Intelligent Interactive Visual Exploration of Temporal Associations among Multiple Time-oriented Patient Records

D. Klimov; Y. Shahar; M. Taieb-Maimon
Medical Informatics Research Center, Ben Gurion University of the Negev, Beer-Sheva, Israel

Keywords
Multiple patients, intelligent visualization, interactive visual data mining, temporal association, temporal abstraction, knowledge-based systems

Summary
Objectives: To design, implement and evaluate the functionality and usability of a methodology and a tool for interactive exploration of time and value associations among multiple-patient longitudinal data and among meaningful concepts derivable from these data.
Methods: We developed a new, user-driven, interactive knowledge-based visualization technique, called Temporal Association Charts (TACs). TACs support the investigation of temporal and statistical associations within multiple patient records among both concepts and the temporal abstractions derived from them. The TAC methodology was implemented as part of an interactive system, called VISITORS, which supports intelligent visualization and exploration of longitudinal patient data. The TAC module was evaluated for functionality and usability by a group of ten users, five clinicians and five medical informaticians. Users were asked to answer ten questions using the VISITORS system, five of which required the use of TACs.
Results: Both types of users were able to answer the questions in reasonably short periods of time (a mean of 2.5 ± 0.27 minutes) and with high accuracy (95.3 ± 4.5 on a 0–100 scale), without a significant difference between the two groups. All five questions requiring the use of TACs were answered with similar response times and accuracy levels. Similar accuracy scores were achieved for questions requiring the use of TACs and for questions requiring the use only of general exploration operators. However, response times when using TACs were slightly longer.
Conclusions: TACs are functional and usable. Their use results in a uniform performance level, regardless of the type of clinical question or user group involved.

Correspondence to:
Denis Klimov
Medical Informatics Research Center
Department of Information Systems Engineering
Ben Gurion University of the Negev
P.O. Box 653
Beer-Sheva 84105
Israel
E-mail: klimov@bgu.ac.il

Methods Inf Med 2009; 48: 254–262
doi: 10.3414/ME9227
prepublished: April 20, 2009

1. Introduction: Interactive Exploration and Mining of Time-oriented Data of Multiple Patients

1.1 Exploration of Multiple Patient Data

Analysis of large patient population data, such as clinical trial results, quality assessments of clinical management processes, and the search for the temporal patterns requires a tool that provides aggregate views of clinically meaningful interpretations of the longitudinal data of multiple patients.

The current visual data mining methods [1, 2] and visual exploration systems for multiple patient data, particularly in medicine typically, focus only on *raw* patient data, either via the design of visualization techniques for data exploration [3, 4] or through visual mining [5]. The user, however, requires additional cognitive and computational mechanisms to derive meaningful conclusions from the results of the analysis. Aigner et al. [6] overview several visualizations of time-oriented data and emphasize the importance of the integration of visual, analytical and user-centered methods.

To derive meaningful patterns and interpretations, called *temporal abstractions*, from the raw time-oriented patient data, we have been using the knowledge-based temporal-abstraction (KBTA) method [7]. We separate the concepts in the domain ontology into *raw concepts* (e.g., hemoglobin value, or age), and *abstract concepts* (*temporal abstractions*) (e.g., levels of anemia), which are derivable from raw concepts. Through a domain-specific temporal-abstraction knowledge base acquired from a domain expert by using appropriate tools [8, 9], this method derives interval-based temporal abstractions, such as the pattern "a period of more than two months of grade I or higher bone-marrow toxicity, followed within three months by a decrease in liver functions", as this pattern is defined in the context of a particular oncology therapy protocol. The temporal abstractions computed by the KBTA method for an *individual patient* or for a small number of patients (typically less than ten patients) can be efficiently visualized through an ontology-driven interface, called KNAVE-II [10], which has been shown to be functional and usable [11].

To completely support the exploration of *multiple*, time-oriented patients' data (including both raw data and derivable temporal abstractions), we designed and developed an enhanced extension of the KNAVE-II system, a new system called VISualizatIon of Time-Oriented RecordS (VISITORS), which integrates knowledge-based temporal reasoning mechanisms for deriving temporal patterns and abstractions with information visualization methods for the display, exploration, and analysis of associations among *multiple patients'* records. Unique to the VISITORS system is its temporal focus, supported by the VISITORS time-oriented exploration capabilities, which we refer to as *general exploration operators*: unlike the KNAVE-II system, data of one or of multiple patients can be both aggregated at, and explored within, various temporal granularities, such as hour, day, and month (or other specific time periods). The VISITORS system supports both a calendaric (absolute) timeline and a timeline relative to special events (e.g., the months following a bone-marrow transplantation event) (►Fig. 1).

Moreover and quite differently, the VISITORS system enables users to interactively specify temporal and knowledge-based constraints, through a graphical query-building module, which define the patient subsets selected for exploration (e.g., the lists of patients displayed in panel A of Fig. 1). Underlying the query-building module is the *ontology-based temporal-aggregation* (OBTAIN) query language. The query-building interface enables three types of queries supported by the OBTAIN query language: *Select Patients* (Who?), *Select Time Intervals* (When?), and *Get Patient Data* (What?). Each query re-

