Temporal Knowledge Representation for Scheduling Tasks in Clinical Trial Protocols

Chunhua Weng, M.S.,[1] Michael Kahn, M.D., Ph.D.,[2] & John Gennari, Ph.D.[1]
[1]Biomedical and Health Informatics, University of Washington
Seattle, WA, 98195-7240 <{cweng, gennari}@u.washington.edu>
[2]Fast Track Systems, Inc. San Mateo, CA. <mkahn@fast-track.com>

Abstract

Clinical trial protocols include detailed *temporal* constraints on treatment and associated tasks. Unlike health-care guidelines, protocols are highly prescriptive. Therefore, informatics applications that enforce such temporal constraints are more directly useful with protocols than with guidelines. Although there are some temporal knowledge representation efforts for health-care guidelines, we find these to be insufficiently expressive for clinical trial protocols. In this paper, we focus on temporal knowledge representation for clinical trial protocols and the task of patient-specific scheduling in protocols. We define a temporal ontology, use it to encode clinical trial protocols, and describe a prototype tool to carry out patient-specific scheduling for the tasks in protocols. We predict that an expressive temporal knowledge representation can support a number of scheduling and management tasks for protocol-based care.

Motivation

Clinical trial protocols are the defining documents used in medical research. They provide detailed plans and temporal constraints that health-care providers must follow, so that statistically valid data are collected about the efficacy and safety of a particular treatment regimen or new agent. Unfortunately, clinical trial protocols are currently disseminated and managed in a paper-based manner, which can lead to errors and inefficiencies in interpretation and execution. Thus, one focus of medical informatics research is to automate and make electronic and computational the elements of guidelines and protocols.

However, although good progress has been made in the domain of standard knowledge representations for guidelines,[1-3] knowledge representation specifically for clinical trial protocols remains largely ignored. We argue that there are crucial differences between health-care guidelines and clinical trial protocols, and that there are scheduling applications for protocols that require sophisticated knowledge modeling of *temporal* information.

Protocol scheduling tasks

When a particular individual enrolls in a protocol, care providers need an appropriate schedule to carry out the protocol tasks across a set of patient visits. One study shows that clinicians have a strong desire for help with scheduling and co-ordination of interventions during maintenance therapy for acute lymphoblastic leukemia.[4] During the course of patient care, this schedule might need to be adjusted dynamically based on changes in the patient state. Large multi-site protocols describe procedures and schedules for patient care in a site-independent manner, and this makes them inappropriate for direct use in scheduling visits for a particular patient. The paper-based manual process of generating visit schedules for patients tends to be time-consuming and error-prone. Worse, some protocols may contain either ambiguity or inconsistency about their temporal requirements across different sections in the protocol.[5] Regardless of the source of errors, inappropriate schedules can decrease the quality of protocol data, lengthen the duration of the study, and hamper effective patient management.

By definition, scheduling involves the specification and application of temporal information about tasks. If we could precisely specify and correctly apply temporal information in clinical trial protocols, we would have the potential to greatly improve the quality and cost-effectiveness of protocol-based care. In this paper, we describe the temporal knowledge requirements for such a model, and present both a prototype model that conforms to those requirements and an example of a scheduler that uses this model.

Temporal information in protocols

In Table 1, we provide some examples of protocol text that convey temporal information. These examples are drawn from a review of dozens of phase 2 and phase 3 protocols that were sponsored either by pharmaceutical companies or by cooperative groups including the Gynecologic Oncology Group (GOG), the Southwest Oncology Group (SWOG), and the National Surgical Adjuvant Breast and Bowel Project (NSABP).

