A temporal analysis of QMR: abstracted temporal representation and reasoning and initial assessment of diagnostic performance trade-offs

Constantin F. Aliferis, M.D., Gregory F. Cooper, M.D., Ph.D., and Richard Bankowitz, M.D.*
Section of Medical Informatics & Intelligent Systems Program
University of Pittsburgh, Pittsburgh PA
*University Hospital Consortium, Oak Brook, IL

ABSTRACT

Explicit temporal representation and reasoning (TRR) in medical decision-support systems (MDSS) is generally considered to be a useful but often neglected aspect of system design and implementation. Given the great burden of explicit TRR both in knowledge acquisition and computational efficiency, developers of general-purpose large-scale systems typically utilize implicit (i.e., abstracted) forms of TRR. We are interested in understanding better the trade-offs of not incorporating explicit TRR in large general-purpose MDSS along the dimensions of system expressive power and diagnostic accuracy. In particular, we examine the types of abstracted TRR employed in QMR, a diagnostic system in the domain of general internal medicine, and the high-level effects of such an implicit treatment of time in the system's diagnostic performance. We present our findings and discuss implications for MDSS design and implementation practices.

INTRODUCTION

In a review of TRR in MDSS, Kahn [1] proposes an empirical classification which separates systems in two main categories, one based on *temporal ignorance*, and one on explicit representation and utilization of temporal concepts. More specifically he demonstrates how the earlier systems overrode the need for explicit Knowledge Representation and Reasoning (KRR) by incorporating temporal information into ordinary atemporal formalisms. For instance, the INTERNIST-I system would ask questions of the type "did the patient have a history of disease x?", that clearly correspond to an *implicit* or *abstracted* form of TRR. It is obvious that the systems' developers were operating on the assumption that the user of the system would abstract relative data from historical observations and provide it to the program.

Explicitness in TRR practically entails the existence of two main components. First we need a model of time (i.e., a description of temporal primitive entities from which time is composed, plus a set of properties of time, as for example linearity and finiteness). Second, we need an association of entities (objects and relations) to the model of time, such that we can reason about them in one or more temporal contexts, and represent/infer useful temporal knowledge. For all but the most trivial medical domains the use of temporal ignorance is a heuristic approximation to modelling time explicitly. It was applied in a number of influential systems like MYCIN, PIP, DxPlain, CASNET, and ABEL [2]. This way, the developers of MDSS were able to avoid explicit reasoning about temporal concepts, an area within which AI was not well developed in the 70's, or even the early 80's.

On the other hand, a well-known problem in KRR is what Levesque and Brachman call the "fundamental trade-off in KRR" [3]: *more expressive power generally means less computational tractability (and the opposite)*. Given the extensive burden of TRR in terms of efficiency, as well as knowledge acquisition, we believe that any effort to incorporate explicit TRR in large-scale MDSS should be well-justified in terms of expected gains in system performance and/or knowledge engineering (i.e., expressivity of a temporal model vs the atemporal one). In other words, we need to examine *why* and *how* important is the ability to reason explicitly (as contrasted to an abstracted manner) about temporal processes and entities. The importance of TRR has been considered more or less "obvious" and thus has been inadequately explored in the medical AI and medical informatics (MI) literature, especially with respect to *quantifying* its importance.

The basic arguments that have been offered in favor of the necessity of explicit TRR in medical DSSs are :

(a) The *epistemological argument*: observations of physicians diagnostic and therapeutic problem-solving suggest that temporal models of normal and abnormal processes are used, intricate temporal abstractions are created and used to generate and validate or rule-out competing hypotheses. Additionally, physicians are able to utilize temporal planning for either diagnosis (e.g., "watchful expectancy") or therapy [4].

(b) The *linguistic argument:* analysis of discharge summaries and other medical texts indicate an impressive amount of TRR [5].

(c) *Pragmatics argument #1*: certain medical domains

0195-4210/94/$5.00 © 1994 AMIA, Inc.

are based on the premise of a time-evolving process, and TRR is fundamental for them (characteristic examples include the protocol-based therapy management, ICU real-time monitoring and intervention, signal processing as in EKG and EEG interpretation) [6,7].

(d) *Pragmatics argument #2*: evaluation of DSSs diagnostic performance shows that some failures to reach the proper diagnosis is attributed to lack of TRR capabilities [8,9].

