

Using Clinical Preferences in Argumentation about Evidence from Clinical Trials

Anthony Hunter
Dept. of Computer Science
University College London
London, WC1E 6BT, UK
a.hunter@cs.ucl.ac.uk

Matthew Williams
Dept. of Academic Oncology
Royal Free Hospital
London, NW3 2QC, UK
matt.williams@nhs.net

ABSTRACT

Medical practice is increasingly based on the best available evidence, but the volume of information requires many clinicians to rely on systematic reviews rather than the primary evidence. However, these reviews are difficult to maintain, and often do not appear transparent to clinicians reading them. In a previous paper [8], we have proposed a general language for representing knowledge from clinical trials and a framework that allows reasoning with that knowledge in order to construct and evaluate arguments and counterarguments that aggregate that knowledge. However, clinicians need to feel that such a framework is responsive to their assessment of the strengths and weaknesses of different types of evidence. In this paper, we use a specific version of this existing framework to show how we can capture clinical preferences over types of evidence, and we evaluate this in a pilot study, comparing our system against the choices made by clinicians. This pilot study shows how individual clinicians aggregate evidence based on their preferences over the relative significance of the items of evidence, and it shows how our argumentation system can replicate this behaviour.

Categories and Subject Descriptors

J.Computer Applications [J.3 LIFE AND MEDICAL SCIENCES]: Medical information systems

General Terms

Theory

Keywords

Medical decision-support, Evidence-based decision making, Argument systems, Aggregation of information, Biomedical knowledge representation.

1. INTRODUCTION

The systematic use of evidence is already established in healthcare. However, the volume of new knowledge on a

subject means that it is difficult for a decision maker to locate evidence that is relevant to their needs. In addition to the difficulty presented by the sheer volume of information, the evidence is often conceptually complex, heterogeneous, incomplete and inconsistent. As a result, clinicians have developed various approaches (e.g. systematic reviews, meta-analyses, clinical guidelines) to summarize the available evidence. However, these summaries often lack a clear and specified link with the underlying primary evidence, and may suppress conflict within it, and as a result the summaries are not transparent to the user. Thus, if more evidence becomes available, it must be manually incorporated into an updated version of the summary (through a time-consuming process of repeating the summary analysis), and if a clinician disagrees with some of the evidence (e.g. local variations in diseases or treatments, or new information about the veracity of a published study), she has no way of seeing whether that disagreement has already been incorporated into the summary, or whether that disagreement would be important in determining the overall result.

We therefore seek ways to link the summaries more closely with the primary evidence, with the ultimate aim of being able to offer on-the-fly, automatically updated reviews which can be manipulated and queried by the user and which preserve the conflict present in the primary evidence. Not least, is the imperative to abstract away from the details of individual items of evidential knowledge, and to aggregate the evidence in a way that reduces the volume, complexity, inconsistency and incompleteness. Thus, it would be helpful to have a method for automatically analyzing and presenting the clinical trial results and the possible ways to aggregate those in an intuitive form, highlighting the points of agreement and conflict present within that evidence.

Our proposal in [9, 8] aims to suggest such a method. The first part of our proposal is a language that can be used to encode the published results in a semantically appropriate way, and methods for constructing a knowledge base from the encoded results. The second part of our framework allows the construction of arguments and counterarguments concerning possible aggregations of the evidence, and mechanisms for evaluating these in order to determine the winning arguments. The net result of this process is an automated on-the-fly synthesis of the evidence in the form of recommendations for which is the better treatment.

However, capturing and reasoning with the information in clinical trials is not enough. In order to improve the transparency of the system, we want to capture users' preferences over types of evidence (in terms of study design, geograph-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IHI'10, November 11–12, 2010, Arlington, Virginia, USA.
Copyright 2010 ACM 978-1-4503-0030-8/10/11 ...\$10.00.

Table 1: Four results obtained from the NICE Hypertension Guideline (GC34, pages 36-37) concerning angiotensin-converting inhibitors (ACE) and calcium channel blockers (CCB)

	Outcome	Result	Type
e_1	Mortality	ACE < CCB	MA
e_2	Stroke	ACE < CCB	MA
e_3	Heart failure	ACE > CCB	MA
e_4	Diabetes	ACE > CCB	MA

ical location or outcome measure), and we need to be able to explore the link between a clinician’s preferences and the output of such a system. In part, this builds upon existing work within the medical domain that aims to summarize key aspects of studies that should be considered when weighing evidence [10]. In this paper, we present an abbreviated version of our previous proposal, but concentrate on understanding and representing clinical preferences over types of evidence. We present a simple formal technique to representing preferences, and report the results of a pilot study that uses these techniques to capture clinicians’ preferences with respect to three different sets of trials, and then aims replicate the clinicians’ reasoning in our system.

Our framework builds on more general developments in the area of computational models of argument. These models aim to reflect how human argumentation uses conflicting information to construct and analyze arguments. There is a number of frameworks for computational models of argumentation. They incorporate a formal representation of individual arguments and techniques for comparing conflicting arguments (for reviews see [2, 3]). By basing our framework on these general models, we can harness theory and adapt implemented argumentation software.

We proceed as follows: (Section 2) We discuss how we can represent evidence from clinical trials in a tabular format; (Section 3) We review an abstract model of argumentation that we will incorporate in our general framework; (Section 4) We show how we can represent and compare arguments based on available evidence; (Section 5) We present a pilot study involving a group of clinicians; (Section 6) We discuss the viability of our framework from the point of view of obtaining appropriate data and harnessing existing software tools; (Section 7) We conclude with a discussion of our proposal and how it relates to the literature.

2. CLINICAL TRIALS

We represent evidence in a table. Each row is an item of evidence taken from a study such as a randomized clinical trial or a meta-analysis. The choice of columns depends on the available information and the criteria that will be used for aggregating the evidence. We give a very simple example in Table 1 concerning patients who require treatment for hypertension (data from www.nice.org.uk). The table incorporates the columns that would minimally be required for our framework, and we explain them as follows.

- The **Outcome** is the specification of the particular outcome that is being considered when comparing the two treatments. In the table, in each row, it is the proportion of patients who have the event or condition

(i.e. “mortality”, “stroke”, “heart failure” or “diabetes”) with the period of the trial.