As Table 1 shows, temporal constraints in clinical trial protocols are expressed at a much finer level of detail than is typical for health-care guidelines. Temporal information in guidelines is often vague and approximate. In contrast, the temporal constraints in

Context within protocol	*Textual examples that include temporal knowledge*
1. Overall schedule of the study in the schema of a protocol	"External Beam, Whole Pelvis Radiation Therapy, 180 cGy daily fraction Monday through Friday x 25 days."
2. Time constraints in eligibility criteria	"Measurable Disease: must have measurable tumor by MRI that has not received radiation within 6 weeks prior to study entry."
3. Dose schedule and duration	"Concurrent Weekly Cisplatin and Paclitaxel on Mondays Cisplatin 40mg/m2/week x 6 – (70 mg max) Paclitaxel 30 mg/m2/week x 6 – (Starting dose) "
4. Dose modification schedule (which depends on adverse event monitoring schedule)	"Administration of AC will be delayed on day 1 if granulocytes < 1500; resume treatment if granulocytes >= 1500 ".
5. Treatment schedule	"Arm 1, dexamethasone at 20mg/dose, q 6 hours * 3 doses at day #1 and day #2 in q 21day cycle for 12 cycles in maximum".
6. Study calendar (1)	"Chest X-ray will done at 6 day +/- 1, 12 day +/- 1, 18 day +/- 1, etc during the treatment period".
7. Study calendar (2)	CBC/Differential/Platelets test is scheduled on Weeks 4, 7 and 10 during treatment period.
8. Assessment schedule	"Clinical assessment will be done while On-study, q 4 month * 1 yr, q 6 month * 2 yr, and q year after that".

Table 1. Temporal knowledge examples in clinical trial protocols.

Table 1 are rigid, and allow for little variability—in most situations, when protocol constraints are violated; the corresponding clinical data is invalid for use. For these reasons, work in standard knowledge representation for health care guidelines (e.g., GLIF)[3] does not include sufficient detail for automating protocol scheduling.

Therefore, we take a broader look at the knowledge-level requirements of a model for temporal information in protocols. We expect that if temporal knowledge is modeled well, then this model could be used by a set of different decision-support tools for use in the clinical trial protocol lifecycle. Although we are especially motivated by the patient scheduling task presented above, we try to anticipate the temporal modeling needs for other tasks associated with protocol-based care and management.

Protocol requirements for temporal KR

From a knowledge representation standpoint, it is useful to view clinical trial protocols as examples of a *plan*.[6] Viewed in this manner, all of the temporal specifications in various sections of the protocol (the dosing schedule, the treatment schedule, the study calendar, etc.) simply define which activities occur when, under what conditions. Based on our review of protocols, we identified the following knowledge-level requirements for a temporal model that is sufficiently expressive to support patient scheduling tasks in clinical trial protocols.

Represent relative time information Temporal information about clinical data is often naturally expressed in relative terms.[7] In Table 1, row 5, the phrases "day 1" and "day 2" are relative to each cycle's start time. By using relative time, we can provide useful context information for clinical data. In clinical trial protocols, there is a great deal of temporal inter-dependence of activities. Absolute time stamps do not capture the temporal relationships between activities, nor are they intuitive for clinical trial protocol authors. Ultimately, to produce a patient schedule, we need to provide absolute time points. However, temporal information in the protocol, is best expressed relative to a few critical "anchor points", such as the start of treatment. Given the notion of an "anchored time point", and a representation of relative time, we can compute the actual occurrence of an activity once the anchor point has an absolute time stamp for a particular patient.

Handle temporal indeterminacy Although temporal constraints in clinical trial protocols are prescriptive and rigorously defined, they also allow for some flexibility. A knowledge representation for the temporal specification of protocol events must include indeterminacy.[5] However, in order to enforce adherence to the protocol, this indeterminacy must be specified precisely.

There are several types of indeterminacy. First, the occurrence of an activity might allow for flexibility. For example, in Table 1, row 6, the patient's visits have an indeterminacy expressed as "+/- 1 day". We can capture this type of indeterminacy with concepts such as "expected occurrence", and "earliest/latest occurrence". Second, the duration of a temporal interval in the protocol may be indeterminate. For example, a treatment period may have an expected du-

