = 27$.
[26]: $X_{LS} \sim N(4.40, 1.82)$ *with* $n = 54$.
[43]: $X_{VJ} \sim N(-4.00, 1.26)$ *with* $n = 130$.

Here we just delete the SEM value of X_{EW} *(others are similar), when applying the prognostic method, we get the merged sampling distribution as*

$$X_P \sim N[-0.99, 0.91].$$

Alternatively, if we use the interval method, we get

$$X_I \sim N([-1.07, -0.92], [0.89, 0.94]).$$

The traditional meta-analysis with full statistical data, i.e., $SEM_{EW} = 2.49$ *is known, gives* $X_{full} \sim N(-1.01, 0.93)$ *from these four trials, and traditional meta-analysis for trials with incomplete information, i.e., abandoning* X_{EW}*, gives* $X_{trad} \sim N(-1.11, 1.00)$*. Obviously, we have that* X_P *is closer to* X_{full} *than* X_{trad}*.*

The XML output by the prognostic method (Definition 9) is as follows.

...
⟨Result⟩
...
⟨PatientsNum⟩ 258 ⟨/PatientsNum⟩
⟨SampleDist ref = "off−pump vs on−pump coronary revascularisation"⟩
⟨Mean⟩ − 0.99 ⟨/Mean⟩
⟨SEM⟩ 0.91 ⟨/SEM⟩
⟨/SampleDist⟩
⟨/Result⟩
...

The XML output by the interval method (Definition 10) is as follows.

...
⟨Result⟩
...
⟨PatientsNum⟩ 258 ⟨/PatientsNum⟩
⟨SampleDist ref = "off−pump v on−pump coronary revascularisation"⟩
⟨MeanInv⟩
⟨Min⟩ − 1.07 ⟨/Min⟩

```
        ⟨Max⟩ − 0.92 ⟨/Max⟩
      ⟨/MeanInv⟩
      ⟨SEMInv⟩
        ⟨Min⟩ 0.89 ⟨/Min⟩
        ⟨Max⟩ 0.94 ⟨/Max⟩
      ⟨/SEMInv⟩
    ⟨/SampleDist⟩
⟨/Result⟩
...
```

6 Conclusion

In this paper, we proposed an XML based framework to represent clinical trials information and then to merge them which make this framework an automatic tool for meta-analyses. The main task is to represent and merge the statistical information in XML documents. Moreover, we used two case studies, the Type 2 diabetes case and neurocognitive outcomes after off-pump versus on-pump coronary revascularisation, to verify our framework.

Dealing with missing data in statistics, especially in meta-analysis is a very important topic (e.g., [24], [6], [44]). However, there are hardly any papers focusing on missing standard errors. [28] proposed some important results about how to deal with this situation. This paper used the methods in [28] to create a formal XML merging framework.

There are a number of issues we will further look at. First, improvements can be made on the XML document structure to cover a wider range of clinical trials reports. Second, although we had a brief discussion in Section 2.4 about dealing with semantic heterogeneity, the role of ontologies, indexing schemes, and restricted vocabularies, etc, for both the definitions of the XML tags and for the text entries, should be further studied. The creation of an application oriented ontology should facilitate the automated merging. Third, we will examine information extraction tools to see how information from clinical trials reports can be efficiently extracted, in order to generate XML documents automatically.

Appendix

In this appendix, we will provide a full structure of our XML documents. To accommodate our special need of recording clinical trials information, we investigated many clinical trials reports and therefore adapted the DTD of XML documents as follows.

First, information in each clinical report will be put in a Trial element. We define the Trial element and its children as in Fig. 2. Here the ? sign shows that the SideEffect child-element is optional.

The Source element is used to provide some general information for a clinical trial report. We define it as in Fig. 3.

```
⟨!ELEMENT Trial (Source, Objective, MainOutcome, SideEffect?, Conclusion)⟩
```

Fig. 2. DTD adaption of the Title element

```
⟨!ELEMENT Source (URL, Title, Author)⟩
⟨!ELEMENT URL (#PCDATA)⟩
⟨!ELEMENT Title (#PCDATA)⟩
⟨!ELEMENT Author (#PCDATA)⟩
```

Fig. 3. DTD adaption of the Source element

Example 5 *The following is a Source element.*

⟨Source⟩
⟨URL⟩*http://www.neurology.org/cgi/content/abstract/65/9/1415*⟨/URL⟩
⟨Title⟩*Prevalence and size of directly detected patent foramen ovale in migraine with aura*
⟨/Title⟩
⟨Author⟩*Schwerzmann M, Nedeltchev K, Lagger F, Mattle HP, Windecker S, Meier B, et al*
⟨/Author⟩
⟨/Source⟩

