

Model Formulation ■

Modeling Electronic Discharge Summaries as a Simple Temporal Constraint Satisfaction Problem

GEORGE HRIPCSAK, MD, MS, LI ZHOU, MS, SIMON PARSONS, PhD, AMAR K. DAS, MD, PhD, STEPHEN B. JOHNSON, PhD

Abstract **Objective:** To model the temporal information contained in medical narrative reports as a simple temporal constraint satisfaction problem.

Design: A constraint satisfaction problem is defined by time points and constraints (inequalities between points). A time interval comprises a pair of points and a constraint. Five complete electronic discharge summaries and paragraphs from 226 other discharge summaries were studied. Medical events were represented as intervals, and assertions about events were represented as constraints. Through a consensus process, a set of encoding procedures and a list of issues related to encoding were generated.

Measurements: Instances of temporal disjunction and contradiction and distribution of temporal constraints were used.

Results: An average of 95 medical events (range, 46–151) and 234 temporal assertions (range, 118–388) were identified per complete discharge summary. Nondefinitional assertions were explicit (36%) or implicit (64%) and absolute (17%), qualitative (72%), or metric (11%). Implicit assertions were based on domain knowledge and assumptions, e.g., the section of the report determined the ordering of events. Issues included linking events, intermittence, periodicity, granularity, vagueness, ambiguity, uncertainty, and plans. Abstractions such as intermittence were not represented explicitly. The temporal network was sparse: Only 0.80% (range, 0.42%–1.38%) of possible constraints were instantiated. No instances of discontinuous temporal disjunction were found in the complete summaries or the 226 paragraphs. One instance of temporal contradiction was found (intrareport rate of 0.2 with a 95% confidence interval of 0.005–1.114).

Conclusion: A simple temporal constraint satisfaction problem appears sufficient to represent most temporal assertions in discharge summaries and may be useful for encoding electronic medical records.

■ *J Am Med Inform Assoc.* 2005;12:55–63. DOI 10.1197/jamia.M1623.

The electronic medical record is a temporally rich document filled with assertions about the timing of medical events, such as visits, laboratory values, symptoms, signs, diagnoses,

and procedures. Allen,¹ in his review of temporal formalisms, describes how such temporal information can be specified with the use of points anchored in absolute time, qualitative constraints among intervals, and metric relations that specify explicit durations between points. Using such formalisms, past researchers in medical informatics have attempted to model temporal assertions found in the electronic medical record.^{2–4} The expressiveness of these methods has not been tested with data from actual medical reports, however. Determining and validating a suitable temporal formalism for the electronic medical record remain a challenge in medical informatics.

Much of the work in temporal querying and temporal reasoning in medical informatics has used temporal formalisms based on points anchored in absolute time.^{2,5–10} The available data in such work consist of discrete coded events with known times, such as visit information, laboratory test results, and medication administration logs. Typically, events are represented as points or pairs of points, in which each point is a time stamp stored in a database. The primary focus of this research has been on temporal abstraction: recognizing temporal patterns among the time stamps and determining what can be inferred medically based on the temporal semantics of the events. For example, given a series of hemoglobin

Affiliations of the authors: Department of Biomedical Informatics, Columbia University, New York, NY (GH, LZ, SBJ); Department of Computer and Information Science, Brooklyn College, Brooklyn, NY (SP); Department of Psychiatry, Columbia University, New York, NY (AKD). Dr. Das is currently with Stanford Medical Informatics, Stanford University, Stanford, CA.

Abstract presented at the meeting of Intelligent Data Analysis in Medicine and Pharmacology, Stanford University, Palo Alto, CA, September 2004.

Supported by National Library of Medicine grants R01 LM06910, "Discovering and Applying Knowledge in Clinical Databases"; R01 LM07659, "Capturing and Linking Genomic and Clinical Information"; and R01 LM07268, "Using Narrative Data to Enrich the Online Medical Record."

Correspondence and reprints: George Hripcsak, MD, MS, Department of Medical Informatics, Columbia University, 622 West 168th Street, VC5, New York, NY 10032; e-mail: <hripcsak@columbia.edu>.

Received for publication: 05/24/04; accepted for publication: 09/08/04.

tests that occur at discrete time points, one can infer the presence or absence of anemia, a condition that persists over time. Medical informatics researchers have used temporal intervals to support temporal abstraction and temporal querying.^{3,9,11–14} One method, called TNET¹³ and designed for chemotherapy planning, used qualitative constraints among intervals to represent temporal relationships for stored data. Qualitative constraints can be defined by Allen's¹⁵ interval relationships. One interval can occur before, meet, overlap, start, occur during, finish, or equal another interval. In addition to this set of relationships, one can create composite disjunctions. For example, a composite disjunction in a medical narrative may consist of an episode of chest pain that occurred before or after (but not during) an intensive care unit stay.

Metric relations specify durations—or limits of duration—between pairs of time points.¹⁶ For example, it might be stated that chest pain began exactly two days before transfer to the intensive care unit or that it began at least two days before transfer. Temporal querying methods for clinical data^{3,7} have incorporated metric relations into extended representations of relational data, but these methods do not directly support metric relations in natural language statements. Johnson¹⁷ and others have applied metric relations to temporal assertions found in narrative clinical reports, but the difficulty of natural language processing has limited its use. Recent strides in natural language processing are leading to increased need for something beyond absolute time points.

Work on temporal formalisms for nonmedical data has demonstrated that the three types of temporal representations enumerated by Allen in his review could be unified as a general temporal constraint satisfaction problem.^{18–20} Unfortunately, this work has shown that drawing conclusions from such representations is NP-hard.

A subset of representations based on general temporal constraint satisfaction is known as a simple temporal constraint satisfaction problem or, more succinctly, a simple temporal problem.¹⁸ It can be solved in polynomial time (n^3). A simple temporal problem supports points anchored in absolute time, metric relations among points, and primitive Allen interval relationships. It differs from a general temporal constraint satisfaction problem, however, because it does not support discontinuous temporal disjunction: It cannot represent something that occurred during one of two or more disconnected time intervals. Discontinuous temporal disjunction becomes important in planning, for example. It might be stipulated that a patient needs to go for two tests within a month, but they must not occur on the same day. One test must occur before or after but not during the other one. This set of events cannot be represented as a simple temporal problem.