By carefully examining these arguments, we can make the following remarks: the epistemological and linguistic arguments are purely *descriptive* and do not justify directly the importance of TRR. The first pragmatics argument is certainly true, but refers to a clearly defined, limited subset of DSSs with few or no implications for the majority of systems that perform diagnosis/treatment selections in wide areas of medicine such as INTERNIST-I, MYCIN, etc. These latter systems' need for TRR could be substantiated by the second pragmatics argument, in the sense that, ceteris paribus, if TRR accounts for a substantial number of diagnostic failures and the problems can not be fixed in a reasonable way (i.e., by respecting KA and efficiency constraints), then we can conclude that TRR is indeed necessary.

Unfortunately, support for the second pragmatics argument comes in the form of anecdotal evidence rather than from planned experiments designed to prove or disprove the validity of this hypothesis. One famous example is the 1982 NEJM evaluation of INTERNIST-I (one of the most often cited instances of the second pragmatic argument), which on the basis of 3 cases (out of a total of 19 diagnostic problems) supposedly indicates that explicit TRR is indeed necessary. But the 99% confidence limits of 3/19 (16%) are between 2 and 47%, suggesting that no strong conclusion can be reached from this data regarding the effects of TRR. Even more importantly, the cases were not representative of the average encountered clinical case, since they were CPC cases that were selected on the basis of being very challenging [8].

The previous discussion indicates the need for further investigation and quantitative analysis of the importance of TRR in general-purpose DSSs. It should be clear on the other hand, that for a variety of DSS domains this need is well justified by the nature of the domain (i.e., the nature of the entities represented is so deeply temporal, that either we can not reason about it without taking into account time, or it is grossly ineffective to utilize some implicit/abstracted form of KRR). These domains/tasks include:

- Protocol therapy management,
- Biomedical signal processing,
- "Deep" causal models of diseases/physiology which are grounded on dynamic systems.
- Intensive care unit (ICU) decision-support.

Additionally, it might be more parsimonious and/or natural to represent some normal or abnormal processes in temporal terms (for instance hormonal cycles or compartmental models [6,7,9-11]).

The research presented here intends to investigate the following hypothesis: *In systems operating as aids to clinicians (as opposed to systems that are fully automated) the human operator of the system can provide the necessary abstraction reasoning from temporal entities to the system's atemporal knowledge representation language, such that the system would not have to explicitly incorporate TRR to achieve equal levels of performance as in the temporally explicit case.*

There are a number of additional interesting questions associated with this conjecture:

(i) What constitutes an appropriate collection of abstracted (atemporal) knowledge representations, corresponding to the domain to be modelled?

(ii) Are there specific temporal entities that are crucial to DSS performance? What is the proper level of description of those entities?

(iii) How would these results be useful for systems that operate in automated mode?

In the following sections, we investigate this hypothesis and the related questions by means of an analysis of QMR, a well known MDSS, operating in the domain of diagnosis for general internal medicine, along the dimensions of :

- expressive power (what types of abstracted TRR can be handled by the system)
- performance degradation due to temporal complexities of patient cases.

METHODS

1. Temporal analysis of QMR's terms

We devised a series of variables that correspond to what previous theoretical and empirical work suggests are important temporal reasoning and representation attributes [12-17]. These were used by the first author to classify each finding in QMR as temporal or not based on the following criterion. A QMR finding is temporal if *any* of the following is true: *explicit* reference to either time points/intervals or units, temporal relationships/reasoning, events or facts described in some temporal context, processes occurring over time (explicitly static/evolving, or in sequence/overlapping), or patterns (temporal or spatio-temporal).

We additionally recorded the QMR type (history, physical, simple-inexpensive lab, intermediate lab, advanced-expensive lab) and importance (the "import" value of QMR indicating "need for a finding to be explained if found" [8]), for all findings,