The Objective element tells the objective of a clinical trial. We define it as in Fig. 4.

```
⟨!ELEMENT Objective (Drug, Aim, PatientsType)⟩
⟨!ELEMENT Drug (Name, DrugCategory?)⟩
⟨!ATTLIST Drug id CDATA #REQUIRED⟩
⟨!ELEMENT Name (#PCDATA)⟩
⟨!ELEMENT DrugCategory (#PCDATA)⟩
⟨!ELEMENT Aim (#PCDATA)⟩
⟨!ELEMENT PatientsType (#PCDATA)⟩
```

Fig. 4. DTD adaption of the Objective element

Here the DrugCategory element is for a category of drugs, e.g., Glibenclamide is a kind of second generation Sulfonylureas which is a category of drugs. In addition, so far the PatientsType element is a leaf element. For further study, it may need to be changed to a composite element containing some sub elements like AverageAge, Nationality, etc.

Example 6 *The following is an Objective element.*

⟨Objective⟩
⟨Drug id = "l1"⟩
⟨Name⟩*latanoprost*⟨/Name⟩
⟨/Drug⟩
⟨Aim⟩*To test the drug efficacy for intraocular pressure reduction*⟨/Aim⟩
⟨PatientsType⟩*Black American with Glaucoma*⟨/PatientsType⟩
⟨/Objective⟩

The MainOutcome element is defined as in Fig. 5.

⟨!ELEMENT MainOutcome (Result+)⟩

Fig. 5. DTD adaption of the MainOutcome element

Here the Result element is defined in Section 2.2. The reason why we put the Duration element as a child of the Result element instead of a child of the MainOutcome element is that a clinical trial can have more than one duration period. For example, a 12-month trial may provide results at 3th month, 6th month, 9th month and 12th month.

Example 7 *The following is a MainOutcome element.*

⟨MainOutcome⟩
⟨Result⟩
⟨Drug ref = "l1"/⟩
⟨Duration⟩*3 month*⟨/Duration⟩
⟨Unit⟩*mmol/L*⟨/Unit⟩
⟨PatientsNum⟩*247*⟨/PatientsNum⟩
⟨pValue⟩*0.016*⟨/pValue⟩
⟨SampleDist value = "intraocular pressure reduction"⟩
⟨Mean⟩*4.1*⟨/Mean⟩
⟨SEM⟩*3.8*⟨/SEM⟩
⟨/SampleDist⟩
⟨/Result⟩
⟨/MainOutcome⟩

The SideEffect element contained by the Trial element is defined as in Fig. 6.

The adverse event may be for a single drug or for comparing two drugs and the descriptions for side effects include (but not limited to)

- may (cause)
- * more conjunctival (than)
- * increased incident (than)

```
⟨!ELEMENT SideEffect (Report+)⟩
⟨!ELEMENT Report (AdverseEvent, (Drug, Degree|Compare))⟩
⟨!ELEMENT AdverseEvent (#PCDATA)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
⟨!ELEMENT Degree (#PCDATA)⟩
⟨!ELEMENT Compare (Drug, Degree, Drug)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
⟨!ELEMENT Degree (#PCDATA)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
```

Fig. 6. DTD adaption of the SideEffect element

- * higher percentage (than)
- well tolerated
- not severe

Here Words with * are used for comparison between two drugs.

Example 8 *The following is a SideEffect element.*

```
⟨SideEffect⟩
   ⟨Report⟩
      ⟨AdverseEvent⟩cough⟨/AdverseEvent⟩
      ⟨Compare⟩
         ⟨Drug ref = "p1"/⟩
         ⟨Degree⟩increased incident⟨/Degree⟩
         ⟨Drug ref = "11"/⟩
      ⟨/Compare⟩
   ⟨/Report⟩
⟨/SideEffect⟩
```

The Conclusion element is defined as in Fig. 7.
The descriptions for efficacies include (but not limited to)