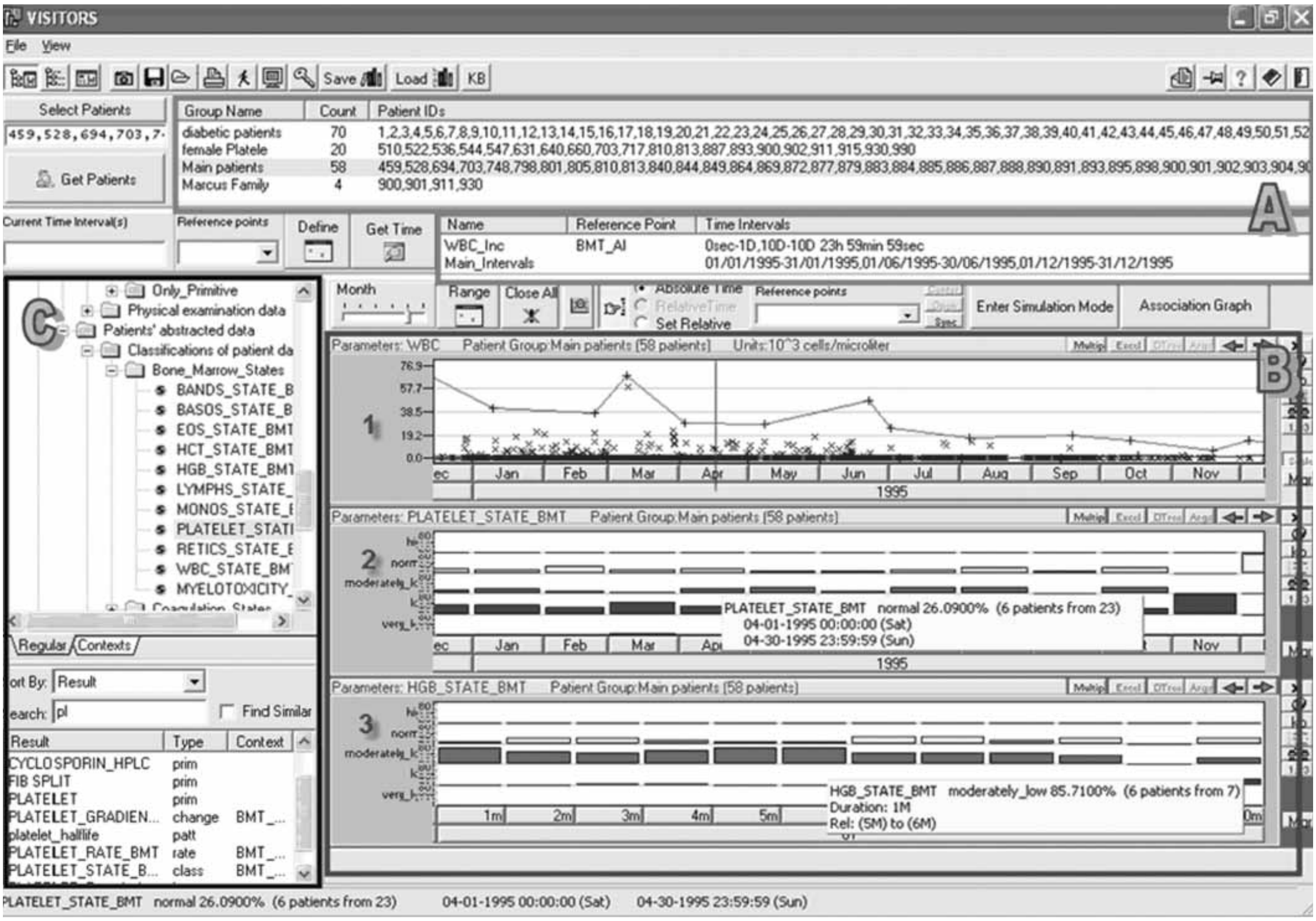

Fig. 1 The VISITORS main interface, in this case, in an oncology domain. The two top panels (denoted by A) display lists of patients and lists of time intervals, returned as answers to previous queries. The graphs (denoted by B) show the data, for a group of 58 patients, of the white blood cell count (WBC) raw concept (graph 1) and of the monthly distribution of the abstract concept Platelet-state values during 1995 (graph 2). Graph 3 shows the monthly distribution of the Hemoglobin-state abstract concept values during the first year (relative time line) following bone-marrow transplantation. The left panel (denoted by C) of the interface includes a browser to the knowledge-base in the oncology domain.

trieves either a list of patients, a list of relevant time intervals, or a list of time-oriented patient data sets, respectively, combinations of which the user can manipulate according to which patients, time periods and data values are to be analyzed further. For example, a typical *Select Patients* query would be "Find patients who had, during the first month following a bone marrow transplantation (BMT), at least one episode of *bone-marrow toxicity* (an abstraction defined by the clinical protocol) of *grade I* or higher, which lasted at least two days"; a typical *Select Time Intervals* query would be "Find periods (relative to the BMT event) during which the *WBC-state* value was less than '*normal*' for at least 50% of the patients".

A full exposition of the OBTAIN language and query building module is outside the scope of this study and can be found elsewhere [12]. Similarly, we refer to Klimov and Shahar [13] for details on the semantics of the VISITORS interface and of its general data exploration operators.

1.2 Exploration of Potentially Meaningful Temporal Associations

One of the more interesting tasks in the analysis of multiple patient data is an investigation of a new, potentially meaningful interrelation, especially a *temporal* interrelation, which we refer to as a *temporal-association task*, within a set of raw patient data and abstract concepts.

An example of a system in the medical domain for visualizing associations among multiple patients is *The Cube* [14]. *The Cube* enables interactive recognition of patient patterns through a 3D display of a set of 2D parallel diagrams (each using a horizontal time axis and a vertical value axis), where each diagram represents a patient attribute (e.g., allergies). Thus, similar patients might have parallel lines connecting the different 2D diagrams. However, *The Cube* works only with *raw data* and does not automatically aggregate similar patients into groups.

To fully support the temporal-association task, we developed the Temporal Association Chart (TAC) module (Sections 2 and 3), which we designed and implemented within the VISITORS system. We demonstrate our ideas using examples from the oncology domain (Section 4). To evaluate the VISITORS system, and in particular the TAC module, five clinicians and five medical informaticians answered ten clinical questions, five of which required the use of TACs (Section 5).