regardless of temporal nature. Temporal types (I and II), ontology, and temporal reasoning were created empirically in an incremental fashion, and refined as new QMR findings were examined. Temporal types-II correspond to simple abstractions over QMR findings. A temporal type-I provides a way to describe covariation of variables (i.e., it is a temporal pattern). In essence it is an abstraction over temporal types-II. Reasoning types, on the other hand, denote fundamental relations and other properties that can be put together to form temporal types I&II (for examples of temporal types, and reasoning types see results sections 1.2-1.4). To ensure consistency in the categorization of temporal QMR findings (according to temporal type, and reasoning type) the following procedure was followed: first temporal findings were identified. In the next step values for the variables for *each temporal type*-II were assigned. Due to the limited number of types-II (<120), consistency checks (with previously established temporal types-II) were easier and less error-prone to carry out than the full set of temporal QMR findings. After the types-II had been characterized, individual findings were categorized as belonging to any of specific temporal types-II. As a consequence, each finding would inherit the variable value assignments of the corresponding abstract type-II. As a final step, each individual finding was examined for differences with the type-II it belonged to (due to the abstraction process), and the necessary adjustments were made to the deviating attributes of the individual findings. Then, temporal types-I and temporal reasoning types were abstracted and classified empirically. Standard descriptive statistics were computed for all variables. Bivariate associations were examined with Likelihood Ratio (G^2) tests of independence, Kendall's tau and the gamma coefficient (for ordinal variables). Multivariate relations were examined with the previous statistics controlling for possible confounders [18].

2. Effects of lack of explicit TRR on diagnostic performance

Ideally we would like to test the following (null) statistical hypothesis: *Lack of explicit TRR in QMR does not cause decreased diagnostic performance, (compared to the case where explicit TRR is employed).* Figure 1 presents an idealized experiment built around a post-test design [19] in which the same group of cases is presented to the system. Assuming that the diagnostic system has explicit TRR and that it can be turned on and off at will, diagnoses are performed twice, once with TRR being active and once with TRR being inactive. The performance in the first case is compared to the second one. Obviously this best-case experiment is unattainable. There is no MDSS employing explicit TRR that operates with a scope comparable to that of general internal medicine. Nor is TRR typically implemented in a manner that can be turned on and off, leaving the system reliability intact.

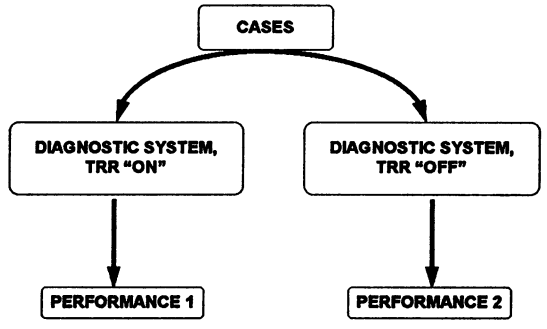

Figure 1: An idealized experiment

Since modifying QMR to incorporate an explicit TRR model is equally infeasible for the purposes of this study, we designed a modified version of the previous experiment, represented in figure 2.

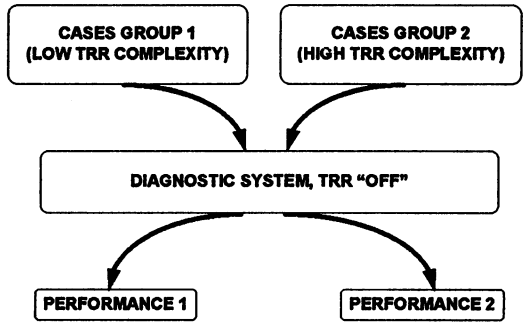

Figure 2: A modified experiment

In this second design two groups are presented to the system. The two groups are similar in every aspect that we would expect to affect performance, except for temporal content. We would like to compare the performance of the system in the two groups. All other things being equal, one would attribute any diagnostic performance discrepancy to the temporal content of the cases (and equivalently to the lack of explicit TRR by QMR). Note that it is important to maintain a prospective design, to avoid a case-control setup (and the associated potential biases with respect to identifying the risk factor, and establishing case-control comparability) [20]. Important considerations in the execution of this design are:

(a) Selection of a representative sample of actual patient cases. We used 105 cases from the latest formal evaluation of QMR (each consisting of history, physical, initial laboratory tests, and discharge summary and diagnoses). The third author (R.B.) is the primary investigator in that study. The coding of the patient information was done by experienced

QMR users, and the cases are considered to be representative of the cases admitted to a large university hospital.

(b) Potential confounders: We considered, and controlled for analytically, several potential confounders (rareness of primary diagnosis, length of case, uncertainty in the description of evidence, spatial information references, causal information references, multiple levels, and number of diseases in true diagnosis).