- * significantly (better than)
- * no significantly difference (with)
- * greater hypotensive (efficacy) (than)
- * more effective (than)
- * not more effective (than)
- * equivalent (to)
- * significantly greater (than)

```
⟨!ELEMENT Conclusion (Efficacy*, CompareEfficacy*)⟩
⟨!ELEMENT Efficacy (Drug, Degree, Duration?, pValue?, FromBaseline?)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
⟨!ELEMENT Degree (#PCDATA)⟩
⟨!ELEMENT Duration (#PCDATA)⟩
⟨!ELEMENT pValue (#PCDATA)⟩
⟨!ELEMENT FromBaseline (#PCDATA)⟩
⟨!ELEMENT CompareEfficacy (Drug, Degree, Drug, Duration?, pValue?)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
⟨!ELEMENT Degree (#PCDATA)⟩
⟨!ELEMENT Drug EMPTTY⟩
⟨!ATTLIST Drug ref CDATA #REQUIRED⟩
⟨!ELEMENT Duration (#PCDATA)⟩
⟨!ELEMENT pValue (#PCDATA)⟩
```

Fig. 7. DTD adaption of the Conclusion element

- significantly
- * superior (than)

Similarly, words with * are used for comparison between two drugs.

Example 9 *The following is a Conclusion element.*

⟨Conclusion⟩
  ⟨Efficacy⟩
    ⟨Drug ref = "p1"/⟩
    ⟨Degree⟩*siginificantly*⟨/Degree⟩
    ⟨Duration⟩*52 weeks*⟨/Duration⟩
  ⟨/Efficacy⟩
⟨/Conclusion⟩

Finally, we provide an example of a clinical trial report and its full XML document as follows.

Example 10 *A clinical trial report entitiled "The effects of prostaglandin analogues on the blood aqueous barrier and corneal thickness of phakic patients with primary open-angle glaucoma and ocular hypertension" can be found at the following link* http://www.ncbi.nlm.nih.gov/pubmed/16936646?dopt=Abstract.

The authors are Arcieri ES, Pierre Filho PT, Wakamatsu TH, Costa VP.

The main summary is in its abstract as:

PURPOSE: To evaluate the effects of topical latanoprost, travoprost, and bimatoprost

on the blood-aqueous barrier and central corneal thickness (CCT) of patients with primary open-angle glaucoma (POAG) and ocular hypertension (OHT). DESIGN: Prospective, randomized, masked-observer, crossover clinical trial. METHODS: A total of 34 phakic patients with POAG or OHT with no previous history of intraocular surgery or uveitis completed the study. Patients were randomized to use latanoprost 0.005%, travoprost 0.004%, or bimatoprost 0.03% once daily (2000 hours) for 1 month, followed by a washout period of 4 weeks between each drug. Aqueous flare was measured with a laser flare metre. CCT was calculated as the average of five measurements using ultrasound pachymetry. All measurements were performed by a masked observer (1000 h). RESULTS: There were no statistically significant differences between baseline mean IOP, mean CCT, and mean flare values among the groups. There was no statistically significant increase in mean flare values from baseline in all groups ($P > 0.05$). There were no statistically significant differences between mean flare values among the groups ($P > 0.05$). All medications significantly reduced the mean IOP from baseline ($P < 0.0001$). IOP reduction obtained with travoprost (7.3+/-3.8 mmHg) was significantly higher than that obtained with latanoprost (4.7+/-4.2 mmHg) (P=0.01). A statistically significant reduction in mean CCT (0.6+/-1.3%) from baseline was observed when patients instilled bimatoprost (P=0.01). CONCLUSIONS: Latanoprost, travoprost, and bimatoprost had no statistically significant effect on the blood-aqueous barrier of phakic patients with POAG or OHT. Bimatoprost may be associated with a clinically irrelevant reduction in mean CCT.

The corresponding full XML document is extracted as follows.

⟨Trial⟩
⟨Source⟩
⟨URL⟩*http://www.ncbi.nlm.nih.gov/pubmed/16936646?dopt=Abstract*⟨/URL⟩
⟨Title⟩*The effects of prostaglandin analogues on the blood aqueous barrier and corneal thickness of phakic patients with primary open-angle glaucoma and ocular hypertension*
⟨/Title⟩
⟨Author⟩*Arcieri ES, Pierre Filho PT, Wakamatsu TH, Costa VP* ⟨/Author⟩
⟨/Source⟩
⟨Objective⟩
⟨Drug id = "drug − a"⟩
⟨Name⟩*latanoprost 0.005%*⟨/Name⟩
⟨/Drug⟩
⟨Drug id = "drug − b"⟩
⟨Name⟩*travoprost 0.004%*⟨/Name⟩
⟨/Drug⟩
⟨Drug id = "drug − c"⟩
⟨Name⟩*bimatoprost 0.03%*⟨/Name⟩
⟨/Drug⟩
⟨Aim⟩*Blood-adqueous barrier and central corneal thickness*⟨/Aim⟩
⟨PatientsType⟩*primary open-angle glaucoma (POAG) and ocular hypertension (OHT)*
⟨/PatientsType⟩