2. Aggregation of Patients' Time-oriented Data and of their Abstractions

To aggregate patients' data at arbitrary temporal granularities or during specific time periods, we defined the concept of a *delegate function* which computes a *delegate value*: Given a single patient's time-oriented data for a specific concept (raw or abstract) over a particular time interval (including a predefined granularity level), we calculate the *delegate value* of the patient's data at each time granule (or at some other particular time period) using a function specific to the concept, the context, and the temporal granularity. Such calculation is performed for each patient in the group. More formally, the input data has the following data structure:

$$input\_data \equiv <Patient_n, Concept_c, Ts_{n,m}, Te_{n,m}, value_{c,n,m}>^*,\ 1 = n = N,\ 1 = m = M_n,$$

where N is the number of patients. M_n is the number of *values* of $Concept_c$ for $Patient_n$; $Ts_{n,m}$ and $Te_{n,m}$ are the start and end times of the m-th observation (or temporal abstraction) for $Patient_n$, with value $value_{c,n,m}$.

The delegate value for $Patient_n$ of $Concept_c$ within a specific aggregation time period $[Ts_{agg}, Te_{agg}]$ is computed by the concept-specific delegate function $DF_{concept}$ from the *input_data* as follows:

$$delegate\_value_{c,n,Tsagg,Teagg} = DF\ [(Ts_{n,1}, Te_{n,1}, value_{c,n,1}), \ldots (Ts_{n,i}, Te_{n,i}, value_{c,n,i}), \ldots\ (Ts_{n,K}, Te_{n,K}, value_{c,n,K})],$$
$$Ts_{agg} \leq Ts_{n,i},\ Te_{n,i} \leq Te_{agg},$$
$$1 \leq i \leq K = K_{c,n,Tsagg,Teagg}$$

where $K = K_{c,n,Tsagg,Teagg}$ is number of instances of $Concept_c$ for $Patient_n$ measured within the $[Ts_{agg}, Te_{agg}]$ period. That is, K varies per each concept, patient, and time period.

The delegate function of each concept is defined in the knowledge base, or is chosen at runtime by the user from several predefined default functions. For example, assume that the results of a patient's three blood glucose (BGL) observations on January 1 revealed the following values: 92 g/dl at 5 a.m., 140 g/dl at 11 a.m. and 182 g/dl at 8 p.m. If *maximum* is the default *daily* delegate function for BGL, then the patient had a **daily delegate value** of 182 g/dl for BGL. However, the user can choose another suitable delegate function (such as the *mode* or the *mean*). Indeed, for granularity of months, it might be preferable to use the mean as a delegate function (applied to *all* raw data within each month).

In the case of interval-based temporal abstractions, such as intervals of different grades of bone-marrow toxicity, we provide additional delegate functions, such as the value of the abstractions that has the maximal cumulative duration during the relevant time period.

Indeed, in theory, almost any function from multiple values into one value (with same units) can serve the role of a delegate function. However, it must be applied to *each time interval* in the *relevant temporal granularity* (e.g., day), and, of course, must make clinical sense, and is thus specific to each clinical concept and medical context.

3. Temporal Association Charts

The core of a Temporal Association Chart (TAC) is an ordered list of raw and/or abstract domain concepts (e.g., platelet state, hemoglobin value, WBC count), in a particular order determined by the user. Each concept is measured (or computed, in the case of a temporal abstraction) for a particular patient group during a concept-specific particular time period. The period can be different for each concept. Between every pair of consecutive concepts in the list, a set of relations amongst the delegate values, for each patient, of these neighboring concepts, will be computed. If one of the concepts is raw, each relation will be between a delegate value of the first concept and a delegate value of the second concept for each patient. If both concepts are abstract, the relations between the delegate values for all patients will be aggregated into a set of extended relations – *temporal association rules*, one rule per each combination of values from both concepts. Each rule represents the *set* of patients who have had this particular combination of values for the two abstract concepts.

TACs are created by the user in two steps. First, the user selects two or more concepts, using an appropriate interface (not shown here), possibly changing the order as necessary; second, the user selects the group of patients (e.g., from a list of groups retrieved earlier by *Select Patients* queries).

In the current version, the VISITORS system does not recommend which concepts to select, nor the time periods in which to examine their, nor group of patients. However, as we explain in the Discussion, we intend to combine the VISITORS systems with pure computational tools (that we have been developing) for detection of sufficiently common temporal associations.

3.1 Temporal Association Templates

A Temporal Association Template (TAT) is an ordered list of *time-oriented concepts* (TOCs) (|TOCs|≥2), where each TOC denotes a combination of a raw or derived domain concept (such as a hemoglobin value or a bone-marrow toxicity grade) and a time interval $<t_{start}, t_{end}>$. Each TAT is thus a list L_{TOCs} of TOCs, that is:

$$L_{TOCs} = <TOC_1, ...TOC_i, ... TOC_I>, \forall i, 1 = i = I, \quad TOC_i \equiv <C_i, t_i^{start}, t_i^{end}>,$$

where $C_i \in C$ (the set of domain concepts). The time stamps t_i^{start} and t_i^{end} define the time interval of concept C_i (using either absolute or relative time stamps). I is the number of concepts in the TAT. A concept can appear more than once in the TAT, but only within different time intervals. An example of a TAT listing the hemoglobin-state and WBC-state abstract (derived) concepts, and the platelet-count raw-data concepts, and their respective time intervals, would be <(Hemoglobin-state, 1/1/95, 31/1/95), (WBC-state, 1/1/95, 31/1/95), (Platelet count, 1/1/95, 31/1/95), (WBC-state, 1/2/95, 28/2/95)>. Note that once a TAT is defined, it can be applied to different patient groups.