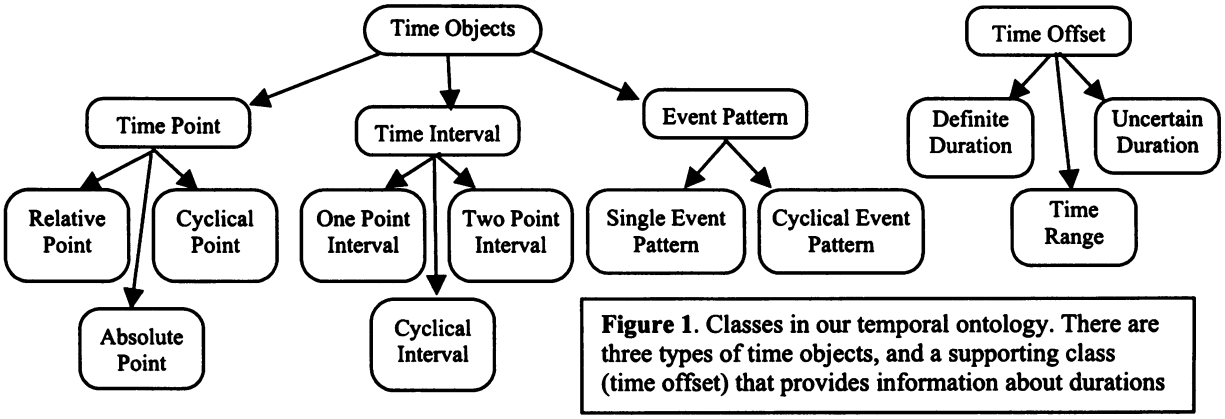

Figure 1. Classes in our temporal ontology. There are three types of time objects, and a supporting class (time offset) that provides information about durations

ration of 5 weeks, but may be delayed due to patient toxicities to become 6 weeks or more. Finally, an entire sequence of events may occur at an indeterminate time. For example, protocols may have a schedule of tasks (such as an exit interview, or final disease assessment) associated with protocol termination. These events may be triggered at an unknown time—patients are always allowed to leave for any reason, or they may be forced to leave the protocol due to disease progression or toxic reactions. In spite of their indeterminacy, once such events are triggered (and an anchor point is established), the protocol may specify temporal constraints on a set of exit tasks.

Represent cyclical event patterns It is necessary to represent cyclical event patterns simply because there are many recurrent temporal activities in clinical trial protocols. In Table 1, rows 5 and 8 are examples of explicit cyclical activities. Sometimes cyclical patterns may be implicit. For example, row 7 of Table 1 might be interpreted as a cyclical pattern of testing, namely, "every 3 weeks, starting with week #4 of treatment". When generating patient-specific schedules, a significant task is understanding and "unrolling" these cyclical patterns. Therefore, representations for cyclical patterns should support efficient reasoning over such patterns.

Represent both time points and time intervals In scheduling tasks for clinical trial protocols; there is usually a large gap between the temporal granularities of an activity's duration and the intervals between activities. Therefore, usually it is important to decide when a task happens, and it is not as important to know how long a task lasts. For example, a protocol might specify "Patients should have a physical examination every 2 weeks during the treatment period". The duration of the physical examination is brief compared to 2 weeks or the duration of the treatment period. Therefore, using time points to represent tasks or activities as instantaneous objects is appropriate and can simplify calculations. On the other hand, clinical studies are usually divided into different periods such as "pre-study", "treatment period" and "follow-up". These longer intervals are used to define the duration of some activity patterns. Thus, it is also important to represent time intervals.

A Temporal Ontology

We have built an example of a temporal ontology that satisfies the above requirements and can be used to support scheduling tasks for clinical trial protocol applications. Our ontology includes the temporal concepts and distinctions necessary for time-related decision-support applications for clinical trial protocols. Although we cannot know *a priori* what an arbitrary application might need, our model at least satisfies the requirements of prototype systems we have built to date. We constructed our ontology with the Protégé-2000 environment,[8] and in conjunction with a more general model for clinical trial protocols. Figure 1 shows the major classes in our temporal ontology: time point, time interval, and event pattern.

A time point is an instantaneous object, without duration. An absolute time point is associated with an absolute calendar date. We use two features, "anchored time point" and "offset from the anchor" to define relative time information for "Relative Time Point". For example, for a treatment period of 6 months, we would define "treatment start" as an absolute time point, and then "treatment end" as a relative time point, anchored to "treatment start" with an offset of 6 months. (This time offset could be any of the three types shown in Figure 1.)