- The **Result**, is a binary relation, denoted $>$ (superior), \sim (equal), and $<$ (inferior), over two treatments that is determined from the value of the outcome and an evaluation of whether the outcome indicator is desirable or undesirable for the patient class. For the first row, mortality is undesirable, and the expression ACE < CCB means that ACE is inferior to CCB, with respect to the outcome of interest (Mortality in this study). Note, the treatment on the left (respectively right) of the $>$ relation is the left arm (respectively right arm) of the trial or analysis.
- The **Evidence Type**, abbreviated by Type, specifies the type of study undertaken, e.g. randomized clinical trial (RCT), meta-analysis (MA), systematic review of cohort studies (SRC), cohort studies (Cohort), network analysis (NA), etc. It is an indicator of the quality of the evidence.

The set of attributes we have discussed here is only indicative. Normally, a small number of further attributes are useful for assessing and aggregating evidence (e.g. the number of patients involved in each trial, the geographical location for each trial, the drop-out rate for the trial, the methods of randomization for ensuring patients and clinician do not know which arm a patient is in, etc). For a general introduction to the nature of clinical trials, and a discussion of a wider range of attributes, see [11].

The patient class is an important attribute that can be captured about an item of evidence. In the above table, the patient class is people with “persistent raised blood pressure of 160/100 mmHg or more”. In our previous work, we showed how the patient class may involve a conjunction and/or disjunction of terms from a medical ontology and therefore description logics can be used to provide inferencing (see [1]). Similarly, treatments presented in the left arm and right arm can be composed for a conjunction and/or disjunction terms from an ontology. Again, medical ontologies cater for this by providing categories and relationships on treatments, substances used, and other characteristics. See [19, 9] for proposals for using a medical ontology in argumentation about clinical trials.

For simplicity, in the rest of this paper, we assume that the evidence concerns a particular, sensible patient class, and that each treatment in the left arm and right arm is atomic, and so we do not consider the ontological aspects of patient class or treatment further in the rest of this paper.

3. ARGUMENTATION

In this section, we review the proposal for abstract argumentation by Dung [4]. The simplest way to formalize a collection of arguments consists of just naming arguments (so, in a sense, treating them as atomic) and merely representing the fact that an argument is challenged by another (and so not indicating what the nature of the challenge is). In other words, a collection of arguments can be formalized as a directed binary graph.

DEFINITION 1. *An abstract argument graph is a pair $(\mathcal{A}, \mathcal{R})$ where \mathcal{A} is a set and \mathcal{R} is a binary relation over \mathcal{A} (in symbols, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$).*

Each element $a \in \mathcal{A}$ is called an **argument** and $a\mathcal{R}b$ means that a **attacks** b (accordingly, a is said to be an **attacker** of b). So a is a **counterargument** for b when $a\mathcal{R}b$ holds.

EXAMPLE 1. Consider arguments $A_1 = \text{“Patient has hypertension so prescribe diuretics”}$, $A_2 = \text{“Patient has hypertension so prescribe betablockers”}$, and $A_3 = \text{“Patient has emphysema which is a contraindication for betablockers”}$. Here, we assume that A_1 and A_2 attack each other because we should only give one treatment and so giving one precludes the other, and we assume that A_3 attacks A_2 because it provides a counterargument to A_2 . Hence, we get the following abstract argument graph.

$$A_1 \rightleftharpoons A_2 \leftarrow A_3$$

Arguments can work together as a coalition by attacking other arguments and by defending their members from attack as follows.

DEFINITION 2. Let $S \subseteq \mathcal{A}$ be a set of arguments.

- S **attacks** $b \in \mathcal{A}$ iff there is an argument $c \in S$ such that c attacks b .
- S **defends** $a \in \mathcal{A}$ iff for each argument $b \in \mathcal{A}$, if b attacks a then S attacks b .

EXAMPLE 2. Continuing Example 1, $\{A_1, A_3\}$ attacks A_2 and $\{A_1, A_3\}$ defends A_1 .

The following gives a requirement that should hold for a coalition of arguments to make sense. If it holds, it means that the arguments in the set offer a consistent view on the topic of the argument graph.

DEFINITION 3. A set $S \subseteq \mathcal{A}$ of arguments is **conflict-free** iff there are no a and b in S such that a attacks b .

EXAMPLE 3. Continuing Example 1, the conflict free sets are $\{\}$, $\{A_1\}$, $\{A_2\}$, $\{A_3\}$, and $\{A_1, A_3\}$. The other sets (i.e. $\{A_1, A_2\}$, $\{A_2, A_3\}$ and $\{A_1, A_2, A_3\}$) are not conflict free.

Now, we consider how we can find an acceptable set of arguments from an abstract argument graph. The simplest case of arguments that can be accepted is as follows.

DEFINITION 4. A set $S \subseteq \mathcal{A}$ of arguments is **admissible** iff S is conflict-free and defends all its elements.

EXAMPLE 4. Continuing Example 1, the admissible sets are $\{\}$, $\{A_1\}$, $\{A_3\}$, and $\{A_1, A_3\}$. The other sets (i.e. $\{A_2\}$, $\{A_1, A_2\}$, $\{A_2, A_3\}$ and $\{A_1, A_2, A_3\}$) are not admissible.

The intuition here is that for a set of arguments to be accepted, we require that, if any one of them is challenged by a counterargument, then they offer grounds to challenge, in turn, the counterargument. There always exists at least one admissible set: The empty set is always admissible.

Clearly, the notion of an admissible set of arguments is the minimum requirement for a set of arguments to be accepted. We will focus on the following classes of acceptable arguments.

DEFINITION 5. Let Γ be a conflict-free set of arguments, and let $\text{Defended} : \wp(\mathcal{A}) \mapsto \wp(\mathcal{A})$ be a function such that $\text{Defended}(\mathcal{A}) = \{A \mid \Gamma \text{ defends } A\}$.

1. Γ is a **complete extension** iff $\Gamma = \text{Defended}(\Gamma)$
2. Γ is a **grounded extension** iff it is the minimal (w.r.t. set inclusion) complete extension.
3. Γ is a **preferred extension** iff it is a maximal (w.r.t. set inclusion) complete extension.