(c) TRR assessment: The history and physical (H&P) portion of each patient case was separated into a number of individual pieces of information (POI). A POI was defined as the smallest piece of clinically relevant information that could be meaningful if stated in the given document context (thus a POI could be either a stand-alone statement or a qualification of a previously established statement). Each POI was characterized as temporal or not based on the criteria described for QMR findings. For each POI the values of the confounder attributes were assessed. The percentage of temporal POIs divided by the total number of POIs in the case, constituted our measure of TRR content for that case. For each POI the temporal attributes utilized in the assessment of explicit TRR (methods section 1) were evaluated and summarized for each case. We utilized principal components analysis to identify summary linear combinations of those measures as more detailed metrics of the case temporal content. Similar measures of complexity and temporal content were assessed for the QMR encodings for each case. Finally we identified TRR types in the cases that exceeded the expressive capacity of the QMR abstractions (these are discussed in [21]). The temporal attributes' value assignments (for both cases and QMR inputs) were blinded with respect to the case outcome.

(d) Performance assessment: Our criterion was the percentage of cases for which QMR found the primary diagnosis. The following matching criterion was used:

- A match occurred iff the gold standard (GS) primary diagnosis is clinically *equivalent* to one of the q (%) first diagnoses in the QMR differential diagnosis (DD) list. The primary discharge diagnosis (ICD9 primary diagnosis) was considered to be the GS.
- q is defined to be a percentage of diagnoses from QMR's differential diagnosis list. As discussed later in the results section, we chose a q that gave us a mathematically convenient diagnosis rate in the assessment of the effects of TRR content of each case.
- Equivalence is one of the following: *identity, synonymy,* or a *close match.*
- Close match was taken to mean a significantly overlapping disease category or a disease which is at most one level down or up in a recognized clinical classification as those found in major textbooks of medicine (e.g., Harrison's or Cecil's textbook of medicine etc.).

Cases with no established true diagnoses were excluded. When the first (primary) diagnosis in the GS differential was asserted, or given as a finding in QMR, or was not in QMR's KB, the next diagnosis would be the primary one (with a recursive application of the exclusion and skipping rules).

(e) Analysis: All variables associated potentially with diagnostic performance were descretized (based on their 50th percentile as a single cutoff point). Odds-ratios of correct diagnoses were computed between the explanatory variable categories [18].

Logit models (using the continuous variable versions and a standard statistical package) and Bayesian models (through the application of the K2 inductive learning algorithm) were built to assess quantitatively the impact of TRR in the cases to the system's diagnostic accuracy. The interrelations of TRR content and the rest of the explanatory variables were also examined with respect to diagnostic accuracy [18,22].

RESULTS

1. Temporal analysis of QMR's terms

1.1. Ontology

It was found that QMR utilizes the following temporal ontology to express temporal findings.

(a) Entities :

(i) Generic: disease, syndrome, finding, symptom, laboratory value, test result, medical procedures, drugs, causal factors, diagnostic factors.

(ii) Temporal: periods, points of time, seasons, parts of the day, disease intervals, EKG - related intervals, systolic/diastolic periods, units of time.

(b) Relationships/Properties: History of, during, before, after, coincides with, repeating, properties (frequency, speed, rhythm, regularity).

Also Boolean combinations of the above are used to derive more complex propositions.

1.2. Temporal types-I

Forty-nine different TRR types-I were found (and combinations of those). Table 1 presents the most frequent ones, with frequencies.

1.3. Temporal types-II

A total of 116 temporal types-II were found to be employed by QMR within its findings. Table 2 presents the most frequent ones, together with frequencies (% of total number of temporal findings).

1.4. Temporal reasoning

A total of 20 different temporal reasoning types were identified (as well as combinations of those). Table 3 presents the most frequent ones, and their frequencies, together with the relation that they

correspond to.

Table 1: Most frequent of temporal types-I (higher level of abstraction over QMR findings)

(1) Hx of disease prior to current one [19.5%]
(2) Hx of event/factor exposure/finding before disease [15.8%]
(3) Event/factor exposure/finding before disease [13.0%]
(4) Hx of remote event/factor exposure/finding before disease [8.2%%]
(5) Finding during period [5.6%]
(6) Finding/disease after drug/medical procedure [4.8%]
(7) Increase/decrease in measurement [4.5%]
(8) Finding at onset/early/middle/late period (e.g., season) [3.5%]
(9) Hx of recurrent/chronic symptom/disease/ finding [2.7%]
(10) Abrupt onset [2.1%]

Table 2: Most frequent of temporal types-II (lower level of abstraction over QMR findings)