⟨/Objective⟩
⟨MainOutcome⟩
⟨Result⟩
⟨Drug ref = "drug − a"/⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨Unit⟩*mmHg*⟨/Unit⟩
⟨PatientsNum⟩*34*⟨/PatientsNum⟩
⟨pValue⟩*0.01*⟨/pValue⟩
⟨SampleDist value = "IOP Reduction"⟩
⟨Mean⟩*4.7*⟨/Mean⟩
⟨SEM⟩*4.2*⟨/SEM⟩
⟨/SampleDist⟩
⟨/Result⟩
⟨Result⟩
⟨Drug ref = "drug − b"/⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨Unit⟩*mmHg*⟨/Unit⟩
⟨PatientsNum⟩*34*⟨/PatientsNum⟩
⟨pValue⟩*0.0001*⟨/pValue⟩
⟨SampleDist value = "IOP Reduction"⟩
⟨Mean⟩*7.3*⟨/Mean⟩
⟨SEM⟩*3.8*⟨/SEM⟩
⟨/SampleDist⟩
⟨/Result⟩
⟨/MainOutcome⟩
⟨SideEffect⟩
⟨Report⟩
⟨AdverseEvent⟩*irrelevant reduction in mean CCT*⟨/AdverseEvent⟩
⟨Drug ref = "drug − c"/⟩
⟨Degree⟩*may*⟨/Degree⟩
⟨/Report⟩
⟨/SideEffect⟩
⟨Conclusion⟩
⟨Efficacy⟩
⟨Drug ref = "drug − a"/⟩
⟨Degree⟩*significantly*⟨/Degree⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨FromBaseline⟩*yes*⟨/Duration⟩
⟨/Efficacy⟩
⟨Efficacy⟩
⟨Drug ref = "drug − b"/⟩
⟨Degree⟩*significantly*⟨/Degree⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨FromBaseline⟩*yes*⟨/Duration⟩
⟨/Efficacy⟩

⟨Efficacy⟩
⟨Drug ref = "drug − c"/⟩
⟨Degree⟩*significantly*⟨/Degree⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨FromBaseline⟩*yes*⟨/Duration⟩
⟨/Efficacy⟩
⟨CompareEfficacy⟩
⟨Drug ref = "drug − a"/⟩
⟨Degree⟩*no significant difference*⟨/Degree⟩
⟨Drug ref = "drug − b"/⟩
⟨Duration⟩*1 month*⟨/Duration⟩
⟨/CompareEfficacy⟩
⟨/Conclusion⟩
⟨/Trial⟩

References

1. Abiteboul S, Segoufin L, and Vianu V. Representing and querying XML with incomplete information. *ACM Trans. Database Syst.*, 31(1):208-254, 2006.
2. Barbara D, Garcia-Molina H, and Porter D. The management of probabilistic data. *IEEE Trans. on Knowledge and Data Engineering*, 4(5):487-502, 1992.
3. Bolen S, Wilson L, Vassy J, Feldman L, Yeh J, Marinopoulos S, Wilson R, Cheng D, Wiley C, Selvin E, Malaka D, Akpala C, Brancati F, and Bass E. Comparative effectiveness and safety of oral diabetes medications for adults with type 2 diabetes. *Comparative effectiveness review*, No. 8, 2007.
4. Chiselita D, Antohi I, Medvichi R, and Danielescu C. Comparative analysis of the efficacy and safety of latanoprost, travoprost and the fixed combination timolol-dorzolamide; a prospective, randomized, masked, cross-over design study. *Oftalmologia*, 49(3):39-45, 2005.
5. Crangle CE, Cherry JM, Hong EL, and Zbyslaw A. Mining experimental evidence of molecular function claims from the literature. *Bioinformatics*, 23, 3232-3240, 2007.
6. Copas J B and Eguchi S. Local model uncertainty and incomplete-data bias. *J. R. Statist. Soc. B*, 67(4):459-513, 2005.
7. Cantor LB, Hoop J, Morgan L, Wudunn D, and Catoira Y. Bimatoprost-Travoprost Study Group, Intraocular pressure-lowering efficacy of bimatoprost 0.03% and travoprost 0.004$ in patients with glaucoma or ocular hypertension. *Br J Ophthalmol*, 90(11):1370-3, 2006.
8. Cowie J and Lehnert W. Information extraction. *Communications of ACM*, 39:81C91, 1996.
9. Charbonnel B H, Matthews D R, Schernthaner G, Hanefeld M, Brunetti P, for the QUARTET Study Group. A long-term comparison of pioglitazone and gliclazide in patients with Type 2 diabetes mellitus: a randomized, double-blind, parallel-group comparison trial. *Diabetic Medicine*, 22:399C405, 2004.
10. Cunningham H, Maynard D, Bontcheva K, and Tablan V. Gate: A framework and graphical development environment for robust nlp tools and applications. *In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.