We illustrate these definitions with the running example. We classify each subset of the set of arguments according to the definitions as follows (where \times means that the classification holds).

	conflict free	admissible	complete	grounded	preferred
$\{\}$	\times	\times			
$\{A_1\}$	\times	\times			
$\{A_2\}$	\times				
$\{A_3\}$	\times	\times			
$\{A_1, A_2\}$					
$\{A_1, A_3\}$	\times	\times	\times	\times	\times
$\{A_2, A_3\}$					
$\{A_1, A_2, A_3\}$					

As another example, consider the situation where we have just two arguments A_4 and A_5 that attack each other. We classify each subset of arguments as follows.

	conflict free	admissible	complete	grounded	preferred
$\{\}$	\times	\times	\times	\times	
$\{A_4\}$	\times	\times	\times		\times
$\{A_5\}$	\times	\times	\times		\times
$\{A_4, A_5\}$					

As can be seen from the examples, the grounded extension provides a skeptical view on which arguments can be accepted, whereas each preferred extension take a credulous view on which arguments can be accepted.

The formalization we have reviewed in this section is abstract because both the nature of the arguments and the nature of the attack relation are ignored. In particular, the internal (logical) structure of each of the arguments is not made explicit. Nevertheless, Dung’s proposal for abstract argumentation is ideal for clearly representing arguments and counterarguments, and for intuitively determining which arguments should be accepted (depending on whether we want to take a credulous or skeptical perspective).

We harness abstract argumentation in our general framework for aggregating evidence. We will introduce mechanisms for generating arguments from the evidence, and for generating the attacks relation based on the preferences over the arguments. In this way, we will instantiate abstract argumentation with logical arguments. This means that we can use Dung’s definitions for determining which sets of arguments are acceptable, and thereby determine which aggregations of the evidence are acceptable.

4. ARGUMENTS FROM EVIDENCE

Here we present a general framework for evidence aggregation that involves constructing and comparing arguments from items of evidence where the evidence involves multiple outcome indicators.

We start with a set of evidence $EVIDENCE = \{e_1, \dots, e_n\}$. Each item in $EVIDENCE$ is a result from a study that compares a pair of treatments. We partition $EVIDENCE$ into three sets $SUPERIOR$, $EQUITABLE$, and $INFERIOR$. Those in $SUPERIOR$ are the trials for which τ_1 was shown to be superior to τ_2 with respect to some outcome indicator μ . By superior, we mean that if the outcome is desirable for the patient, then τ_1 is shown to be more efficacious for positive outcome than τ_2 , and if the outcome is undesirable for the patient, then τ_1 is shown to be less susceptible to this negative outcome than τ_2 . Similarly, those in $EQUITABLE$ are the trials for which τ_2 was shown to be equitable with τ_1 with respect to an outcome indicator μ , and those in $INFERIOR$ are the trials for which τ_2 was shown to be superior to τ_1 with respect to an outcome indicator μ .

Given treatments τ_1 and τ_2 , there are three possible interpretations of a set of items of evidence (i.e. a set of rows from an evidence table such as Table 1):

1. $\tau_1 > \tau_2$, meaning the evidence supports the claim that treatment τ_1 is superior to τ_2 .
2. $\tau_1 \sim \tau_2$, meaning the evidence supports the claim that treatment τ_1 is equivalent to τ_2
3. $\tau_1 < \tau_2$, meaning the evidence supports the claim that treatment τ_1 is inferior to τ_2 .

Any formula of the form $\tau_1 > \tau_2$, $\tau_1 \sim \tau_2$, and $\tau_1 < \tau_2$, we will call a **claim**, denoted by ϵ . We assume $\tau_1 > \tau_2$ is equivalent to $\tau_2 < \tau_1$, and $\tau_1 \sim \tau_2$ is equivalent to $\tau_2 \sim \tau_1$.

We use inference to derive a claim from a set of evidence. We use inference rules to define what are the allowed inferences. In this paper, we use three inference rules

DEFINITION 6. An **inference rule** is one of the following three forms, where $\Phi \subseteq EVIDENCE$.

1. If $\Phi \subseteq SUPERIOR$, then $\tau_1 > \tau_2$.
2. If $\Phi \subseteq EQUITABLE$, then $\tau_1 \sim \tau_2$.
3. If $\Phi \subseteq INFERIOR$, then $\tau_1 < \tau_2$.

For example, in the evidence given in Table 1, there is a subset $\{e_3, e_4\}$ of the evidence for which each items states that ACE is superior to CCB. From this subset, we may draw the conclusion that ACE is superior to CCB in general.

Given a set of results $EVIDENCE$ one can informally think of an argument comprising of a set of evidence (i.e. a subset of $EVIDENCE$), and a conclusion or claim that has been derived from the set of evidence using an inference rule.

DEFINITION 7. An **argument** is a tuple $\langle \Phi, \epsilon \rangle$ such that ϵ follows from Φ using one of the three inference rules given in Definition 6. We call Φ the support and ϵ the claim of the argument.

EXAMPLE 5. Returning to the evidence in Table 1, concerning treatments ACE and CCB, we have $EVIDENCE = \{e_1, e_2, e_3, e_4\}$, $SUPERIOR = \{e_3, e_4\}$, and $INFERIOR = \{e_1, e_2\}$.

From this, together with the inference rules, we get the following arguments with non-empty support.

$$\begin{array}{ll} \langle \{e_3\}, ACE > CCB \rangle & \langle \{e_1\}, ACE < CCB \rangle \\ \langle \{e_4\}, ACE > CCB \rangle & \langle \{e_2\}, ACE < CCB \rangle \\ \langle \{e_3, e_4\}, ACE > CCB \rangle & \langle \{e_1, e_2\}, ACE < CCB \rangle \end{array}$$

Informally, for instance $\langle \{e_1, e_2\}, ACE < CCB \rangle$ is an argument that says “based on evidence e_1 and e_2 , we can infer that ACE is inferior to CCB”.

In the above example, we see intuitively that the arguments with differing claims are in conflict. We capture this kind of conflict with the following definition.

DEFINITION 8. If $A = \langle \Phi_A, \epsilon_A \rangle$ and $B = \langle \Phi_B, \epsilon_B \rangle$ are two arguments then we say that A **conflicts** with B whenever:

1. $\epsilon_A = \tau_1 > \tau_2$, and $(\epsilon_B = \tau_1 \sim \tau_2$ or $\epsilon_B = \tau_1 < \tau_2)$.
2. $\epsilon_A = \tau_1 \sim \tau_2$, and $(\epsilon_B = \tau_1 > \tau_2$ or $\epsilon_B = \tau_1 < \tau_2)$.
3. $\epsilon_A = \tau_1 < \tau_2$, and $(\epsilon_B = \tau_1 > \tau_2$ or $\epsilon_B = \tau_1 \sim \tau_2)$.

Note that this definition is symmetric, i.e., if A conflicts with B then B conflicts with A .