It is our belief that discontinuous temporal disjunction is not necessary to represent the vast majority of temporal assertions that are found in the electronic medical record and that a simple temporal problem will suffice. It is clear that time stamp data can be represented as a simple temporal problem,¹⁸ but the temporal assertions found in narrative text are likely to be more complex.

The goal of this study was to test the hypothesis that the temporal information commonly available in narrative clinical reports can be represented with a temporal formalism defined as a simple temporal problem. If this is true, then discrete

time stamp data and metric assertions derived from narrative reports can be integrated into a single temporal representation, and a computationally feasible algorithm can be used to draw conclusions from this representation. Furthermore, recent discoveries about simple temporal problems, such as a new efficient algorithm for sparse networks,²¹ can be applied to the electronic medical record.

Background

Consider the following short example:

Example 1: A patient developed fevers and then developed a rash. The rash appeared three weeks before admission, which occurred at midnight, January 9, 2000 (2000-01-09-00:00:00). The patient was discharged some time on the 13th, but the patient was readmitted within two weeks.

We adopted the following model to represent the temporal information contained in this excerpt.

A simple temporal problem includes the following primitive entities: (1) time points, which are instants in time; (2) the origin, which is a special time point anchored in absolute time¹⁸; (3) durations, which are lengths of time; and (4) constraints, which represent the relations among the other primitive entities. We denote time points as X_i . The origin is denoted X_o , which we arbitrarily set to 2000-01-01-00:00:00 (midnight on January 1, 2000) for this paper. We express durations as days, with hours, minutes, and seconds expressed as fractional days. Assertions about durations of years are mapped to the average number of days per year in the Gregorian calendar (365.2425), and months are mapped to 12ths of a year. Kautz and Ladkin¹⁹ extended the original formalism of Dechter et al.¹⁸ to include two types of inequalities. Two time points, X_i and X_j , and a duration, d , can be constrained by one of the following:

$$X_j - X_i \leq d \quad (1)$$

$$X_j - X_i < d \quad (2)$$

With this very simple formalism, we can represent points in absolute time and metric relations. A pair of equations where one constrains $X_j - X_i$ and the other constrains $X_i - X_j$, determines the limits of duration between two points, X_i and X_j . For example,

$$X_j - X_i \leq b \text{ and } X_i - X_j \leq a, \quad (3)$$

is equivalent to the following:

$$-a \leq X_j - X_i \leq b \quad (4)$$

The other three combinations of "<" and "≤" behave analogously. To avoid contradiction, $-a < b$ or $-a = b$ (when both inequalities are "≤") must be true.

One can anchor a point in absolute time with the following pair of constraints:

$$X_j - X_o \leq d \text{ and } X_o - X_j \leq -d \quad (5)$$

where d is the duration from 2000-01-01-00:00:00 to point X_j , and d is negative if X_j occurred before 2000.

Intervals are represented by a pair of points where the end points are constrained not to contradict. Therefore, given

interval h with start point $X_{h,start}$ and finish point $X_{h,finish}$, $X_{h,start} - X_{h,finish} \leq 0$. The primitive Allen relationships between intervals can be defined in terms of constraints, as illustrated by Kautz and Ladkin.¹⁹

Allen and Ferguson²² define a temporal event as a concept that agents use to classify useful and relevant patterns of change. They state that events “must involve at least one object over some stretch of time or involve at least one change of state,” and they discuss the distinctions among events, states, and actions.²² We adopted a broad definition of medical events to include a process with a start and finish time (e.g., an operation), the presence of a state for some period of time (e.g., the presence of a rash), and an instantaneous change (e.g., death). We model events as intervals and encode assertions about events as constraints. Our model is illustrated in Figure 1. Events can be constrained to be instantaneous by constraining the end points to be simultaneous.

The events and assertions in Example 1 can be represented as illustrated in Table 1. To simplify the example, we treat assertions about days and weeks as exact, but we cover vagueness of assertions later in the paper.

Three types of constraints are shown. The first group is based on explicit assertions in the example; they require the least medical knowledge or interpretation. The second group is based on implicit assertions using domain knowledge and assumptions. For example, admission, discharge, and readmission are modeled as instantaneous events (e.g., the instant that the clerk presses enter in the registration system), although they could be modeled as extended intervals that cover the admission and discharge processes. The last group of constraints is based on the definition of intervals.

Pairs of constraints can be aggregated. Two equations that both constrain $X_j - X_i$ can be reduced to the more restrictive of the two: Pick the constraint with the smaller (more negative) duration and favor “<” over “≤” for a tie. For example, given

$$X_j - X_i \leq 3 \text{ and} \quad (6)$$

$$X_j - X_i < 3, \quad (7)$$

one need only keep Equation 7.

Constraints are transitive, so a new constraint can be composed from a pair that shares a point in common. For example, given two constraints,

$$X_k - X_i < a \text{ and} \quad (8)$$

$$X_j - X_k \leq b, \quad (9)$$

one can compose the following new constraint:

$$X_j - X_i < a + b \quad (10)$$

If either source constraint includes “<” then the resulting constraint will use “<”; otherwise it will use “≤.”