11. Combi C, Oliboni B, and Rossato R. Merging multimedia presentations and semi-structured temporal data: a graph-based model and its application to clinical information. *Artificial Intelligence in Medicine*, 2005.
12. Cavallo R and Pittarelli M. The theory of probabilistic databases. *Proc. of VLBD'87*, 71-81, 1987.
13. Clegg A and Shepherd A. Benchmarking natural-language parsers for biological applications using dependency graphs. *BMC Bioinformatics*, 8:24, 2007.
14. Ernest C S, Worcester M U, Tatoulis J, Elliott P C, Murphy B M, Higgins R O, LeGrande M R, and Goble A J. Neurocognitive outcomes in off-pump versus on-pump bypass surgery: a randomized controlled trial. *Ann Thorac Surg*, 81(6):2105-14, 2006.
15. Gracia-Feijo J, Martinez-de-la-Casa JM, Castillo A, Mendez C, Fernandez-Vidal A, and Garcia-Sanchez J. Circadian IOP-lowering efficacy of travoprost 0.004$ ophthalmic solution compared to latanoprost 0.005%. *Curr Med Res Opin*, 22(9):1689-97, 2006.
16. Greenhalgh T. *How to Read a Paper: The Basics of Evidence-Based Medicine*. BMJ Press, 1997.
17. Hunter A and Liu W. Fusion rules for merging uncertain information. *Information Fusion*, 7:97-114, 2006.
18. Hunter A and Liu W. Merging uncertain information with semantic heterogeneity in XML. *Knowledge and Information Systems*, 9(2):230-258, 2006.
19. Hunter A and Liu W. A logical reasoning framework for modelling and merging uncertain semi-structured information. *Modern Information Processing: From Theory to Applications*, B. Bouchon-Meunier, G. Coletti and R.R. Yager (eds), Elsevier, 345-356, 2006.
20. Hunter L, Lu Z, Firby J, Baumgartner WA Jr, Johnson HL, Ogren PV, and Cohen KB. An open-source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-specific gene expression. *BMC Bioinformatics*, 31 9(1):78, 2008.
21. Howard S, Silvia ON, Brian E, John S, Sushanta M, Theresa A, and Michael V. The Safety and Efficacy of Travoprost 0.004%/Timolol 0.5% Fixed Combination Ophthalmic Solution. *Ame J Ophthalmology*, 140(1):1-8, 2005.
22. Hirschman L, Yeh A, Blaschke C, and Valencia A. Critical assessment of information extraction for biology. *BMC Bioinformatics*, 6 (Suppl 1):S11, 2005.
23. Keulen M van, Keijzer A de and Alink W. A probabilistic XML approach to data integration. *Proceedings of ICDE'05*, 459-470, 2005.
24. Lu G and Copas J B. Missing at Random, Likelihood Ignorability and Model Completeness. *The Annals of Statistics*, 32(2):754-765, 2004.
25. Lee J D, Lee S J, Tsushima W T, Yamauchi H, Lau W T, Popper J, Stein A, Johnson D, Lee D, Petrovitch H, and Dang C R. Benefits of off-pump bypass on neurologic and clinical morbidity: a prospective randomized trial. *Ann Thorac Surg*, 76(1):18-25, 2003.
26. Lund C, Sundet K, Tennoe B, Hol P K, Rein K A, Fosse E, Russell D. Cerebralischemic injury and cognitive impairment after off-pump and on-pump coronary artery bypass grafting surgery. *Ann Thorac Surg*, 80:2126-31, 2005.
27. Lawrence J, Reid J, Taylor G, Stirling C, and Reckless J. Favorable Effects of Pioglitazone and Metformin Compared With Gliclazide on Lipoprotein Subfractions in Overweight Patients With Early Type 2 Diabetes. *Diabetes care*, 27(1):41-46, 2004.
28. Ma J, Liu W, Hunter A, and Zhang W. Performing meta-analysis with incomplete statistical information in clinical trials. *BMC Informatics*, Aug 18;8(1):56, 2008.