EXAMPLE 6. Continuing with Example 5, there are a number of conflicts such as $\langle \{e_3, e_4\}, ACE > CCB \rangle$ conflicts with $\langle \{e_1, e_2\}, ACE < CCB \rangle$

We organize the arguments into a graph. To do this, we first consider the conflict relation given above. It is easy to see that the graph induced is tripartite, and its independent sets are given by those arguments with claim $\tau_1 > \tau_2$, those arguments with claim $\tau_1 \sim \tau_2$, and those arguments with claim $\tau_1 < \tau_2$.

EXAMPLE 7. Consider the following evidence table. From this, we get the argument graph below using the arguments with non-empty support.

	Outcome	Result	Type
e_{τ_1}	Mortality	$\tau_1 > \tau_2$	RCT
e_{τ_2}	Palpitations	$\tau_1 < \tau_2$	NA

$$\langle \{e_{\tau_1}\}, \tau_1 > \tau_2 \rangle \sqsubset \langle \{e_{\tau_2}\}, \tau_1 < \tau_2 \rangle$$

Since the argument graph is by definition symmetric (if we use the conflict relation), it would be beneficial to allow breaking the symmetry with user-defined preferences. We do this by defining preference rules.

DEFINITION 9. A **preference rule** is a condition on a pair of arguments $\langle \Phi_A, \epsilon_A \rangle, \langle \Phi_B, \epsilon_B \rangle$. When the condition is satisfied, A is said to be preferred to B , otherwise, we say that A is not preferred to B .

EXAMPLE 8. Consider the evidence from the table in example 7. If we want to express the fact that we prefer arguments based on meta-analyses, then we would write:

$$\langle \Phi_A, \epsilon_A \rangle \text{ is preferred to } \langle \Phi_B, \epsilon_B \rangle \text{ iff} \\ \text{for all } e_1 \in \Phi_A, \text{ the Type of } e_1 \text{ is MA} \\ \text{and not for all } e_2 \in \Phi_B, \text{ the Type of } e_2 \text{ is MA}$$

We use the preference rules chosen by the user in breaking the symmetry present in the conflict relation, and capture the attack relation as follows.

DEFINITION 10. *For any pair of arguments A and B , A attacks B iff A conflicts with B and A is preferred to B and it is not the case that B is preferred to A .*

The motivation here is that if A and B conflict with each other and A is preferred to B then B 's conflict with A is cancelled. However, this wording leads to problems when A is preferred to B according to a preference rule and B is preferred to A according to a preference rule. In this case, cancelling both attacks will give the misleading impression that A and B are consistent together. For this reason we give the above, more complicated definition, which only cancels an attack if exactly one argument is preferred to the other.

Now we combine these components by defining an argument graph based on a set of trial results, a set of inference rules, and a set of preference rules as follows.

DEFINITION 11. *Given a pair of treatments τ_1 and τ_2 , and a set EVIDENCE concerning these treatments, an **argument graph** is a graph where the set of nodes is the set of arguments formed by Definition 7 and the set of arcs is the attacks relation given by Definition 10.*

We illustrate the use of an informally defined preference rule to get the following argument graph by applying the preference rule to the arguments in our running example.

EXAMPLE 9. *Continuing Example 7, suppose we use the preference rule given in Example 8, then we get the following argument graph.*

$$\langle \{e_{71}\}, \tau_1 > \tau_2 \rangle \rightarrow \langle \{e_{72}\}, \tau_1 < \tau_2 \rangle$$

Compared to the argument graph in the previous example, we see that our preferences have resulted in there only being one remaining attack between the two arguments.

We can directly use the dialectical semantics (i.e. the definitions for acceptability such as preferred and grounded extensions) given by Dung [4] that we reviewed in the previous section to decide extensions of argument graphs. In Example 9, there is one grounded and preferred extension and it contains just the argument $\langle \{e_{71}\}, \tau_1 > \tau_2 \rangle$.

We regard an extension as an interpretation of a EVIDENCE (i.e. an aggregation of the evidence in EVIDENCE). So if X is an extension of the argument graph, and $A \in X$, and ϵ is the claim of A , then ϵ is a possible aggregation of the evidence. Furthermore, we regard a grounded extension as a higher quality interpretation than a preferred extension.

This section has provided a general framework for aggregating evidence concerning a pair of treatments according to multiple outcomes. To use the framework, a specific set of preference rules needs to be specified. In the next section, we consider a specific set of preference rules that are based on clinicians' preferences over evidence.

5. CLINICAL PILOT STUDY

In order to assess how well our system captures clinical requirements we conducted a small pilot study. We wanted to assess both the ability of our formalism to represent clinicians' preferences and to see how stable clinicians' preferences are in different circumstances.

Table 2: Details of the three sets of study summaries given to clinicians. DFS: Disease-free Survival, OS: Overall Survival, RCT: Randomized Controlled Trial, MA: Meta-analysis of RCTs, SRC: Systematic review of Cohort studies. See the section "Evidence for pilot study" for further information.

Set A					
Study	Outcome	Result	Setting	Size	Type
a_1	OS	$\tau_1 > \tau_2$	UK	Large	MA
a_2	OS	$\tau_2 > \tau_1$	USA	Small	RCT
a_3	DFS	$\tau_2 > \tau_1$	World	Large	MA
a_4	DFS	$\tau_1 > \tau_2$	UK	Small	SRC
a_5	OS	$\tau_2 > \tau_1$	World	Large	MA

Set B					
Study	Outcome	Result	Setting	Size	Type
b_1	OS	$\tau_2 > \tau_1$	UK	Large	MA
b_2	OS	$\tau_2 > \tau_1$	USA	Small	Cohort
b_3	DFS	$\tau_1 > \tau_2$	World	Large	MA
b_4	DFS	$\tau_1 > \tau_2$	UK	Small	SRC
b_5	DFS	$\tau_1 > \tau_2$	World	Large	MA

Set C					
Study	Outcome	Result	Setting	Size	Type
c_1	OS	$\tau_1 > \tau_2$	USA	Large	MA
c_2	OS	$\tau_2 > \tau_1$	USA	Small	RCT
c_3	DFS	$\tau_2 > \tau_1$	World	Large	MA
c_4	DFS	$\tau_2 > \tau_1$	UK	Large	SRC
c_5	OS	$\tau_2 > \tau_1$	USA	Large	SRC

Our presentation of the pilot study proceeds as follows: (1) We describe the evidence tables we used for the pilot study; (2) We introduce the notion of expressed preference schemes as a simpler way for clinicians to represent their preferences over evidence; (3) We describe the methods we used for the pilot study; (4) We explain how we translated the preferences given by the clinicians into preference rules; (5) We present the results we obtained from the pilot study; and (6) We discuss the pilot study and its implications for our general framework.