Points and constraints can be represented as a network or graph, where points are nodes, and constraints are directed paths between nodes. Picking the more restrictive constraint between two nodes (e.g., Equation 7 over Equation 6) is analogous to picking the shorter path, and the transitive operation

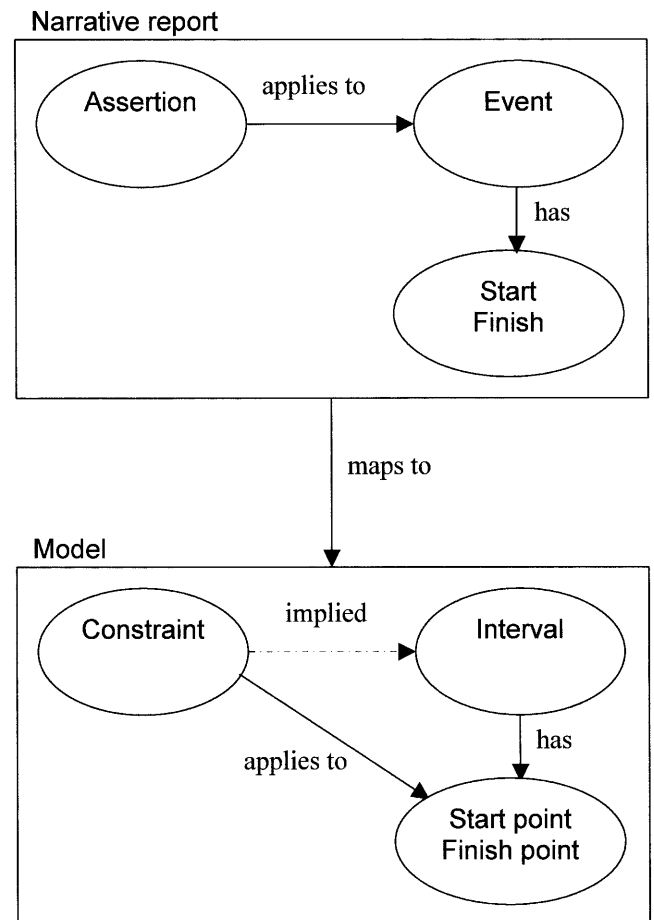


Figure 1. Narrative reports map to the temporal model. We map the temporal information in narrative reports into our model. Narrative reports describe medical events and contain temporal assertions about those events. The events have a start and a finish (whose time may or may not be known). Our model contains time points; constraints, which specify the temporal relationship between pairs of points; and temporal intervals, which comprise a pair of time points and a definitional constraint that the start point must not follow the finish point. We map medical events to intervals and assertions to constraints. Individual constraints apply to points, which imply constraints on the intervals.

is analogous to deriving a new path between two nodes by adding paths through intermediate nodes.

Given the information in Table 1, a simple algorithm can calculate the transitive closure by exploiting the transitive relation and aggregation. It will produce all the information that can be inferred about pairs of points without abstraction or additional domain knowledge. The resulting network is known as the minimum network representation of the problem. Dechter et al.¹⁸ illustrate the algorithm in this context, and Kautz and Ladkin¹⁹ extend it to include “<” and “≤”. The so-called all-pairs-shortest-paths algorithm runs in n^3 time. The result for Example 1 is shown in Table 2.

Constraints can contradict each other, leading to an inconsistent network. In Example 1, an assertion that the patient was discharged within two days of admission would render the network inconsistent because it would contradict the date assertions. The all-pairs-shortest-paths algorithm computes

Table 1 ■ Example 1 as a Simple Temporal Problem

Constraint	Meaning
Based on explicit assertions stated in the example	
$X_{\text{fever,start}} - X_{\text{rash,start}} < 0$	Fever episode began before rash episode.
$X_{\text{admit,start}} - X_{\text{rash,start}} \leq 21$ and $X_{\text{rash,start}} - X_{\text{admit,start}} \leq -21$	Rash episode began 3 weeks before admission.
$X_{\text{admit,start}} - X_o \leq 8$ and $X_o - X_{\text{admit,finish}} \leq -8$	Admission occurred at 2000-01-09-00:00:00 (the mentioned instant occurred during admission).
$X_{\text{discharge,finish}} - X_o < 13$ and $X_o - X_{\text{discharge,start}} \leq -12$	Discharge occurred some time on 2000-01-13 (discharge occurred during the day 2000-01-13).
$X_{\text{readmit,start}} - X_{\text{discharge,finish}} \leq 14$ and $X_{\text{discharge,finish}} - X_{\text{readmit,start}} < 0$	Patient was readmitted after discharge but no more than two weeks after discharge ("within two weeks").
Based on implicit assertions, domain knowledge, and assumptions	
$X_{\text{admit,finish}} - X_{\text{admit,start}} \leq 0$	Admission is instantaneous.
$X_{\text{discharge,finish}} - X_{\text{discharge,start}} \leq 0$	Discharge is instantaneous.
$X_{\text{readmit,finish}} - X_{\text{readmit,start}} \leq 0$	Readmission is instantaneous.
Based on the definition of intervals	
$X_{\text{fever,start}} - X_{\text{fever,finish}} \leq 0$	Fever interval is well formed (start never follows finish).
$X_{\text{rash,start}} - X_{\text{rash,finish}} \leq 0$	Rash interval is well formed.
$X_{\text{admit,start}} - X_{\text{admit,finish}} \leq 0$	Admission interval is well formed.
$X_{\text{discharge,start}} - X_{\text{discharge,finish}} \leq 0$	Discharge interval is well formed.
$X_{\text{readmit,start}} - X_{\text{readmit,finish}} \leq 0$	Readmit interval is well formed.

consistency as a by-product, but faster algorithms can be used if the goal is only to check consistency.¹⁸ Determining the cause and eliminating the inconsistency are computationally hard problems and the subject of recent research.²³ For the purpose of this study, we will determine the magnitude of the problem rather than solve it.

Methods

The goal of our study was to test the hypothesis that narrative data in the electronic medical record can be encoded as a simple temporal constraint satisfaction problem. We focused on discharge summaries because they present a complex clinical picture covering the patient's history, course, and planning. We defined medical events as any events found in the discharge summary, and we attempted to encode all the temporal assertions about events as temporal constraints in our model.