At runtime, a relation $<val_i, val_{i+1}>$ will be created between each pair TOC_i and TOC_{i+1}, *for each patient*, such that the delegate value of concept C_i for that patient during $[t_i^{start}, t_i^{end}]$ is a value val_i and the delegate value of concept C_{i+1} for that patient during $[t_{i+1}^{start}, t_{i+1}^{end}]$ is val_{i+1}.

3.2 Application of a Temporal Association Chart Template to the Set of Patient Records

When applying a TAT to a set P of patient records including N patients, we get a *Temporal Association Chart* (TAC). A TAC is a list of instantiated TOCs and association relations (AR), in which each instantiated TOC is composed of the original TOC of the TAT upon which it is based, and the patient-specific delegate values for that TOC within its respective time interval, based on the actual values of the records in P. Each set of the associations denotes the associations between pairs of consecutive instantiated TOCs $<TOC_i, TOC_{i+1}>$, $1 \le i < I$. To be included in a TAC, a patient P_n $(1 = n = N)$ must have at least one value from each TOC of the TAT defining the TAC. The group of such patients is the *relevant group* (relevant patients).

In the resulting TAC, each instantiated TOC^*_i includes the original TAT TOC_i and the set of delegate values (one delegate value for each patient) of the concept C_i, computed using the delegate function appropriate to C_i from the set of patient data included within the respective time interval $[t_i^{start}, t_i^{end}]$ as defined in the TAT:

$$TOC^*_i \equiv \{<P_n, val_n^i>, [Dist]\}, \quad 1 \le n \le N, 1 \le i \le I,$$

where val^i_n is the delegate value of C_i within the period $[t_i^{start}, t_i^{end}]$ for patient P_n, N is the number of patients in the relevant group, and I is the number of concepts in the TAT. *Dist* is an optional distribution data structure $\{<val^i_l, >Prop^i_l>\}$, where val^i_l is the l-th value of concept C_i, and $Prop^i_l$ is its proportion within the group of patients P. The optional *Dist* structure is useful only for abstract concepts and supports the visualization of the relative proportion (i.e, distribution) of all the values of C_i for the N relevant patients, within the time interval of the instantiated TOC^*_i (see Section 4).

Given a pair of instantiated TOCs $<TOC^*_i, TOC^*_{i+1}>$, the set of association relations (AR) between them is:

$$AR \equiv \{<P_n, val^i_n, > val^{i+1}_n>\}, 1 \le n \le N, 1 \le i < I$$

where N is the number of relevant patients, and I is the number of concepts in the TAC.

When at least one of the concepts is raw, the number of ARs between each pair of TOCs is equal to the number of relevant patients. Each AR connects the delegate values val^i_n and val^{i+1}_n of the pair of concepts C_i and C_{i+1}, during the relevant period of each concept, for one specific patient P_n.

In the case of an abstract-abstract concept pair, we aggregate the ARs between two consecutive TOCs into groups, where each group includes a set of identical pairs of delegate values (one pair for each concept). Each such group denotes a *temporal association rule* (TAR) and includes:

- **Support:** the proportion of patients who have the combination of delegate values $<val^{i,j}_n, val^{i+1,k}_n>$, $1 \le j \le J$, $1 \le k \le K$, where $val^{i,j}_n$, $val^{i+1,k}_n$ are the j-th and k-th allowed values of C_i and C_{i+1}, respectively. J and K are the numbers of different values of the concepts C_i and C_{i+1}. (We assume a finite number of (symbolic) values for each abstract concept.)
- **Confidence:** the fraction of patients that, given the delegate value $val^{i,j}_n$ of the concept C_i for patient P_n, the delegate value of concept C_{i+1} will be $val^{i+1,k}_n$, i.e., $P[val^{i+1,k}_n | val^{i,j}_n]$.
- **Actual number of patients:** the number of patients who have this combination of values.

The number of possible TARs between two consecutive TOCs is thus $J * K$. The support and confidence measures are calculated as follows:

$$\text{support}\,(val^{i,j}_n, val^{i+1,k}_n) \equiv \{P^{i,j}{}_{i+1,k}\}|/N$$

$$\text{confidence}\,(val^{i,j}_n, val^{i+1,k}_n) \equiv |\{P^{i,j}{}_{i+1,k}\}|/M$$

$$1 \le n \le N \qquad 1 \le i < I \qquad 1 \le j \le J \qquad 1 \le k \le K$$

where $|\{P^{i,j}_{i+1,k}\}|$ is the number of patients whose delegate value for concept C_i was $val^{i,j}_n$ and the delegate value for concept C_{i+1} was $val^{i+1,k}_n$, N is the number of relevant patients, *M* is the number of patients whose delegate value for concept C_i was $val_n^{i,j}$, *I* is the number of concepts in the TAC, and *J* and *K* are the number of symbolic values of concepts C_i and C_{i+1}, respectively.

4. Display of TACs and Interactive Data Mining Using TACs

►Figure 2 presents an example of a TAC computed by applying a TAT (*user-defined* on the fly, using another interface (not shown here) that enables the user to select TAT concepts). The TAT includes three hematological concepts (platelet state, hemoglobin (HGB) state abstractions, and the white blood cell (WBC) count raw concept) and two hepatic concepts (total bilirubin and Alk-phosphatase state abstraction) applied to a group of 58 patients *selected earlier by the user*. The visualization in Figure 2 shows the distribution of the values, using the optional *Dist* structure (see Section 3.2), for the abstract concepts HGB, Platelet, and Alk-phosphatase states; it also shows each patient's mean values for WBC count and total bilirubin during the year 1995. Delegate values of all adjacent concept pairs for each patient are connected by lines, denoting the ARs. Only 49 patients in this particular group happen to have data for all concepts during 1995.