Evidence for pilot study

We used three different sets of evidence (A, B, C), with five items of evidence in each set.

Each item of evidence gave information on the patient class (uniform across all trials), outcome indicator (Overall Survival (OS) or Disease-free Survival (DFS)), which of two treatments was found to be superior, the geographical area the trial was set in, the size of the study population and the type of study. Each set of trials explored different interactions: A and B concentrated on the clinical preferences for

OS vs. DFS, while C looked at the impact of the country and the type of study.

The details of the sets of trials are given in Table 2. The columns are as follows:

- The **Outcome** refers to the outcome measure used in the study, and in this case is restricted to either Disease-free Survival or Overall Survival.
- The **Result** summarizes which treatment, τ_1 or τ_2 was found to have greater impact on the outcome measure, as described earlier in Section 2.
- The **Setting** is the geographical location from which most of the patients were drawn. Here, we assume that all patients come from either UK, USA, or a (larger) World population.
- The **Size** is a measure of the study population size, where the dividing line between the two is calculated in terms of the statistical power of the study (based on sample size).
- The **Type** specifies the type of study, as described earlier in Section 2.

For instance, consider the study a_1 that is summarized in Table 2. We can see that this study measured overall survival (OS) as an outcome measure, showed that treatment τ_1 was more beneficial than treatment τ_2 (wrt OS), was based on patients from the UK, was a large study and was a meta-analysis of randomized controlled trials.

Expressed preference schemes

Before we explain how we conducted the pilot study, we introduce the notion of expressed preference schemes. None of the clinicians participating in the pilot study had familiarity with formal logic, and so we required a simple way for each clinician to represent their preferences. For this we introduced the notion of an **expressed preference scheme** (EPS). This is less expressive than preference rules, but for the pilot study they were better suited to capture each clinician’s preferences for each set of evidence.

Essentially, an EPS is a list of values v_1, \dots, v_n that may appear in the evidence table. The meaning of this list is that a set of evidence X is preferred over a set of evidence Y if

- all evidence in X has value v_1 but not all evidence in Y has value v_1 ;
- or all evidence in X has value v_1 and all evidence in Y has value v_1 and all evidence in X has value v_2 but not all evidence in Y has value v_2 ;
- or all evidence in X has value v_1 and all evidence in Y has value v_1 and all evidence in X has value v_2 and all evidence in Y has value v_2 and all evidence in X has value v_3 but not all evidence in Y has value v_3 ;
- or all evidence in X has value v_1 and all evidence in Y has value v_1 and ... and all evidence in X has value v_i and all evidence in Y has value v_i and all evidence in X has value v_{i+1} but not all evidence in Y has value v_{i+1} .

Table 3: Details of the 8 different Expressed Preference Schemes (EPS). See section on “Expressed preference schemes” for details of how to interpret each row and Table 4 for the distribution in the use of EPS by each clinician.

EPS	Criteria for Preferred Studies
1	Large, MA, World
2	OS, MA, Large
3	OS
4	Large, MA, World \equiv UK
5	OS, Large, MA, USA \equiv UK
6	Number of MA
7	MA
8	DFS, MA, World

In other words, the values given in an EPS are used in a lexicographic order. So if both sets of evidence, v_1, \dots, v_i appear in each item of evidence in each set (i.e. the two sets of evidence cannot be differentiated on values v_1, \dots, v_i , then consider v_{i+1} to differentiate them. However, if v_1, \dots, v_n appear in each item of evidence in each set, then they cannot be differentiated (i.e. they are equally preferred).

We illustrate the meaning of some of the EPS in the following examples.

EXAMPLE 10. Consider EPS 3 in Table 3. Here, there is just one value which is OS. So a set of evidence X is preferred to Y , if all items in X have value OS and not all items in Y have value OS. So if we consider Set A in Table 2, then $\{a_1, a_2, a_5\}$ is preferred to $\{a_1, a_3, a_4\}$.

EXAMPLE 11. Consider EPS 1 in Table 3. Here, there are three values which are Large, MA, and World. So if we consider Set A in Table 2, then $\{a_3, a_5\}$ is preferred to $\{a_1\}$: For both sets, each item has the value Large and the value MA, but only $\{a_3, a_5\}$ has the value World for each item of evidence.

There are two additional refinements we use for defining an EPS. First, two or more values can be regarded as equivalent. For instance, in EPS 5, we have USA \equiv UK. So when using EPS 5, if both sets of evidence have the values OS, Large, and MA for each item of evidence, then we need to try to use the fourth value to differentiate the sets. So suppose for one of the sets each item of evidence has either USA or UK, and the other does not, then the first set is preferred over the second.

EXAMPLE 12. Consider the EPS with the value USA \equiv UK. So for Set A in Table 2, then $\{a_1, a_2, a_4\}$ is preferred to $\{a_1, a_3, a_5\}$ since the first set has either the value USA or UK in each item of evidence whereas the second set does not.

The second refinement is that we can compare sets of evidence on the number of items of a particular value. For instance, in EPS 6, the criterion is the number of items with

value MA. So for two sets of evidence X and Y , X is preferred to Y , if all items of evidence in X have value MA and not all items of evidence in Y have value MA, or all items of evidence in X and in Y have value MA but X has more items of evidence than Y .

We explain how we obtained the EPSs in Table 3 in the next section and the explain how we formalized the preference rules from the EPSs in the subsequent section.

Methods for pilot study

Each clinician was presented with a written sheet giving a brief background to the exercise and explaining our motivation, and presenting three sets of evidence, where each set contained summary details on five different studies. These sets of evidence are presented in Table 2. We used abstract treatments (τ_1, τ_2) rather than real treatment names in order to avoid clinicians using pre-existing knowledge about treatment efficacy while deciding on preferences.

For each set of evidence, each clinician was asked to decide on which treatment they would recommend, and to explain their reasoning as to weighing up the differing evidence from the studies. Answers as to treatment choice and reason were recorded by one of the investigators (MW), and checked at the end of the exercise. Any uncertainties in the data collected were clarified within 3 days of the first meeting.

From these interviews with clinicians, we generated some EPSs to summarize the way in which each clinician explained their choices of treatment. We then formalized each EPS as a preference rule. We explain how to do this in the next section.