We used an iterative process to develop the encoding procedure. One author (LZ, an informatics PhD candidate with a medical degree) encoded one discharge summary, and she and another author (GH, an informatics faculty member and internist) developed an initial procedure. The procedure and encoding issues were presented to the other authors and to a clinical data mining research group, and the procedure was modified. The author (LZ) encoded a second discharge summary and reviewed it with the other author (GH) and then with the group, modified the procedure, and then re-encoded both discharge summaries. This was continued until five discharge summaries were encoded using the ultimate encoding procedure. The five discharge summaries came from four unique patients. Two of the discharge summaries came from the same patient but covered different clinical problems; they were included to check whether contradiction would arise between the two summaries.

Both our coders had medical knowledge. We allowed simple medical reasoning and knowledge of English grammar to recognize and link events; for example, we handled pronouns, synonyms, and classification. In this way, it could be recog-

nized that two assertions referred to the same event. We disallowed complex medical reasoning such as diagnosis. For example, even if a symptom that is pathognomonic for a disease was noted, a temporal assertion about the symptom was not applied to the disease.

The expected result was a procedure for encoding assertions, including special issues and how to deal with them; a list of assertions that could not be encoded within a temporal constraint satisfaction problem; an assessment of the need for discontinuous disjunction, which would require a general rather than simple temporal problem; the frequency of temporal contradiction within reports; an estimate of the number of events and assertions per report; an estimate of the sparseness of assertions (the proportion of possible pairwise constraints that were actually instantiated); and the time to encode reports manually.

To further verify that the temporal assertions in discharge summaries constitute a simple temporal problem, we reviewed a broad sample of discharge summaries. We randomly sampled one paragraph from each of 226 discharge summaries on 226 different randomly chosen patients. We selected paragraphs at least 175 characters in length (the minimum paragraph length varied slightly with the section label length). One author (GH) reviewed each paragraph looking for discontinuous disjunction, contradiction, or anything that could not be encoded as a simple temporal problem. The difficulties were tallied.

This study was approved by the Institutional Review Board of the Columbia University Medical Center.

Results

The five discharge summaries were encoded. Each summary had 826 words on average with a range from 346 to 1637. The process initially took weeks to encode each discharge summary, but after the basic encoding procedures were established, the time decreased to six to 12 hours per report. We developed the following general procedures and encountered a number of issues along the way.

Table 2 ■ The Minimal Network Representation (Transitive Closure) of Example 1

Constraint	Meaning
$X_{\text{fever,start}} - X_o < -13$	Fever began before 1999-12-19-00:00:00.
$X_{\text{rash,start}} - X_o \leq -13$ and $X_o - X_{\text{rash,start}} \leq 13$	Rash began 1999-12-19-00:00:00.
$X_o - X_{\text{rash,finish}} \leq 13$	Rash ended no earlier than 1999-12-19-00:00:00.
$X_{\text{admit,start}} - X_o \leq 8$ and $X_o - X_{\text{admit,start}} \leq -8$ and $X_{\text{admit,finish}} - X_o \leq 8$ and $X_o - X_{\text{admit,finish}} \leq -8$	Admission occurred at 2000-01-09-00:00:00.
$X_{\text{discharge,start}} - X_o < 13$ and $X_o - X_{\text{discharge,start}} \leq -12$ and $X_{\text{discharge,finish}} - X_o < 13$ and $X_o - X_{\text{discharge,finish}} \leq -12$	Discharge occurred some time on 2000-01-13.
$X_{\text{readmit,start}} - X_o < 27$ and $X_o - X_{\text{readmit,start}} < -12$ and $X_{\text{readmit,finish}} - X_o < 27$ and $X_o - X_{\text{readmit,finish}} < -12$	Readmission occurred after 2000-01-13-00:00:00 and before 2001-01-28-00:00:00.
$X_{\text{fever,start}} - X_{\text{rash,start}} < 0$ and $X_{\text{fever,start}} - X_{\text{rash,finish}} < 0$	Fever began before rash.
$X_{\text{fever,start}} - X_{\text{admit,start}} < -21$ and $X_{\text{fever,start}} - X_{\text{admit,finish}} < -21$	Fever began before 3 weeks before admission.
$X_{\text{fever,start}} - X_{\text{discharge,start}} < -25$ and $X_{\text{fever,start}} - X_{\text{discharge,finish}} < -25$	Fever began before 25 days before discharge.
$X_{\text{fever,start}} - X_{\text{readmit,start}} < -25$ and $X_{\text{fever,start}} - X_{\text{readmit,finish}} < -25$	Fever began before 25 days before readmission.
$X_{\text{admit,start}} - X_{\text{rash,start}} \leq 21$ and $X_{\text{rash,start}} - X_{\text{admit,start}} \leq -21$ and $X_{\text{admit,finish}} - X_{\text{rash,start}} \leq 21$ and $X_{\text{rash,start}} - X_{\text{admit,finish}} \leq -21$	Rash began 3 weeks before admission.
$X_{\text{admit,start}} - X_{\text{rash,finish}} < 21$ and $X_{\text{admit,finish}} - X_{\text{rash,finish}} \leq 21$	Rash ended no earlier than 3 weeks before admission.
$X_{\text{discharge,start}} - X_{\text{rash,start}} < 26$ and $X_{\text{rash,start}} - X_{\text{discharge,start}} \leq -25$ and $X_{\text{discharge,finish}} - X_{\text{rash,start}} < 26$ and $X_{\text{rash,start}} - X_{\text{discharge,finish}} \leq -25$	Rash began 25 to (but not including) 26 days before discharge.
$X_{\text{discharge,start}} - X_{\text{rash,finish}} < 26$ and $X_{\text{discharge,finish}} - X_{\text{rash,finish}} < 26$	Rash ended no earlier than 26 days before discharge.
$X_{\text{readmit,start}} - X_{\text{rash,start}} < 40$ and $X_{\text{rash,start}} - X_{\text{readmit,start}} < -25$ and $X_{\text{readmit,finish}} - X_{\text{rash,start}} < 40$ and $X_{\text{rash,start}} - X_{\text{readmit,finish}} < -25$	Patient was readmitted after 25 days after rash began and before 40 days after rash began.
$X_{\text{readmit,start}} - X_{\text{rash,finish}} < 40$ and $X_{\text{readmit,finish}} - X_{\text{rash,finish}} < 40$	Rash ended no earlier than 40 days before readmission.
$X_{\text{discharge,start}} - X_{\text{admit,start}} < 5$ and $X_{\text{admit,start}} - X_{\text{discharge,start}} \leq -4$ and $X_{\text{discharge,finish}} - X_{\text{admit,start}} < 5$ and $X_{\text{admit,start}} - X_{\text{discharge,finish}} \leq -4$ and $X_{\text{discharge,start}} - X_{\text{admit,finish}} < 5$ and $X_{\text{admit,finish}} - X_{\text{discharge,start}} \leq -4$ and $X_{\text{discharge,finish}} - X_{\text{admit,finish}} < 5$ and $X_{\text{admit,finish}} - X_{\text{discharge,finish}} \leq -4$	Patient was discharged 4 to (but not including) 5 days after admission.
$X_{\text{readmit,start}} - X_{\text{admit,start}} < 19$ and $X_{\text{admit,start}} - X_{\text{readmit,start}} < -4$ and $X_{\text{readmit,finish}} - X_{\text{admit,start}} < 19$ and $X_{\text{admit,start}} - X_{\text{readmit,finish}} < -4$ and $X_{\text{readmit,start}} - X_{\text{admit,finish}} < 19$ and $X_{\text{admit,finish}} - X_{\text{readmit,start}} < -4$ and $X_{\text{readmit,finish}} - X_{\text{admit,finish}} < 19$ and $X_{\text{admit,finish}} - X_{\text{readmit,finish}} < -4$	Patient was readmitted more than 4 days after admission and before 19 days after admission.
$X_{\text{readmit,start}} - X_{\text{discharge,start}} \leq 14$ and $X_{\text{discharge,start}} - X_{\text{readmit,start}} < 0$ and $X_{\text{readmit,finish}} - X_{\text{discharge,start}} \leq 14$ and $X_{\text{discharge,start}} - X_{\text{readmit,finish}} < 0$ and $X_{\text{readmit,start}} - X_{\text{discharge,finish}} \leq 14$ and $X_{\text{discharge,finish}} - X_{\text{readmit,start}} < 0$ and $X_{\text{readmit,finish}} - X_{\text{discharge,finish}} \leq 14$ and $X_{\text{discharge,finish}} - X_{\text{readmit,finish}} < 0$	Patients was readmitted after discharge but no more than 2 weeks after discharge.
$X_{\text{admit,finish}} - X_{\text{admit,start}} \leq 0$	Admission is instantaneous.
$X_{\text{discharge,finish}} - X_{\text{discharge,start}} \leq 0$	Discharge is instantaneous.
$X_{\text{readmit,finish}} - X_{\text{readmit,start}} \leq 0$	Readmission is instantaneous.
$X_{\text{fever,start}} - X_{\text{fever,finish}} \leq 0$	Fever interval is well formed.
$X_{\text{rash,start}} - X_{\text{rash,finish}} \leq 0$	Rash interval is well formed.
$X_{\text{admit,start}} - X_{\text{admit,finish}} \leq 0$	Admission interval is well formed.
$X_{\text{discharge,start}} - X_{\text{discharge,finish}} \leq 0$	Discharge interval is well formed.
$X_{\text{readmit,start}} - X_{\text{readmit,finish}} \leq 0$	Readmission interval is well formed.