As described above, ARs among values of temporal abstractions provide additional statistical information. For example, the AR's width indicates to the user the support for each combination of values, while the color saturation represents the level of confidence: a deep shade of red signifies high confidence while pink denotes lower confidence. The support, confidence, and the number of patients in each association are displayed numerically on the edge. For example, the widest edge in Figure 2 represents the relation between the "low" value of the platelet state and the "moderately low" value of the HGB state during the respective time periods: 55.8% of the patients in the relevant patient group had this combination of values during these periods (i.e. support = 0.558), with "low" platelet state values, 92.6% of the patients exhibited "moderately low" HGB state values (i.e. confidence = 0.926), and this association was valid for 25 patients. Note that in this case both time periods are similar.

Using *direct manipulations* [15], the user can dynamically apply a *value* and *time lens* in the TACs. Generally, the term *direct manipulation* is defined as a human-computer interaction style which involves continuous representation of objects of interest, and rapid, reversible, incremental actions and feedback. In our case, the direct manipulations enable the user to interactively analyze the time and value associations among multiple patients' data:

- Dynamic application of a *value lens* enables the user to answer the question "*how does constraining the value of one concept*

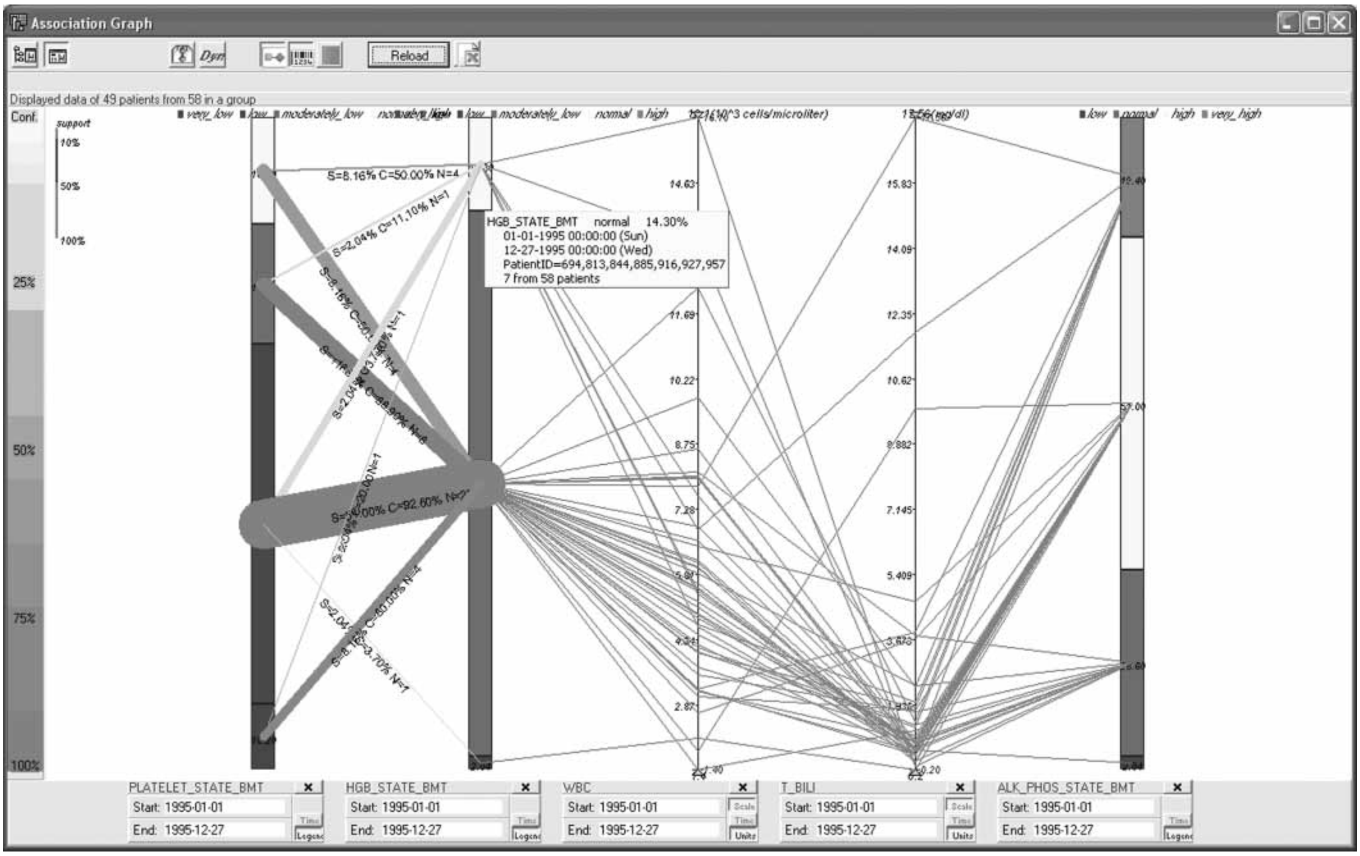

Fig. 2 Visualization of associations among three hematological and two hepatic concepts for 49 patients during the year 1995. Association rules are displayed between the Platelet-state and HGB-state abstract concepts. The confidence and support scales are represented on the left.

during a particular time period affect the association between multiple concepts during that and/or during additional time periods". The user can either select another range of values for the data of the raw concepts using trackbars, or select a subset of the relevant values in the case of an abstract concept. In the next versions, we are planning to allow the user to vary also the delegate function to enable additional analyses.
- The system also supports the application of a *time lens* by changing the range of the time interval for each instantiated TOC, including ranges on the relative time line. The time lens can be especially useful for clinical research involving longitudinal monitoring.

In addition, the user can change the order of the displayed concepts, export all of the visualized data and associations to an electronic spreadsheet, and add or remove displayed concepts.