We then used the evidence in each set (i.e. the sets in Table 2) to generate an argument graph for each clinician using the preference rule corresponding to their choice of EPS. We compared the arguments contained in the extensions of the argument graphs to the actual treatment choices made by the clinicians in order to assess the ability of our argumentation system to capture the behavior of the clinicians. We also explored the significance of different EPS by assessing when the EPSs produced different results with the evidence in each set. We explain in detail how we did this in the results section.

Generating preference rules

As explained above, each clinician's preferences were represented as an EPS. As we will explain in the results section, there were eight EPSs identified in the pilot study as represented in Table 3. In order, to use each EPS with our argumentation system, we need to generate a preference rule for each EPS. Here, we show how this is done.

In order to represent an EPS as a preference rule, we require the following subsidiary definition which says when one argument is better than another based on an individual criterion.

DEFINITION 12. *Let $\langle \Phi_A, \epsilon_A \rangle$ and $\langle \Phi_B, \epsilon_B \rangle$ be arguments, let v , v_1 , and v_2 be values occurring in the evidence table.*

- $\langle \Phi_A, \epsilon_A \rangle$ is **superior** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. v iff for all e in Φ_A , there is an attribute with value v in e and there exists e' in Φ_B where no attribute in e' has value v .
- $\langle \Phi_A, \epsilon_A \rangle$ is **superior** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. $v_1 \equiv v_2$ iff for all e in Φ_A , there is an attribute with value v_1 or v_2

in e and there exists e' in Φ_B where no attribute in e' has value v_1 or v_2 .

- $\langle \Phi_A, \epsilon_A \rangle$ is **superior** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. number of v iff (for all e in Φ_A , there is an attribute with value v in e and there exists e' in Φ_B where no attribute in e' has value v) or (for all e in Φ_A , there is an attribute with value v in e and for all e' in Φ_B , there is an attribute with value v in e' and the cardinality of Φ_A is greater than the cardinality of Φ_B).
- $\langle \Phi_A, \epsilon_A \rangle$ is **equal** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. value v iff for all e in Φ_A , there is an attribute with value v in e and for all e' in Φ_B , there is an attribute with value v in e' .
- $\langle \Phi_A, \epsilon_A \rangle$ is **equal** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. $v_1 \equiv v_2$ iff for all e in Φ_A , there is an attribute with value v_1 or v_2 in e and for all e' in Φ_B , there is an attribute with value v_1 or v_2 in e' .
- $\langle \Phi_A, \epsilon_A \rangle$ is **equal** to $\langle \Phi_B, \epsilon_B \rangle$ w.r.t. number of v iff for all e in Φ_A , there is an attribute with value v in e and for all e' in Φ_B , there is an attribute with value v in e' and the cardinality of Φ_A is equal to the cardinality of Φ_B .

The above subsidiary definitions allow us to directly formalize each EPS as a preference rule as follows where A and B are arguments, and EPS 1 is captured by EPS1', EPS 2 is captured by EPS2', etc.

EPS1' A is preferred to B iff A is superior to B wrt Large or (A is equal to B wrt Large, and A is superior to B wrt MA) or (A is equal to B wrt Large and MA, and A is superior to B wrt World).

EPS2' A is preferred to B iff A is superior to B wrt OS or (A is equal to B wrt OS, and A is superior to B wrt MA) or (A is equal to B wrt OS and MA, and A is superior to B wrt Large).

EPS3' A is preferred to B iff A is superior to B wrt OS.

EPS4' A is preferred to B iff A is superior to B wrt Large or (A is equal to B wrt Large, and A is superior to B wrt MA) or (A is equal to B wrt Large and MA, and A is superior to B wrt World \equiv UK).

EPS5' A is preferred to B iff A is superior to B wrt OS or (A is equal to B wrt OS, and A is superior to B wrt Large) or (A is equal to B wrt OS and Large, and A is superior to B wrt MA) or (A is equal to B wrt OS, Large, and MA, and A is superior to B wrt USA \equiv UK).

EPS6' A is preferred to B iff A is superior to B wrt the number of MA.

EPS7' A is preferred to B iff A is superior to B wrt MA.

EPS8' A is preferred to B iff A is superior to B wrt DFS or (A is equal to B wrt DFS, and A is superior to B wrt MA) or (A is equal to B wrt DFS and MA, and A is superior to B wrt World).

Table 4: Details different Expressed Preference Schemes (EPS) by clinician and trial. Details of individual EPS are given in Table 3. So for instance, for evidence set A, clinician RP used EPS 2.

Clinician	Set A	Set B	Set C
NG	1	2	3
GD	4	2	5
DW	6	7	2
RP	2	2	2
DKW	8	8	8
CG	2	3	2

So each EPS, which represents a preference over sets of evidence, corresponds directly to a preference rule over arguments, and this correspondence is based on the evidence used as support in the arguments. Whilst we have only considered a preference rule for each EPS arising in the pilot study, it is possible to capture a far wider range of possible EPSs as preference rules using this approach.

Results from pilot study

From the interviews with our six clinicians, we obtained eight EPSs which are presented in Table 3. The use of each EPS by clinician and set of studies is detailed in Table 4. Only two clinicians (RP and DKW) used the same EPS on all three sets of evidence. The other four clinicians used a different EPS for some of the sets of evidence. It is also interesting to note that some EPSs were used at least once by the majority of clinicians. For instance, EPS 2 was used by five clinicians.

As explained in the previous section, each EPS has been represented by a preference rule. This means that we can then use our argumentation system to automatically aggregate the evidence, and then compare the result with aggregation by each clinician for each set of evidence.

EXAMPLE 13. *Consider the set of evidence given by Set A in Table 2. From this, we can generate the following arguments. Note that the arguments in the left column conflict with all of those in the right in a symmetrical fashion.*

$$\begin{array}{l}
 \langle \{a_1\}, \tau_1 > \tau_2 \rangle \\
 \langle \{a_4\}, \tau_1 > \tau_2 \rangle \\
 \langle \{a_1, a_4\}, \tau_1 > \tau_2 \rangle
 \end{array}
 \begin{array}{l}
 \langle \{a_2\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_3\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_5\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_2, a_3\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_3, a_5\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_2, a_5\}, \tau_1 < \tau_2 \rangle \\
 \langle \{a_2, a_3, a_5\}, \tau_1 < \tau_2 \rangle
 \end{array}$$

Now, if we apply either EPS2' or EPS4', then we get two preferred extensions: One extension with the left hand arguments and the other extension with the right hand arguments. Whereas, if we apply EPS1', or EPS6', or EPS8', then we get one grounded extension which contains the right hand arguments. This means if we use EPS2' or EPS4', then τ_1 and τ_2 are equally preferred (because we have $\tau_1 > \tau_2$ as the claim in one preferred extension and $\tau_1 < \tau_2$ as the claim in the other preferred extension) whereas if we apply EPS1', or EPS6', or EPS8', then τ_2 is preferred over τ_1 (i.e.