(1) Hx of syndrome/disease [11.2%]
(2) Hx of drug administration prior to current illness [8%]
(3) Improvement/worsening of function after/during test/medical procedure/state [7.3%]
(4) Hx of familial disease/behavior [7.2%]
(5) Abnormal/normal finding/syndrome after drug/medical procedure [4.4%]
(6) Hx of exposure to animals/factors [3.6%]
(7) Hx of recent medical procedure [3%]
(8) Hx of recent exposure to factor environment /food/behavior [2.7%]
(9) Increased/decreased rhythm/rate/speed [2.4%]
(10) Measurement per unit of time > C [2.3%]

Table 3: Most frequent temporal reasoning types (in parentheses the fundamental relation each type corresponds to)

(1) Hx of [42.6%] *(before)*
(2) Hx of recent [12.7%] *(before with qualitative distance)*
(3) Finding during period [8.5%] *(during)*
(4) Finding after or during event [8%] *(after or during)*
(5) Properties (e.g., duration etc.) of a temporal primitive [7.7%] *(arbitrary property)*

Figure 3 depicts a multiple-inheritance hierarchical classification of temporal types I & II that captures their main features. A similar classification was developed for temporal reasoning (not shown here).

1.5. Frequency and importance of temporal entities

Of all 4431 QMR findings 17.5% were classified as temporal. Table 4 presents frequency distributions for some of the findings attributes.

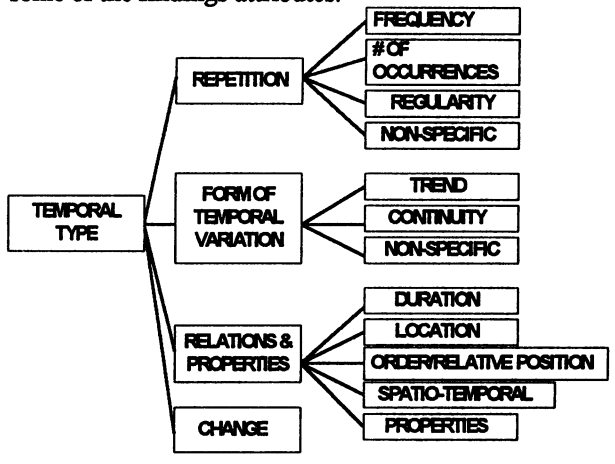

Figure 3: Temporal type (I & II) abstraction

Interesting associations include: Findings that reference temporal units have higher importance (G^2 p<.0001, gamma=-.51 with t-value=-3.6). Similarly when we have explicit reference to procedures or patterns, the importance is higher (G^2 p=.015 and G^2 p=.0005, gamma=.64 with t-value=3.44 respectively). Overall, however, temporal findings have less import than non-temporal ones (G^2 p<.0001, gamma=.65 with t-value23.2). Symptoms and signs had less import than more advanced lab findings (G^2 p<.0001, tau=0.11 with p<.0001). At the same time though, temporal findings are characterized by smaller values in the QMR TYPE scale of diagnostic sophistication (see methods) (G^2 p<.0001, tau=.49 with p<.0001). When we control for QMR TYPE, the relationship between temporality and importance vanishes.

Table 4: Frequency distributions for main attributes

Among all findings:
TEMPORAL : yes 17.5%, no 82.5%
QMR TYPE : history 11.5% , symptom 5 %, sign 25.4%, lab simple 6.5%, lab intermediate 30.7%, lab expensive/invasive 20.9%
IMPORT : low 2.3%, medium-low 15.8%, medium 35.5% medium-high 32.4%, high 14%
Among temporal findings only:
TIME PRIMITIVES : implicit 93.3 %, explicit points 0.3%, explicit intervals 6.4%
TIME UNITS : yes 5.5%, no 94.5%
TEMPORAL UNCERTAINTY : no 97%, yes 3%
PROCESSES : yes 45.9%, no 54.1%
REPEATING PATTERNS : yes 22.5%, no 77.5%

2. Effects of lack of explicit TRR on diagnostic performance

Based on our diagnostic success criteria, we had to discard a number of patient cases, for any of the following reasons: the diagnosis was a finding in QMR, the diagnosis was not part of the QMR's KB, the cases did not represent a straightforward diagnostic problem (but a therapeutic or "rule-out" problem), QMR did not produce a diagnostic list, diagnoses were asserted (i.e., given to the system as fact), or all the necessary information was not available in the patient record. Thus 35 out of 105 cases were excluded from subsequent analyses. We experimented with various values for the q parameter (see methods), and decided to use q=100% to provide a better balance of sample size between successful and non-successful diagnostic groups (as it turns out, our results are insensitive to this parameter for the tested range of 20 to 100%).