5. Evaluation of the Functionality and Usability of Temporal Association Charts

5.1 Research Questions

We envision TACs as potentially useful for two user types: clinicians and medical informaticians. We also envision them using the system to answer different clinical motivated questions, while using also the general exploration operators of the VISITORS system. Thus we defined the following **three** research questions.

5.1.1 Functionality and Usability

Are clinicians and medical informaticians able to answer *clinical questions*, which require the use of TACs, at a high level of accuracy and within a reasonably short time? Furthermore, is the integrated VISITORS/TAC system usable, when assessed using the SUS score [16]?

5.1.2 The Effect of the Clinical Question

Are there significant differences in accuracy and time when answering different clinical questions that required the use of TACs?

5.1.3 The Effect of the Interaction Mode

Are there significant accuracy or time to answer differences when answering questions requiring only the use of general VISITORS exploration operators, as opposed to answering questions requiring the use of TACs?

5.2 Measurement Methods and Data Collection

When evaluating a new tool such as TACs, it is difficult to produce a control group. As far as we know, no known method duplicates the effect of using either VISITORS or TACs. Furthermore, the potential users simply cannot answer the complex questions (for which TACs are designed) other than by laborious computations. Thus, we have chosen an *objective-based approach* [17]. In such an approach, certain reasonable objectives are defined for a new system, and the evaluation strives to demonstrate that these objectives have been achieved. In this case, we strove to prove certain functionality and usability objectives of the TACs system when evaluated within the context of a larger framework (VISITORS) for exploration of the time-oriented data of multiple patients. Our evaluation measures and specific research questions, listed below, reflected these objectives.

The evaluation of the TACs was performed in the oncology domain. Ten participants, five medical informaticians, i.e., information system engineers who work in the medical domain, and five clinicians with different medical training levels, were each asked to answer five clinical questions that require using TACs (five questions are listed in the Results section). None of the study participants was a member of the VISITORS development team. The five questions were selected by consultation with oncology domain experts. They represented typical questions relevant when monitoring a group of oncology patients, or when performing an analysis of an experimental protocol in an oncology. The order of the questions was randomly permuted across participants.

Each evaluation session with a participant started with a 20-minute tutorial that included a brief description of the VISITORS general exploration operators and of the TAC operators. A demonstration was given of the general and TAC operators showing how several typical clinical questions are answered. The scope of the instruction was predetermined and included (after the demo) testing each participant by asking them to answer three clinical questions, one of which included the use of TAC. When the participant could answer the questions correctly, he/she was considered ready for the evaluation.

The TACs evaluation study was performed as a part of an overall feasibility and usability assessment of the VISITORS system. Another part of the evaluation involved testing the feasibility and usability of the general exploration operators by asking clinical questions such as "What were the maximal and mean monthly values of the WBC count during August 1995?".

Throughout the evaluation, we used a retrospective database of more than 1000 oncology patients who had a BMT event. The knowledge source used for the evaluation was an oncology knowledge base specific to the bone-marrow transplantation domain.

Our goals for the TACs objectives-based evaluation were manifested in our evaluation measures.

The *functionality* was assessed using two parameters: the time in minutes needed to answer the question, and the accuracy of the resultant answer. The accuracy score assigned to each *possible* answer (in each case, measured on a scale of [0 ... 100], 0 being completely wrong and 100 being completely right), was predetermined by a medical expert.

To test the *usability* of TACs, we used the system usability scale (SUS)[16], a common validated method to evaluate interface usability. The SUS is a questionnaire that includes ten predefined questions regarding the effectiveness, efficiency, and satisfaction of an interface. SUS scores have a range of 0 to 100. Informally, a score higher than 50 is considered to indicate a usable system.

5.3 Analysis Methods

5.3.1 Functionality and Usability

User capability in answering clinical questions using the TACs was assessed by calculating the means and standard deviations of the answers accuracy and of the answers response time.

5.3.2 The Effect of the Clinical Question

The effects of five clinical questions and of two groups of participants (i.e., medical informaticians and clinicians) on the dependent variables of response time and accuracy of answer were examined using two different two-way ANOVA tests with repeated measures (one for each dependent variable). The clinical question was a within-subject independent variable, and the group of participants was a between-subjects independent variable.

5.3.3 The Effect of the Interaction Mode

The effects of the interaction mode (i.e., general exploration operators and TACs) and the group of participants (i.e., medical informaticians and clinicians) on the dependent variables of response time and accuracy of answer were examined using two different two-way ANOVA tests with repeated measures (one for each dependent variable). The interaction mode was a within-subject independent variable and the group of participants was a between-subject independent variable. Since we did not find statistically significant differences among the response times (and among the resultant accuracy levels) of the different clinical questions of the same interaction mode, the mean value of the response time (and of the accuracy) of the five clinical questions of the same interaction mode was used as the dependent variable.

5.4 Results

This section summarizes the evaluation results of the TAC in terms of the research questions.

5.4.1 Functionality and Usability

►Table 1 summarizes the results according to the clinical questions used in the evaluation. The mean accuracy was 97.9 ± 3.4 (the median was 100 for all five questions, with interquartile range of zero for questions one to four and 10 for question five). All participants successfully answered the clinical questions with mean accuracies greater than 90, while six of them achieved mean accuracies of 100. The range of mean accuracies per participant across all questions was [90 ... 100].

The mean response time was 2.7 ± 0.4 minutes. Only two participants needed more than 3 minutes (3.2 and 3.6 minutes) to answer. The range of mean response times per participant across all questions was [2.2 ... 3.0] minutes.

The mean SUS score for all operators, across all participants, was 69.3 (over 50 is usable). The results of a t-test analysis showed that the mean SUS score of the medical informaticians (80.5) was significantly higher than that of the clinicians (58): [$t(8) = 3.88$, $p < 0.01$].