Medical Events

We encoded each event as a descriptive phrase and a unique event identifier and modeled it as an interval. Instantaneous events were constrained to have equal start and finish times.

In general, a discharge summary represents an observation (at the time of dictation) of earlier observations (e.g., at admission) of previous or current events. Patient-reported symptoms have three levels of observation (dictation observation of physician observation of patient observation of symptoms). Where possible, we attempted to simplify the representation, directly representing the underlying event rather than fully representing the observation of an observation.

Explicit Temporal Assertions

Assertions about events were translated to temporal constraints. We found that explicit temporal assertions generally followed one of three forms. (1) Assertions anchored events in absolute time. We represented "In June of 1995 he developed obstructive jaundice" as an episode of obstructive jaundice whose start point occurred no earlier than 1995-06-01-00:00:00 and earlier than 1995-07-01-00:00:00. (2) Assertions linked two events qualitatively. We represented "He developed sepsis following the procedure" as an episode of sepsis and a procedure, in which the start of sepsis occurred after the finish of the procedure. (3) Assertions linked two events quantitatively (metrically). We represented "Over the next 48 hours in the hospital he remained afebrile" as an interval of no fever whose finish point was at least 48 hours after the start.

In addition to these explicit temporal assertions, a number of implicit temporal assertions could be inferred based on domain knowledge and further assumptions. We cover several types of implicit assertions in the following subsections.

Visit Information

Particular medical events are basic to the temporal structure of the report and warrant special handling. For example, many of the events in the report are explicitly linked to the admission, discharge, or hospitalization of the patient. We modeled admission and discharge as instantaneous events and constrained admission to equal the start of hospitalization and discharge to equal the finish of hospitalization.

Context

A number of temporal assertions can be inferred from the context of the report.¹⁷ The events in the chief complaint, medical history, and present illness occurred at or before admission, unless explicitly stated otherwise. Some sections, such as the hospital course, follow a chronology. Events can be assumed to have occurred at the same time or after events mentioned in preceding sentences, unless stated otherwise. Other sections, such as the physical examination, do not follow such a chronology. Diseases mentioned among the discharge diagnoses occurred some time during the admission. The discharge plan is set in the future with respect to the discharge and remains hypothetical.

Linking Events

One of the main challenges is determining that two different assertions are actually referring to the same event. For example, two mentions of a myocardial infarction may refer to the same or different episodes. Events can sometimes be linked using linguistic knowledge or using knowledge of medicine.

For example, in the sentence "The patient was initially short of breath; however, this resolved with diuresis," the episode of shortness of breath mentioned in the first half can be linked to the "this" in the second half using knowledge of linguistics. Domain knowledge is required to link the following phrase "and he was therefore treated with ampicillin, gentamicin, and Flagyl," with the termination of the treatment: "and therefore IV antibiotics will be stopped."