A "Cyclical Time Point" is a special type of relative time point, with two additional features: repeat times and repeat interval. To represent "q 4 month * 1 yr" (as seen in Table 1, row 8), we can define a cyclical time point with a repeat interval of 4 months and a repeat times of 3.

	During Radiation, Weekly	Every 3 months after therapy (1st 2 years)	Every 6 months after therapy (next 3 years)
CBC, Differential, Platelets	X	X	X
Creatinine	X	X	X
Tumor Measurement		X	X

Table 2. A portion of the study calendar for protocol GOG #160.

We define three types of time intervals. The most common is a time interval bounded by two end points. However, for clinical trial protocols, we found it useful to also define a "one point time interval", where we ignore one end of the interval due to uncertainty. Examples of this type of time intervals include "pre-study", where the end point is known, but not the start point, and "follow up", where the end point may not be known, i.e. "until death" or "until disease progression". The third type of time interval is a "cyclical time interval", which is anchored to a cyclical time point and has a repeat duration.

An Event Pattern defines patterns for temporal activities in clinical trials. There are two types of temporal event patterns: single and cyclical. These specify how temporal activities are anchored to time points and how often they will happen. We use cyclical event pattern together with cyclical time point to represent cycles such as those seen in Table 1.

Scheduling Task Scenario: Study Calendaring

Based on our temporal ontology, we built a prototype scheduling decision-support tool, which mainly makes use of the study calendar in clinical trial protocols to support patient visit scheduling. In paper clinical trial protocols, the study calendar is a 2-dimensional table, where rows define different study tests and columns define time patterns or temporal constraints. As an example, Table 2 shows a small portion of the study calendar for the clinical trial protocol GOG #160.[9] When a patient enrolls in this protocol, the scheduling task is to generate a patient-specific schedule of visits and clinical tasks.

In our prototype system, we first need to encode or capture the temporal information about protocol events using our time entity model. Using the portion shown in Table 2 we would proceed as follows:

1) Define "therapy starts" and "therapy ends" as time points. "Therapy ends" can be an anchored time point referring to "therapy starts", with an offset equal to the duration of the therapy.
2) Define three time patterns: "weekly cycle", "every 3 months" and "every 6 months". The "weekly cycle" pattern is anchored to the "therapy starts"; the other patterns are anchored to "therapy ends". Their repeat times need to be computed. For example, the repeat times for pattern #1 is equal to the number of weeks of treatment; whereas pattern #2 repeats eight times, and pattern #3, six times.
3) Associate clinical tasks with different time patterns. "CBC, Differential, and Platelets" is associated with all three patterns; "Tumor Measurement" is associated just with patterns #2 and #3.

To capture all temporal constraints, sometimes we need information beyond that specified in the calendar: in this protocol, we need the duration of radiation treatment (five weeks). After that, we will incorporate patient data such as enrollment time and therapy ending time to assign time values to some "anchored time points". Then, we can reason over the temporal patterns and generate patient visits schedule showing when "CBC, differential, Platelets" should occur or when "tumor measurement" should happen.

We have entered about a dozen protocols and proved our ability to generate hypothetical patient-specific schedules from these protocols. However, our aim is not to build a user-ready tool for scheduling, but rather to explore the usefulness of an expressive knowledge representation. In fact, we envision a set of additional capabilities that could be built on top of our knowledge model. For example:

- How much do different time patterns overlap? Reasoning about overlapping activities could be important for resource management.
- On a particular date, which tests and tasks need to co-occur? This information is not usually explicit in the protocol document, and must be inferred.
- How many total patient visits are required to complete a protocol and how long will the whole study last (given some assumptions)?

In sum, there are a number of scheduling-related temporal reasoning tasks. To help care providers with these tasks, we envision a series of decision-support tools that would benefit from our temporal ontology.

Discussion

In this paper, we summarize the temporal knowledge representation requirements for scheduling tasks in clinical trial protocols and propose an ontology that meets the temporal reasoning needs of scheduling tasks. Related work in temporal knowledge representation includes Résumé,[10] EON,[1] and ASBRU.[2] In contrast to these more interval-based efforts, we view "clinical tasks" from a scheduling perspective as instantaneous events relative to the context when they happen. This view led us to define "temporal event patterns", which do not exist in EON or ASBRU, to capture temporal features of sets of clinical tasks.