Conclusion: Based on the results of the TAC evaluation, we can determine that after a very short training period, the participants were able to answer the clinical questions with very high accuracies and within short periods of time. The SUS scoring shows that TACs are usable but still need to be improved.

5.4.2 The Effect of the Clinical Question

Both analyses yielded no significant effect. The interaction effect group × clinical question was not significant both in the ANOVA of the accuracy scores [$F(4, 32) = 2.02$, $p = 0.12$], and of the response times [$F(4, 32) = 0.73$, $p = 0.58$]. There was no significant difference between the mean accuracy scores/ response times of the clinicians (96.3 ± 4.3)/ (2.6 ± 0.2 min) and of the medical informaticians (99.5 ± 0.9)/ (2.9 ± 0.5 min); accuracy: [$F(1, 8) = 2.72$, $p = 0.14$, regression coef-

Table 1 Details of the questions that all participants had to answer, mean response times (minutes) and accuracy scores

N	Clinical questions	Response time (mean ± s.d.)	Accuracy score [0..100] (mean ± s.d.)
1	What percentage of the patients has had a "low" delegate value of the Platelet-state concept? What percentage of the patients has had a "moderately low" delegate value of the HGB-state concept? What percentage of the patients has had both a "low" value of the Platelet-state concept and a "moderately low" value of the HGB-state concept?	2.4 ± 0.5	99.9 ± 0.3
2	What delegate value of the HGB-state derived concept was the most frequent among the patients who have had a "low" aggregate value of the Platelet-state?	2.4 ± 0.5	99.8 ± 0.6
3	What were the maximal and minimal delegate values of the WBC count for patients who have the HGB-state delegate value "moderately low"? What were the maximal and minimal delegate values of the RBC for patients who have had a delegate HGB-state value that was "normal"?	2.9 ± 1.0	96.0 ± 9.0
4	What is the distribution of the delegate values of the Platelet-state in patients whose minimal delegate value of the WBC count raw concept was 5000 cells/ml (instead of the previous minimal value)? What were the new maximal and minimal delegate values of the RBC?	3.0 ± 0.7	98.8 ± 4.0
5	What percentage of the patients has had a "low" delegate value of the Platelet-state during both the first and second month following bone-marrow transplantation?	3.0 ± 0.8	96.0 ± 5.0

ficient ± sd = 3.2 ± 1.95]/time: [$F(1, 8) = 1.90$, $p = 0.20$, regression coefficient ± sd = 0.3 ± 0.26]. There was no significant difference also between the mean accuracy scores/response times of the five clinical questions (for mean accuracy ± sd/response times ± sd; see Table 1); accuracy: [$F(4, 32) = 2.29$, $p = 0.08$, maximal estimated effect ± sd = 4.8 ± 3.1]/time: [$F(4, 32) = 2.24$, $p = 0.09$, maximal estimated effect ± sd = 0.6 ± 0.33].

Conclusion: Thus, we can conclude that neither the clinical question nor the group seem to affect the accuracy or the response time of answers provided using the TAC module.

5.4.3 The Effect of the Interaction Mode

The results of the ANOVA of the accuracy scores showed that the interaction effect group × interaction mode was not significant [$F(1, 8) = 2.64$, $p = 0.14$]. There was no significant difference between the mean accuracy scores of the clinicians (97.7 ± 3.0) and of the medical informaticians (99.8 ± 0.4); [$F(1, 8) = 2.37$, $p = 0.16$, regression coefficient ± sd = 2.1 ± 1.4]. There was no significant difference also between the mean accuracy scores of the answers obtained by using the general exploration operators (99.5 ± 1.6) and by using the TACs (97.9 ± 3.4); [$F(1, 8) = 4.80$, $p = 0.07$, regression coefficient ± sd = 2.3 ± 1.6].

With respect to the response time, the results of the analysis showed that the only significant effect was the main effect of the type of interaction mode [$F(1, 8) = 14.96$, $p < 0.01$]: a mean of 2.2 ± 0.18 minutes for answering the clinical questions when using the general exploration operators of VISITORS, and a mean of 2.7 ± 0.43 minutes for answering the clinical questions using the TACs. There was no significant difference between the mean response times of the clinicians (2.6 ± 0.2 min) and the medical informaticians (2.9 ± 0.5 min); [$F(1, 8) = 3.24$, $p = 0.11$, regression coefficient ± sd = 0.3 ± 0.16]. The interaction effect group × interaction mode was also insignificant [$F(1, 8) = 0.42$, $p = 0.54$].

Conclusion: Interaction mode does not seem to affect the accuracy of the answers to clinical questions. The mean time needed to answer the clinical scenarios using the TACs is significantly higher than when using the general exploration operators of VISITORS, but it is still less than three minutes.

5.5 Results of Power Analysis

Since we have not found significant effects in the results of research questions 2 and 3, we performed a statistical power analysis.

For each two-way ANOVA test in the results of research question 2 (for each dependent variable), we performed a power analysis, both for each main effect, namely, the five clinical questions and the two groups of participants, and for the interaction effect (questions X groups).

In the case of research question 3, the main effects for which the power analysis was performed were the two interaction modes (general exploration operators vs. TACs) and the two groups of participants, and for the interaction effect (modes X groups).
The results showed that, in the case of accuracy, assuming a meaningful difference being at least 5 points on a scale of (0 ... 100), the experiment (i.e., $N = 10$, $\alpha = 0.05$) would detect an effect (i.e., a difference between the means of the groups) with a probability of at least 80%, which is considered a reasonable power. Similarly, in the case of response time, assuming a meaningful difference being at least one minute.