29. Matthews D R, Charbonnel B H, Hanefeld M, Brunetti P, and Schernthaner G. Long-term therapy with addition of pioglitazone to metformin compared with the addition of gliclazide to metformin in patients with type 2 diabetes: a randomized, comparative study. *Diabetes Metab Res Rev*, 21:167-174, 2005.
30. Michael T, David W, and Alan L. Projected impact of travoprost versus timolol and latanoprost on visual field deficit progression and costs among black glaucoma subjects. *Trans Am Ophthalmol Soc*, 100:109-118, 2002.
31. Marasco S F, Sharwood L N, and Abramson M J. No improvement in neurocognitive outcomes after off-pump versus on-pump coronary revascularisation: a meta-analysis. *European Journal of Cardio-thoracic Surgery*, 33:961-970, 2008.
32. Noecker RJ, Earl ML, Mundorf TK, Silvestein SM, and Phillips MP. Comparing bimatoprost and travoprost in black Americans. *Curr Med Res Opin*, 22(11):2175-80, 2006.
33. Nierman A and Jagadish H. ProTDB: Probabilistic data in XML. In *Proc. of VLDB'02*, LNCS2590:646-657. Springer, 2002.
34. Nicola C, Michele V, Tiziana T, Francesco C, and Carlo S. Effects of Travoprost Eye Drops on Intraocular Pressure and Pulsatile Ocular Blood Flow: A 180-Day, Randomized, Double-Masked Comparison with Latanoprost Eye Drops in Patients with Open-Angle Glaucoma. *Curr Ther Res*, 64(7):389-400, 2003.
35. Pf*ü*zner A, Marx N, L*ü*ben G, Langenfeld M, Walcher D, Konrad T, and Forst T. Improvement of Cardiovascular Risk Markers by Pioglitazone Is Independent From Glycemic Control Results From the Pioneer Study. *Journal of the American College of Cardiology*, 45(12):1925-1931, 2005.
36. http://protege.stanford.edu/.
37. Parmarksiz S, Yuksel N, Karabas VL, Ozkan B, Demirci G, Caglar Y. A comparison of travoprost, latanoprost and the fixed combination of dorzolamide and timolol in patients with pseudoexfoliation glaucoma. *Eur J Ophthalmol*, 16(1):73-80, 2006.
38. Qi G and Hunter A. Measuring incoherence in description logic-based ontologies. *Proceedings of the International Semantic Web Conference (ISWC'07)*:381-394, Springer.
39. Radev D, Fan W, Qi H, Wu H, and Grewal A. Probabilistic question answering on the Web. *Proc. of WWW'02*, 408-419, 2002.
40. Stefan C, Nenciu A, Malcea C, and Tebeanu E. Axial length of the ocular globe and hypotensive effect in glaucoma therapy with prostaglandin analogs. *Oftalmologia*, 49(4):47-50, 2005.
41. Tan M H, Johns D, Strand J, Halse J, Madsbad S, Eriksson J W, Clausen J, Konkoy C S, Herz M, for the GLAC Study Group. Sustained effects of pioglitazone vs. glibenclamide on insulin sensitivity, glycaemic control, and lipid profiles in patients with Type 2 diabetes. *Diabetic Medicine*, 21:859-866, 2004.
42. Wang Y, Liu W., and Bell D A. Combining uncertain outputs from multiple ontology matchers. *Proceedings of the First International Conference on Scalable Uncertainty Management (SUM07)*:201-214. Springer.
43. van Dijk D, Jansen E W L, Hijman R, Nierich A P, Diephuis J C, Moons K G M,Lahpor J R, Borst C, Keizer A M A, Grobbee D E, de Jaegere P P, and Kalkman C J. Cognitive outcome after off-pump and on-pump coronary artery bypass graft surgery: a randomized trial. *JAMA*, 287:1405-12, 2002.
44. White, I. Missing data and departures from randomised treatment in pragmatic trials. http://www.mrc-bsu.cam.ac.uk/BSUsite/Research/Section11.shtml.
45. Zupan B, Demsar J, Katten M, Ohori M, Graefen M, Bojanec M, and Beck R. Orange and decisions-at-hand: bridging predictive data mining and decision support. *Proc. of*

ECML/PKDD'01 workshop on Integrating Aspects of Data Mining Decision Support and Meta-Learning, 151-162, September 2001.
46. http://en.wikipedia.org/wiki/Sampling_distribution.