From the examined confounding variables, most were characterized by a small worsening of diagnostic performance (odds ratios were between .54 and .74). TRR content had an odds ratio of .7, which means that the odds of getting a correct diagnosis versus an incorrect one in the high TRR complexity group, was 70% the odds of a correct diagnosis in the low TRR complexity group. Unfortunately, our modest sample size did not allow for tight confidence limits (95% c.l. = .27 to 1.83), and all the associations examined were not statistically significant (at the .05 level), so they must be interpreted as indicative only. Finally, we built a Bayesian Network model utilizing the K2 algorithm. In the most probable model found for this data (as shown in Figure 4), diagnostic correctness is determined jointly by TRR content, uncertainty, diseases number, and spatial information.

This model provides an interpretation of the dependency of diagnostic performance on TRR content and the rest of the variables in the form of a conditional probability distribution:

p(correct-diagnosis | diseases, uncertainty, spatial info, TRR-content)

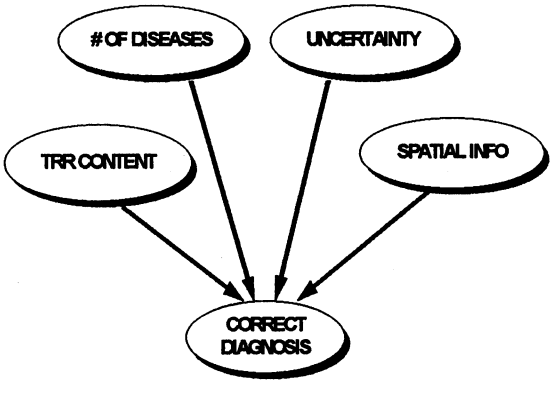

Figure 4: Determinants of diagnostic accuracy

By examining this distribution we concluded that no clear form of covariation exists between TRR content and successful diagnosis, when the rest of the explanatory variables are taken into account. For instance, high TRR content is associated with low probability for correct diagnoses (p=.17) when the other three predictors take the value 'high', while high TRR content is associated with high probability for a correct diagnosis (p=.8) when number of diseases=high and uncertainty=low. Other interesting observations have to do with holding the values of the rest of the three variables constant and observing the probabilities of successful diagnosis: sometimes the probability of a correct diagnoses increases, other times it decreases, when we go from low to high TRR content (depending on the set values of the confounders). Utilizing the principal components - derived measures of TRR case content yielded similar findings. Although the interpretation of these results is complicated and should be viewed with caution in light of the modest sample size, it suggests that *TRR content per se is not a very strong indicator of the diagnostic performance of QMR, in the context of this study, which is evidence in favor of our initial hypothesis.*

DISCUSSION

In this paper we presented an initial analysis of QMR's implicit TRR both in terms of expressive power and performance. At the knowledge-engineering level we were surprised to find out that the QMR knowledge base contains an impressive array of different temporal types, which we identified and classified. The various temporal types are composed of a small number of primitives. We identified this ontology. We additionally abstracted specific temporal patterns and types of temporal reasoning employed, and examined their importance. We believe that the identification of these temporal entities offers three potential benefits:

(a) It *explains the ability of the system* to cope with the rich temporal nature of most patient cases, since it shows an abundance of TRR structures which can be mapped to patient-specific information. In few cases, the patient records were found to contain TRR types that were not in the QMR lexicon [21]. Naturally, this success is highly dependent on the human users of the system, who perform the abstraction from the patient record to the program.

(b) In cases where the system is expected to function independently from human users, it suggests the *types of temporal abstraction mechanisms (and thus intelligent temporal data pre-processors)that should be in place* for the system to function properly. These abstractions complement the set of suggested

mechanisms offered by other researchers who have presented well-defined temporal abstraction mechanisms, aimed at having general applicability [14,16]. In contrast to those mechanisms, the abstractions from QMR can be used independently of specific problem-solving methods and software tools. The expected gains are flexibility and speed of development, whereas the trade-off is in clarity of specification and possibly in domain-independent applicability.