$\tau_1 < \tau_2$ is the claim of the arguments in the grounded extension). Furthermore, when there is a grounded extension, we can select an argument as a representative for the winning claim by choosing an unattacked argument with maximal support (i.e. $\langle \{a_2, a_3\}, \tau_1 < \tau_2 \rangle$) and thereby provide a concise rationale for the inference that τ_2 is preferred over τ_1 . We can also provide the counterargument with maximal support (i.e. $\langle \{a_1, a_4\}, \tau_1 > \tau_2 \rangle$) to summarize the conflict.

We systematically considered each EPS used by the clinicians for each set of evidence. In Table 5, we give the result of comparing what our argumentation systems stated as the preferred treatment with what with the clinician stated as the preferred treatment. In this table, we see that for instance, the clinician NG used EPS 1 on the evidence set A to obtain τ_2 as the preferred treatment, and our argumentation used the preference rule corresponding to EPS 1 (i.e. EPS1') to generate τ_2 as the preferred treatment using the evidence a_3 and a_5 .

Using Table 5, we can measure the precision of our argumentation system. For each clinician and each evidence set, we have the clinician's preferred treatment, and we have the preferred treatment from our argumentation. We say there is agreement if the clinician's preferred treatment agrees with that given by the argumentation system. So there is agreement if both say τ_1 , or if both say τ_2 , or if both say τ_1, τ_2 , otherwise there is disagreement. According to this measure, we see that in 14 out of 18 cases (where each case is a clinician and an evidence set), there is agreement. For the 4 out of 18 cases where there is disagreement, it is where the argumentation system is more sceptical and so chooses τ_1, τ_2 , whereas the clinician had chosen one of τ_1 or τ_2 .

Discussion of pilot study

Although the sets of evidence presented in the pilot study are artificial, we note that the data in Table 2 are very similar to those we have obtained from NCI and NICE Guidelines.

The formalism for the preference rules presented here is deliberately simple. Despite that, we were able to capture all eight of the preference criteria used by the clinicians.

Given the three sets of trials and six clinicians, there were 18 possible points on which to assess agreement between our approach and the clinician's choices. Of these 18, there was agreement on 14. The areas of disagreement occurred because the clinicians' preference criteria, each represented by an EPS, allowed different preferred extensions containing arguments with two different claims, whereas they chose only one claim. One reason for this is that although we asked the clinicians for their treatment choices and to explain their reasoning, we did not feedback the results of the reasoning to them; thus, for example, when clinician DW chose τ_1 in Set B and explained this with reference to EPS 7, we did not demonstrate that using EPS 7 (which only considers whether a study is a MA) with Set B leads to two possible outcomes, based on either b_1 or b_3 and b_5 . If we were able to provide that feedback from our argumentation system, then it would be possible for the clinician to revise their choice of preferred treatment based on the evidence or to use another EPS.

A question raised by this work is to what extent the different EPS actually represent different preference strategies, especially in terms of the preferred treatment output by the system and the evidence used to support that. We explored

Table 5: Comparison of reasoning by clinicians and by argumentation system. Clinician’s Choice is the clinician’s preferred treatment (and if the clinician cannot choose, then both possibilities are given separated by a comma), Support refers to the items of evidence used in the unattacked arguments in the grounded extension and System Choice is the preferred treatment given by the system if it is a grounded extension (i.e. it is τ_i if the claim of the arguments in the grounded extension are such that τ_i is preferred to τ_j), otherwise both possibilities are given separated by a comma.

EPS	Clinician	Clinician’s Choice	Support	System Choice
Set A				
1	NG	τ_2	a_3, a_5	τ_2
2	CG/RP	τ_1	-	τ_1, τ_2
4	GD	τ_1, τ_2	-	τ_1, τ_2
6	DW	τ_2	a_3, a_5	τ_2
8	DKW	τ_2	a_3	τ_2
Set B				
2	NG/GD/RP	τ_2	b_1	τ_2
3	CG	τ_2	b_1, b_2	τ_2
7	DW	τ_1	-	τ_1, τ_2
8	DKW	τ_1	b_3, b_5	τ_1
Set C				
2	DW/RP/CG	τ_1	c_1	τ_1
3	NG	τ_2	-	τ_1, τ_2
5	GD	τ_1	c_1	τ_1
8	DKW	τ_2	c_3	τ_2

this by identifying when different EPSs used the same evidence to draw the same conclusion. This only occurred twice. First, for Set A, when EPS 1 and EPS 6 were used to choose τ_2 with evidence a_3 and a_5 , and second, for Set C, when EPS 2 and EPS 5 were used to choose τ_1 with evidence c_1 . Otherwise, the EPSs used different evidence even if the treatment chosen was the same. This suggests that the clinicians were drawing non-trivial distinctions on how to interpret the evidence. Understanding the sensitivity of the system to different EPSs, and how this should be presented to the user remains an area for further investigation.

Finally, in this pilot study, we have limited the kinds of argument that can be presented in order to focus on the preferences used by clinicians. However, in practice we would need to use more complex evidence tables that take into account the patient class and involve more than two treatments. This would mean that the set of arguments would be more diverse, and hence the argument graphs would be more complex. Nonetheless, we believe that based on the experience from this pilot study, the use of preference criteria from clinicians can be effectively harnessed for these more complex evidence tables.

6. ARGUMENTATION IN PRACTICE

The input to our argumentation system is a table of evidence, and we use simple inference rules to generate arguments. These arguments justify choices about treatments, and we use preference criteria from clinicians to modify the way in which the arguments interact. The output of the system is a grounded or preferred extension with the arguments showing how the evidence used to derive the claim. As such, our argumentation system provides on-the-fly aggregation of clinical evidence to provide evidence-based recommendations. Furthermore, the evidence is organized in terms of arguments and counterarguments that highlights points of agreement and conflict.