Intermittent Events

Intermittent events could not be fully represented as a temporal constraint. For example, "three weeks of fever" implies a three-week period over which fever came and went. We represented this as a simple interval ("intermittent fever") of length three weeks (finish point three weeks after start). Some of the temporal information (intermittence) was therefore folded into the event definition rather than represented explicitly in the temporal model. To draw conclusions about the intermittence of fever, a temporal abstraction system¹¹ would need to work outside the constraint satisfaction network, recognizing intermittence in the event definition.

Periodic Events

Several types of events were periodic. Superficially, medication dosing could be represented as a periodic event. For example, "Lasix 20 mg po qd," could be represented as a conjunction of events. In each 24-hour period for the duration of the prescription, it can be asserted that a dose of Lasix was taken (with each day as a separate event in a simple temporal model). It is not known whether the patient took the medication, however, so we represented it instead as a single interval of a medication at a given dose (Lasix 20 mg po qd).

Some symptoms are periodic: "Intermittent shortness of breath that occurred only in the morning." The patient could have had zero, one, or more episodes of shortness of breath during any of the mornings during the interval. This can be represented in a number of ways. We chose to represent them as a single interval of "intermittent morning shortness of breath," but one could also represent them as a series of intervals (one for each morning in the duration of the symptom) of "intermittent shortness of breath" episodes. The single-interval approach preserves the abstraction (that the symptoms occurred in the morning) at the cost of leaving the information out of our temporal model. The multiple interval approach puts more of the temporal information into our temporal representation (so that we can answer simple questions more accurately), but it loses the abstraction. Extensions to the temporal constraint model to handle periodic events explicitly are also possible.²⁴

Temporal Granularity

Different events are recorded with different temporal granularity. Clinical laboratory tests may be known to the minute, whereas pathologic examination may be known only to the day and some symptoms may be known to the year or even to the decade. We encoded assertions about events using whatever granularity was stated in the report. In Example 1, the discharge is only known to the day and is represented as being constrained to a range of absolute times. Constraints can even represent what is known about the start, finish, and

duration of a given interval with different granularities; this situation arises in medicine.³

Implicit Vagueness

Medical assertions in narrative text are often vague. The statement that a rash began three weeks ago implies a range of possibilities. Although it might be possible to model this assertion as a probability density function over time with the greatest probability around three weeks, calculating conclusions from such a representation is NP-hard.²⁵ We choose instead to model vagueness by widening the limits of the constraints.

We represented an assertion as plus or minus one unit with a minimum of one half unit. "The rash appeared three weeks before admission" was represented as two to four weeks as follows:

$$X_{\text{admit,start}} - X_{\text{rash,start}} \leq 28 \text{ and } X_{\text{rash,start}} - X_{\text{admit,start}} \leq -14 \quad (11)$$

Trailing zeros were assumed to be nonsignificant, so that the "unit" was the last significant digit. Statements like "48 hours" were assumed to refer to a number of days rather than an exact number of hours. Therefore, 1 week would be 0.5 to two weeks, and 22 days would be 21 to 23 days, 20 days would be 10 to 30 days, and 48 hours would be 24 to 72 hours. We based the above heuristics on our experience with the discharge summaries, but we expect to refine the heuristics based on further experience.

Explicit Vagueness

Vagueness qualifiers such as "about" widen the interval: from -50% to +100%. "About three weeks" would be represented as 1.5 to six weeks. The size of the range may depend on the context.

Ambiguity

We found many cases of ambiguity: "Workup by ERCP (endoscopic retrograde cholangiopancreatography) (at an outpatient facility) revealed an obstructing area in the distal common duct. He developed sepsis following the procedure and a transhepatic catheter was placed for drainage. Bilirubin was 17 at the time." It is unclear whether "at the time" refers to workup, development of sepsis, or catheter placement. We considered this an interpretation issue rather than an actual representation issue, although another representation might be able to represent the ambiguity of the assertion.

Uncertainty

The statement that a patient may have had a myocardial infarction three years ago presents a challenge. First, there is the ambiguity of what is uncertain: the temporal assertion or the existence of the event. In this example, it appears that the temporal assertion is known well enough, but whether a myocardial infarction actually happened is uncertain. We do not attempt to model the uncertainty of the event, but instead represent the event as "possible myocardial infarction," which would then need to be handled by a higher level reasoning system.

Clinical Plans and Statements About the Future

We are concerned with historical medical events reported in an electronic medical record. Discharge summaries frequently

refer to therapeutic plans and prognosis. Because these are hypothetical events, we do not wish to represent them as historical events. Instead, we represent the statement of the plan (in effect, the observation of a plan) as the event rather than the hypothetical events within it.

Contradiction

We found one example of a temporal contradiction in the five discharge summaries. The hospital course section of one report stated that a patient received an inpatient medication on "1/7/97" but the patient was admitted 1/6/1998. Such typographical errors are probably common in January. This corresponds to a mean intrareport contradiction rate of 0.2 (95% confidence interval, 0.005–1.114). No contradictions were found in the two reports on the same patient. We believe that on assessing the entire record, contradictions will become more common. For example, the stated age of the patient frequently contradicts the registration system's birth date. Contradictory assertions can each be represented as constraints, but the network as a whole becomes inconsistent and drawing conclusions becomes impossible. Contradiction needs to be eliminated before further reasoning can occur.

Discontinuous Disjunction

We found no examples of actual discontinuous temporal disjunction—that an event was known to occur sometime within a discontinuous time interval—in the reports. We believe that complex plans may contain such disjunction, but we do not represent the events inside a plan.