(c) In an exploratory sense, it is a starting point for identifying *important TRR requirements for the design of formal MDSS models employing explicit TRR* (for example, among other things, they raise interesting questions about the proper temporal granularity of representing disease-finding relations).

In the second part of the experiments described in this paper, we focused at the performance level. We found that *TRR content has a fairly small, and statistically non-significant, effect on the diagnostic performance of QMR.*

Although we demonstrate satisfactory heuristic power for the QMR system/domain with respect to the temporal robustness of its heuristic, implicit handling of time, we believe that by conditioning the diagnostic performance of the system on a set of a well-specified temporal criteria [23], we will be able to gain further insight into the limits of implicit TRR. Such an analysis can be greatly facilitated by using a formal language for TRR. We plan to pursue this direction of research. Finally it should be kept in mind that our findings are specific to the QMR system and domain. We hope that these results will stimulate similar analyses for other medical systems and domains, so that eventually MDSS developers will be able to make more conscious and informed choices when it comes to selecting among explicit and implicit TRR methods, given their operational constraints and domain of application.

ACKNOWLEDGEMENTS

The first author is indebted to Prof. B.G. Buchanan, and Dr. R.A. Miller for their invaluable guidance throughout the design, implementation, and interpretation of the research described in this paper. The authors would also like to thank Dr. N. Giuse for her assistance with the utilization of the QMR diagnostic performance patient cases in this paper, as well as for useful comments.

Reference

(1) Kahn MG "Modelling time in medical decision-support programs" Med Dec Making 1991;11:249-264.

(2) E. Shortliffe and L. Perreault (eds.): Medical Informatics: Computer applications in Health Care. Addison-Wesley 1990.

(3) Levesque HJ, Brachman RJ. A fundamental tradeoff in knowledge representation and reasoning. In: *Readings in Knowledge Representation*. Levesque HJ, Brachman RJ (eds.) Morgan Kauffman Publishers 1985, 42-70.

(4) Kassirer J. Kopelamn R "Learning clinical reasoning" Williams and Wilkins 1991.

(5) Sager N "Medical language processing: computer management of narrative data" Reading Mass: Addison-Wesley 1987.

(6) Fagan LM "VM: representing time-dependent relations in a medical setting" Doctoral dissertation, Stanford 1980.

(7) Ackerman E, Gatewood L "Mathematical models in the health sciences" Univ. of Minnesota Press, 1979.

(8) Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982; 307: 468-476.

(9) Long W, Naimi S, Criscitielo M "Development of a knowledge base for diagnostic reasoning in cardiology" Comput Biomed Res 1992;25: 292-311.

(10) Stefanelli M "Therapy planning and monitoring" (editorial) Artificial Intelligence in Medicine 1992;4: 189-190.

(11) Chandrasekaran B, Wong T, Pryor T " 'Deep' models and their relation to diagnosis" Artificial Intelligence in Medicine 1989;1: 29-40.

(12) Allen JF, Hayes PJ "A common sense theory of time" IJCAI proceedings 1985:528-531.

(13) Berzuini C, Bellazi R, Quaglini S, Spiegelhalter D "Bayesian networks for patient monitoring" Artificial Intelligence in Medicine 1992;4: 243-260.

(14) Shahar Y Musen MA "RESUME: a temporal-abstraction system for patient monitoring". Comput Biomed Res 1993; 26: 255-73

(15) Das A, Tu S, Purcell G, Musen M "An extended SQL for temporal data management in clinical decision-support systems" SCAMC 1992:128-132.

(16) Kohane I. "Temporal reasoning in medical expert systems" MEDINFO 1986: 170-174.

(17) Shoham Y "Temporal logics in AI: Semantical and ontological considerations " Artificial Intelligence 1987;33: 89-104.

(18) Agresti A "Categorical data analysis" Wiley 1990.

(19) Spector P "Research designs" Sage 1981.

(20) Colton T "Statistics in Medicine" Little, Brown 1974.

(21) Aliferis CF, Cooper GF, Buchanan BG, Miller RA, Bankowitz R, Giuse N "Temporal reasoning abstractions in QMR" Report SMI-94-03, 1994.

(22) Cooper GF, Herskovits E "A Bayesian method for the induction of probabilistic networks from data" Machine Learning, 1992; 9: 309-347.

(23) Aliferis CF, Miller RA "On the heuristic nature of Medical Decision Support Systems" Report SMI-94-05, 1994.