In principle our argumentation system could be implemented using existing, well described methodologies and technologies: for example, evidence tables (in either the form used in this paper or as Forest plots) are increasingly common in the medical literature, and there are existing medical ontologies [13, 15] that we could use to allow for further reasoning and abstraction over treatment types. Our approach to capturing and representing clinical preferences over evidence types is simple, and in our pilot study it was easy for the clinicians to use them. The algorithm for generating arguments and the argument graph is simple, and there are existing argumentation engines [16, 6] that we could use to calculate the different extensions. The key output for clinicians is a summary of the evidence in terms of key arguments, although it remains an open question to what extent different extensions of argument frameworks coincide with clinically relevant sets of evidence. Although the argument graph structure is simple, we are also developing techniques to output the same information as textual summaries, which may be more appropriate for clinical users.

In an implemented version of our argumentation system we could include a range of in-built EPS, or allow the user to add their own. Allowing the user to experiment with different EPS and different subsets of evidence would enable them to explore how sensitive their decision-making was to both preferences and the influence of particular studies. They could also investigate the EPS selected by other users to explore different attitudes to aggregating evidence.

7. CONCLUSIONS

The problem of conflicting information is a general issue in handling knowledge and it arises in virtually all real-world domains. It is certainly a key problem in analyzing results from clinical trials. It is common for different trials of the same treatments to have conflicting results. Yet in order to use the results of clinical trials, it is necessary to harness techniques that handle such inconsistent information by highlighting important conflicts and suppressing unimportant conflicts.

Argument systems aim to reflect how human argumentation uses conflicting information to construct and analyze arguments. There is a number of frameworks for computational models of argumentation. They incorporate a formal representation of individual arguments and techniques for comparing conflicting arguments (for reviews see [2, 3]).

In this paper, we have drawn on argumentation techniques (in particular influenced by assumption-based argumentation [5]) to provide a general framework for taking evidence involving multiple outcome indicators and aggregate it in

terms of arguments. In this framework, we instantiate abstract argument graphs with arguments generated by inference rules applied to the evidence, and attacks relationships obtained via the preference rules. For any application of our framework, a specific set of inference rules and preference rules needs to be given. Our pilot study used realistic examples of clinical trial summaries to evoke preferences from clinicians. We have then shown how these summaries can be used to generate arguments, and how we can capture the clinicians' preferences formally in order to determine preferences between arguments. Separate to this, in [12], we have undertaken a case study with evidence taken from 21 meta-analyses concerning 5 treatment options for raised intraocular pressure (raised IOP) and showed that the results we obtained corresponded closely with those presented in the NICE Guideline for Glaucoma. Taken together, these two papers indicate the potential viability of using argument-based techniques for aggregating clinical evidence.

Little work exists that aims to address the problem in focus here. Medical informatics and bioinformatics research rarely address the issues inherent in the analysis of primary evidence, especially from clinical trials. Previous interesting work ([7, 14, 17, 18] and others) exists that uses argumentation as a tool in medical decision support, but starts with a set of hand-crafted arguments. Our work is distinct in that we start with a set of primary evidence and generate the arguments from it.

In future work, we plan to explore the space of preference rules in more detail. In particular, we need to undertake further interviews with clinicians to better understand the preference criteria they use for evidence aggregation. We also would like to explore the robustness of evidence aggregation in the presence of new evidence. In addition, we plan to develop a prototype system (as suggested in the previous section), and investigate its appropriateness for clinical use.

8. ACKNOWLEDGMENTS

We would like to thank Nikos Gkrogiannis and Vivek Patkar for their helpful discussion on this work, and the six clinicians who participated in the pilot study.

9. REFERENCES

- [1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, New York, NY, USA, 2003.
- [2] T. Bench-Capon and P. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.
- [3] Ph.. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.
- [4] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
- [5] P. Dung, R. Kowalski, and F. Toni. Dialectical proof procedures for assumption-based admissible argumentation. *Artificial Intelligence*, 170:114–159, 2006.
- [6] U. Egly, S. A. Gaggl, and S. Woltran. Aspartix: Implementing argumentation frameworks using answer-set programming. In *ICLP*, pages 734–738, 2008.
- [7] J. Fox and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. MIT Press, 2000.
- [8] N. Gorogiannis, A. Hunter, V. Patkar, and M. Williams. Argumentation about treatment efficacy. In *Knowledge Representation for Healthcare (KR4HC)*, volume 5943 of *LNCS*, pages 169–179. Springer, 2010.
- [9] N. Gorogiannis, A. Hunter, and M. Williams. An argument-based approach to reasoning with clinical knowledge. *International Journal of Approximate Reasoning*, 51(1):1 – 22, 2009.
- [10] G. Guyatt, A. Oxman, G. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coelle, and H. Schünemann. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336(7650):924–926, 2008.
- [11] A. Hackshaw. *A Concise Guide to Clinical Trials*. WileyBlackwell, 2009.
- [12] A. Hunter and M. Williams. Argumentation for aggregating clinical evidence. In *Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2010)*. IEEE Computer Society Press, 2010.
- [13] I. H. T. S. D. Organisation. Systematized Nomenclature of Medicine-Clinical Terms.
- [14] V. Patkar, C. Hurt, R. Steele, S. Love, A. Purushotham, M. Williams, R. Thomson, and J. Fox. Evidence-based guidelines and decision support services: adiscussion and evaluation in triple assessment of suspected breast cancer. *British Journal of Cancer*, 95(11):1490–1496, 2006.
- [15] N. Sioutos, S. de Coronado, M. W. Haber, F. W. Hartel, W. L. Shaiu, and L. W. Wright. Nci thesaurus: a semantic model integrating cancer-related clinical and molecular information. *Journal of biomedical informatics*, 40(1):30–43, February 2007.
- [16] M. South, G Vreeswijk, and J Fox. Dungine: A Java Dung Reasoner. In *Computational Models of Argument (COMMA 2008)*, pages 360–369. IOS Press Press, 2008.
- [17] P. Tolchinsky, U. Cortés, S. Modgil, and F. C. A. López-Navidad. Increasing human-organ transplant availability: argumentation-based agent deliberation. *IEEE Intelligent Systems*, 21(6):30–37, 2006.
- [18] R. Walton, C. Gierl, P. Yudkin, H. Mistry, M. Vessey, and J. Fox. Evaluation of computer support for prescribing (CAPSULE). *British Medical Journal*, 315:791–795, 1997.
- [19] M. Williams and A. Hunter. Harnessing ontologies for argument-based decision-making in breast cancer. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, volume 2, pages 254–261. IEEE Computer Society Press, 2007.