Descriptive Statistics and Sparseness

Each summary had an average of 95 medical events (range, 46–151) and 234 temporal assertions (range, 118–388). The latter includes one definitional constraint per event (i.e., every interval has the constraint that the start must not follow the finish). The distribution of nondefinitional constraints to which the assertions mapped is shown in Table 3. About two thirds of nondefinitional constraints were implicit and almost three fourths were qualitative. Compared with a fully connected temporal network, in which every point is related to every other point with a pair of directed paths, only 0.80% (range, 0.42%–1.38%) were instantiated before transitive closure. This constitutes a sparse network.²¹

Broad Review of Discharge Summaries

Among the 226 discharge summary paragraphs, no actual examples of discontinuous temporal disjunction were found. There were several examples of logical disjunction involving differential diagnoses ("the working diagnosis for the temperature was due to a viral or respiratory infection"), alternative plans ("it was decided to discharge the patient to home or at least to care with strong neurologic and psychiatric follow-up"), and uncertainty in a patient-reported history, but these did not involve time. There was one explicit example of temporal conjunction ("the patient is nonambulatory and was nonambulatory upon presenting to the hospital") and several implicit examples (a patient was febrile during two disconnected intervals). These would be modeled as two episodes of nonambulation or fever in our representation. While these examples foreshadow the difficulty of interpreting and reasoning from narrative text, they do not represent temporal disjunction. No new temporal issues were identified in the 226 paragraphs.

Table 3 ■ Distribution of Nondefinitional Constraints in Discharge Summaries

Property of source assertion	Constraints No. (Range)	Constraints/Event Ratio (Range)	Proportion of Constraints % (Range)
Type of assertion			
Explicit: explicitly stated in the report	42 (34–60)	0.53 (0.23–0.79)	35.8 (14.8–55.4)
Implicit: inferred from domain knowledge and assumptions	98 (37–202)	0.96 (0.64–1.34)	64.2 (44.6–85.2)
Total	140 (74–237)	1.49 (1.36–1.61)	100.0
Time reference			
Absolute: anchored in absolute time (i.e., the constraint involves the origin)	20 (12–38)	0.25 (0.11–0.48)	16.8 (6.8–29.7)
Qualitative: relative ordering of pairs of events (e.g., before or after)	104 (48–193)	1.08 (0.96–1.28)	72.4 (64.9–81.4)
Metric (quantitative): relative timing of pairs of events with explicit nonzero duration stated	15 (4–28)	0.16 (0.09–0.21)	10.9 (5.4–14.5)
Total	140 (74–237)	1.49 (1.36–1.61)	100.0

Discussion

The main outcomes of our study were an enumeration of issues that arise when one attempts to encode the temporal information in electronic discharge summaries and a preliminary procedure for encoding the information. Our results demonstrate that most of the temporal assertions found in discharge summaries can be represented with a temporal formalism based on a temporal constraint satisfaction problem.

We uncovered no need for discontinuous temporal disjunction to represent historical facts. Therefore, a *simple* temporal constraint satisfaction problem appeared sufficient, which permits a simpler representation and easier computation. Furthermore, the temporal networks were very sparse, in the sense that, on average, a given event was linked to few other events. New efficient algorithms can calculate the transitive closure of sparse networks very quickly.²¹

We found a single intrareport contradiction among the five reports, and no contradictions between the two reports on the same patient. This is fortunate, for addressing inconsistency is difficult and CPU intensive. While contradiction may be more frequent when the whole electronic record is considered, having local consistency (within reports) can improve overall performance.²³

We believe that our approach may be used to encode the temporal information in coded and narrative data in the electronic medical record. Converting coded data to a simple temporal problem should be straightforward. Converting narrative data would require natural language processing.¹⁷ This is likely to be a significant challenge. If it is feasible, however, we envision the following scenario. A repository of electronic medical records could be parsed and encoded as simple temporal problems. Contradictions would have to be eliminated, possibly by assigning certainty factors to assertions based on the source—indirectly reported times being less certain than direct laboratory evidence, for example—and adjudicating automatically. The result could be stored in a clinical data warehouse, and clinical researchers could query it using a temporal query language, such as that described by Kahn et al.,⁹ Das and Musen,¹² O'Connor et al.,¹⁰ or Combi et al.²⁶ but adapted for this purpose. With further work, the temporal information could become input into a temporal abstraction method similar to that described by Shahar¹¹ for more sophisticated analysis.

The advantages of representing temporal information as a simple temporal problem are that computation is

tractable, the representation is relatively simple, and we can benefit from a large body of research on simple temporal problems. For example, we chose not to define a formal algebra for intervals in this paper because such an algebra is not needed to characterize the information in a discharge summary. The algebra is instead useful for querying the encoded result. Defining an algebra should be straightforward, however, and should benefit from previous work in this area.^{19,20}

Our approach to representing temporal information is not the only possible one. Researchers in medical informatics have proposed several temporal representations, such as conceptual graphs² or object-oriented models,⁴ to encode the types of information found in medical narrative reports. The expressiveness of these approaches has not been validated with clinical data, however.

Our approach is most similar to that of Johnson¹⁷ and Combi and Pozzi.³ Johnson represented the temporal information in medical narrative as a temporal graph and enumerated major issues in encoding narrative data. He did not formalize it as a temporal constraint satisfaction problem, however, and he did not carry out a formal survey of a set of reports. Combi and Pozzi created a model called HMAP to represent intervals with different granularities and indeterminacy from narrative data focusing on a medical example. They defined a set of notations for different interval constructs and predicates to query the resulting temporal database. All intervals were required to have a reference to absolute time, however. They showed that their representation was a subset of a simple temporal problem. Our approach can handle the same issues of granularity and indeterminacy, and we believe that a similar set of predicates could be defined.

We made a number of assumptions in encoding reports. For example, temporal vagueness was represented as determinate ranges. This was simpler and more easily computed than fuzzy assertions or probability density functions. It may be possible to learn the appropriate ranges based on a large clinical repository. Given an assertion about events whose time is known in another way, one could estimate the usual range of vagueness. For example, if an author states that the previous admission was “about three years ago,” one could correlate the statement with the actual previous admission date in the registration system.