In addition, rather than requiring users to explicitly define interval relationships such as DO-ALL-TOGETHER or DO-ALL-SEQUENTIAL such as are defined in ASBRU, we support the automatic reasoning of temporal relationships among time intervals. In this way, we find it is efficient to do temporal reasoning over clinical tasks' event patterns.

Although EON supports temporal representation and abstraction of patterns in clinical *patient data*,[10] it does not address the scheduling requirements for temporal representation and reasoning over clinical *events*. Both EON and ASBRU are designed to represent clinical guidelines as well as protocols. For our purposes, this means that their models include features and classes that are not necessary for protocols. For example, definitions of "intentions" or "preferences" for time intervals in ASBRU are not necessary for the scheduling task.

We expect that our temporal representation of events will enable the development of a number of reasoning tools. For example, we could build a tool to automatically generate optimal patient visit schedules based on our representation of clinical tasks' event patterns and their temporal indeterminacy. In particular, if all tasks specified by the protocol include a temporal window within which the task may be accomplished, then a system could compute an optimal set of visits so that all tasks are carried out while minimizing the number of distinct visits. Besides patient visits scheduling, we also propose possible applications such as "form submission scheduling", "drug modification scheduling", "risk assessment scheduling", etc.

Finally, we argue that an expressive temporal model specifically for clinical trial protocols could support a protocol-authoring tool by manipulating all sorts of complex temporal constructs. Such a tool might be able to detect potential problems as protocols are being written, such as incompleteness, ambiguity, or even inconsistent temporal information across different sections of the protocol. In summary, we see many opportunities for future work and tool development that could benefit from our temporal ontology for clinical trial protocols.

Acknowledgements

This work was partially funded by FastTrack, Inc., and by a grant from the Life Sciences: Informatics Technology program of the University of California. Many thanks to Carol Broverman from FastTrack for her kind support.

References

[1] S.W. Tu & M. A. Musen. Modeling Data and Knowledge in the EON Guideline Architecture. *MedInfo 2001*, London, UK, 2001.

[2] S. Miksch, Y. Shahar, & P. Johnson. Asbru: A Task-Specific, Intention-Based, and Time-Oriented Language for Representing Skeletal Plans, *7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97)*, Milton Keynes, UK, 1997.

[3] L. Ohno-Machado, J. H. Gennari, S. Murphy, N. L. Jain, S. W. Tu, et al., The GuideLine Interchange Format: A Model for Representing Guidelines. *JAMIA*, **5**(4):357–372, 1998.

[4] J.P. Bury, C. Hurt, etc. LISA: A Clinical Information and Decision Support System for Collaborative Care in Childhood Acute Lymphoblastic Leukaemia. *AMIA '02 (to appear)*.

[5] M. Kahn, C. Broverman, N. Wu, et al., A model-based method for improving protocol quality, *Applied Clinical Trials*, pp. 40-50, May 2002.

[6] S. Miksch, Plan Management in the Medical Domain, *AI Communications*, **12**(4): 209-235, 1999.

[7] C. Combi & Y. Shahar, Temporal Reasoning and Temporal Data Maintenance in Medicine: Issues and Challenges, *Computers in Biology and Medicine*, **27**(5):353-368, 1997.

[8] N. F. Noy, R. W. Fergerson, & M. A. Musen. The knowledge model of Protégé-2000: Combining interoperability and flexibility. *2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000)*, Juan-les-Pins, France, 2000.

[9] M.A. Bookman, M.D. (study chair). *GOG #160: A phase II evaluation of Anti-her2 antibody in recurrent or refractory ovarian or primary peritoneal carcinoma.* Open for enrollment 1996–2000.

[10] S. Chakravarty, Y. Shahar, Specification and Detection of Periodic Patterns in Clinical Data, *Workshop of Intelligent Data Analysis in Medicine and Pharmacology on AMIA Symposium, 1999.*