Moreover, the power analysis had shown that with a larger group, including 20 participants (10 clinicians and 10 medical informaticians), a smaller effect of only two points in the mean accuracy (the minimal effect size obtained in our analysis) could be detected with a probability of 80%. However, we consider a difference of two points or less to be relatively insignificant for practical purposes. Similarly in the case of response times for an effect size of 0.3 min with a sample size of 54 participants (26 clinicians and 26 medical informaticians).

6. Discussion

This paper presents the Temporal Association Chart, a computational and interaction module that enables users to graphically explore and analyze the time and value associations among domain concepts that explicitly or implicitly exist within multiple time-oriented patient records. Moreover, it enables the exploration of 1) intelligent interpretations (temporal abstractions) of the raw data derived using the context-sensitive domain-specific knowledge, and 2) temporal aggregations of the patient data summarized within several specific time periods (including the use of temporal granularities) using delegate functions to each concept and each temporal granularity. The associations displayed between pairs of consecutive abstract concepts include *support* and *confidence* measures that can be interactively investigated via manipulation by the user. Note that when the time periods of pair of concepts are the same, these measures are an interval-based extension of the familiar data mining measures, using delegate functions. When the time periods are different, an extension of temporal association rules, commonly called *sequence mining*, emerges, again using delegate functions, which allows multiple time granularities and which does not necessitate the simultaneous existence of different concepts (known as *items*).

The evaluation of the TAC module, which is integrated within the VISITORS system, has demonstrated its functionality and usability. The only significant difference between the TACs and the general exploration operators was a slightly longer response time. A possible reason for not detecting significant differences in the accuracy scores when using different interaction modes was that the evaluation included a relatively small group of participants and questions. However, it should be noted that the variance among accuracy scores was quite low for both interaction modes, all of the participants achieving scores above 90. Thus, the absence of a significant effect could not be attributed to random differences and high variability in each interaction mode. This conclusion was also supported by the results of the power analysis. Although applying ANOVA in the context of a ceiling effect in the accuracy scores is potentially problematic, this phenomenon, in our judgment, did not have the significant effect on the conclusions of the study study.

One possible conceptual limitation of the TAC approach is the use of a *goal-directed* (user-driven) method for temporal data mining. Thus, the user must have a meaningful intuition regarding the selection of the necessary concepts to explore. However, this limitation can be overcome by combining the TAC module with a knowledge-based temporal data mining method, such as the one we have been developing [18]. Associations that have sufficient support are automatically flagged, and in the future, could be visually explored.

To summarize, we conclude that TACs might be described as "intelligent equalizers" that result in a uniform performance level with respect to answering complex time-oriented clinical statistical-aggregation questions, regardless of the questions asked, or of the user type.

Acknowledgments

This research was supported by Deutsche Telekom Labs at Ben Gurion University and the Israeli Ministry of Defence, BGU award No. 89357628-01.

References

1. Bonneau G, Ertl T, Nielson G. Scientific Visualization: The Visual Extraction of Knowledge from Data. New York: Springer-Verlag; 2005.
2. Soukup T, Davidson I. Visual Data Mining: Techniques and Tools for Data Visualization and Mining. New York: John Wiley & Sons, Inc.; 2002.
3. Spenke M. Visualization and interactive analysis of blood parameters with InfoZoom. Artificial Intelligence in Medicine 2001; 22 (2): 159–172.
4. Wang T, Plaisant C, Quinn A, Stanchak R, Shneiderman B, Murphy S. Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records. SIGCHI Conference on Human Factors in Computing Systems, 2008.
5. Chittaro L, Combi C, Trapasso G. Visual Data Mining of Clinical Databases: An Application to the Hemodialytic Treatment based on 3D Interactive Bar Charts. Proceedings of Visual Data Mining VDM'2002, Helsinki, Finland, 2002.
6. Aigner W, Miksch S, Müller W, Schumann H, Tominski C. Visual Methods for Analyzing Time-Oriented Data. IEEE Transactions on Visualization and Computer Graphics 2008; 14 (1): 47–60.
7. Shahar Y. A framework for knowledge-based temporal abstraction. Artificial Intelligence 1997; 90 (1–2).
8. Stein A, Shahar Y, Musen M. Knowledge Acquisition for Temporal Abstraction. 1996 AMIA Annual Fall Symposium, Washington, D.C. Published in 1996.
9. Chakravarty S, Shahar Y. Acquisition and Analysis of Repeating Patterns in Time-oriented Clinical Data. Methods Inf Med 2001; 40 (5): 410–420.
10. Shahar Y, Goren-Bar D, Boaz D, Tahan G. Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. Artificial Intelligence in Medicine 2006; 38 (2): 115–135.
11. Martins S, Shahar Y, Goren-Bar D, Galperin M, Kaizer H, Basso LV, McNaughton D, Goldstein MK. Evaluation of an architecture for intelligent query and exploration of time-oriented clinical data. Artificial Intelligence in Medicine 2008; 4 (3): 17–34.
12. Klimov D, Shahar Y. Intelligent querying and exploration of multiple time-oriented medical records. MEDINFO Annu Symp Proc 2007; 12 (2): 1314–1318.
13. Klimov D, Shahar Y. A Framework for Intelligent Visualization of Multiple Time-Oriented Medical Records. AMIA Annu Symp Proc 2005. pp 405–409.
14. Falkman G. Information visualisation in clinical Odontology: multidimensional analysis and interactive data exploration. Artificial Intelligence in Medicine 2001; 22 (2): 133–158.
15. Shneiderman B, Plaisant C. Designing the user interface: strategies for effective human-computer-interaction. 4th edition. Addison Wesley; March 2004.
16. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. Taylor and Francis; 1996.
17. Friedman C, Wyatt J. Evaluation Methods in Medical Informatics. New York: Springer; 1997.
18. Moskovitch R. and Shahar Y. Temporal Data Mining Based on Temporal Abstractions. ICDM-05 workshop on Temporal Data Mining, Houston, US. 2005.