Wherever possible, we attempted to encode all the temporal information as temporal constraints. This led to the decision to represent instantaneous events as degenerate intervals

rather than as points. In this way, the assertion that an event is instantaneous is encoded as the constraint that the start and finish are simultaneous. If instantaneous events were encoded as points, then the assertion that an event is instantaneous would require a change in the notation for that event.

Our approach does not explicitly represent the temporal semantics of high-level clinical abstractions contained in discharge summaries. That is, abstractions such as intermittence in "recurrent abdominal pain" were not formalized as constraints among episodes of abdominal pain but were rather incorporated implicitly as part of the event descriptor. This issue goes beyond obvious examples like intermittence and periodicity to the very definition of medical events. Syndromes, for example, can be defined as temporal abstractions over a set of signs and symptoms. We believe that it is useful to separate assertions about the timing of events, which we encoded as temporal constraints, from higher level abstractions. The separation allows us to address a more feasible problem. A clinical researcher can work at the abstraction level (e.g., query both for anemia and for sustained low hemoglobin) until more sophisticated temporal reasoning¹¹ can be supported or until the model is extended to handle concepts such as periodic events.²⁴

The small number of discharge summaries that were fully coded limits our study. This was a result of the labor required to manually encode reports. We used only two primary coders. Our approach was one of a consensus process, first between the two coders and then with the rest of the group. We noted, however, that there were often several ways to encode the same set of events even within our defined encoding procedure. For example, for the statement, "The manifestation of the disease has been diarrhea and had been treated in the past with Azulfidine and steroids," one encoding used two events: (1) presence of diarrhea and (2) treatment of diarrhea with Azulfidine and steroids, whereas another encoding used three events: (1) presence of diarrhea, (2) treatment with Azulfidine, and (3) treatment with steroids. The differences often resulted from differences in the granularity of defining events. Nevertheless, the two coders did independently agree on the lack of discontinuous disjunction and rarity of contradiction.

Conclusion

A number of issues arise when one attempts to encode the temporal information in electronic discharge summaries. A temporal constraint satisfaction problem can represent much of the temporal information. Due to the lack of disjunction, a simple temporal problem is sufficient, allowing transitive closure to be calculated in polynomial time. Few intrareport contradictions were found. The model does not explicitly represent some temporal abstractions such as intermittence.

References ■

1. Allen JF. Time and time again: the many ways to represent time. *Int J Intell Systems*. 1991;6:1-14.
2. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. *J Am Med Inform Assoc*. 1994;1:218-32.
3. Combi C, Pozzi G. HMAP—a temporal data model managing intervals with different granularities and indeterminacy from natural language sentences. *VLDB J*. 2001;9:294-311.
4. Dolin RH. Modeling the temporal complexities of symptoms. *J Am Med Inform Assoc*. 1995;2:323-31.
5. Combi C, Shahar Y. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. *Comput Biol Med*. 1997;27:353-68.
6. Keravnou ET. Temporal reasoning in medicine. *Artif Intell Med*. 1996;8:187-91.
7. Das AK, Musen MA. A foundational model of time for heterogeneous clinical databases. *Proc AMIA Annu Fall Symp*. 1997;106-10.
8. Nigrin DJ, Kohane IS. Temporal expressiveness in querying a time-stamp-based clinical database. *J Am Med Inform Assoc*. 2000;7:152-63.
9. Kahn MG, Tu S, Fagan LM. TQuery: a context-sensitive temporal query language. *Comput Biomed Res*. 1991;24:401-19.
10. O'Connor MJ, Tu SW, Musen MA. The Chronus II temporal database mediator. *Proc AMIA Symp*. 2002:567-71.
11. Shahar Y. A framework for knowledge-based temporal abstraction. *Artif Intell*. 1997;90:79-133.
12. Das AK, Musen MA. A temporal query system for protocol-directed decision support. *Methods Inf Med*. 1994;33:358-70.
13. Kahn MG, Fagan LM, Tu S. Extensions to the time-oriented database model to support temporal reasoning in medical expert systems. *Methods Inf Med*. 1991;30:4-14.
14. Kohane IS, Haimowitz IJ. Hypothesis-driven data abstraction with trend templates. *Proc Annu Symp Comput Appl Med Care*. 1993:444-8.
15. Allen JF. Maintaining knowledge about temporal intervals. *Communications ACM*. 1983;26:832-43.
16. Dean TL, McDermott DV. Temporal data base management. *Artif Intell*. 1987;32:1-55.
17. Johnson S. Temporal information in medical narrative. In: Sager N, Friedman C, Lyman MS, (eds). *Medical Language Processing: Computer Management of Narrative Data*. Reading, MA: Addison-Wesley, 1987, pp 175-94.
18. Dechter R, Neiri I, Pearl J. Temporal constraint networks. *Artif Intell*. 1991;49:61-95.
19. Kautz HA, Ladkin PB. Integrating metric and qualitative temporal reasoning. In: *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, Anaheim, CA, 14-19 July 1991. Menlo Park, CA: The AAAI Press; 1991:241-6.
20. Meiri I. Combining qualitative and quantitative constraints in temporal reasoning. *Artif Intell*. 1996;87:343-85.
21. Xu L, Choueiry BY. A new efficient algorithm for solving the simple temporal problem. Presented at the 10th International Symposium on Temporal Representation and Reasoning and Fourth International Conference on Temporal Logic, July 8-10, 2003, Cairns, Queensland, Australia, p 212.
22. Allen JF, Ferguson G. Actions and events in interval temporal logic. *J Logic Comput*. 1994;4:205-45.
23. Koubarakis M. From local to global consistency in temporal constraint networks. *Theoretical Comput Sci*. 1997;173:89-112.
24. Terenziani P. Toward a comprehensive treatment of temporal constraints about periodic events. *Int J Intell Systems*. 2003;18:429-68.
25. Dagum P, Luby M. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artif Intell*. 1993;60:141-53.
26. Combi C, Missora L, Pinciroli F. Supporting temporal queries on clinical relational databases: the S-WATCH-QL language. *Proc AMIA Annu Symp*. 1996:527-31.