TiMeDDx – A Multi Phase Anchor-based Diagnostic Decision-support Model

Yaron Denekamp[1,2,3] and Mor Peleg[4,5]

[1]Galil Center for Medical Informatics, Faculty of Medicine, Technion - Israel Institute of Technology, Haifa, Israel, [2]Medical Informatics Unit, Hospital Division - Corporate Headquarters, Clalit Health Services, Tel-Aviv, Israel, [3]Carmel Medical Center, Haifa, Israel, [4]Department of Management Information Systems, University of Haifa, Haifa, Israel, [5] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA

All communication should be with:

Mor Peleg

Dept. of Management Information Systems, University of Haifa, Israel, 31905

Peleg.mor@gmail.com

Fax: 1 (413) 375-3755

Abstract

Main clinical manifestation (MCM)-oriented diagnosis starts with a chief problem and reasons about possible diagnoses that can be manifested in that way. The reasoning process often starts by considering abstract diagnosis groups (e.g., infectious vs. non-infectious diarrhea) and refines them. Most existing diagnostic decision-support systems (DSSs) are not specially tailored toward assisting non-expert physicians in the proper and efficient investigation workup of MCM-oriented diagnosis. We developed a prototype diagnostic decision-support model called TiMeDDx that is MCM-oriented and follows the hypothetico-deductive clinical reasoning process of differential diagnosis. The model guides users in a phase-by-phase manner regarding abstract diagnosis groups and diagnoses that should be considered and appropriate data that should be collected during the clinical investigation process. TiMeDDx's knowledge base contains, when possible, knowledge derived from MCM-oriented evidence-based sources. We explain the knowledge model and diagnostic algorithms (Bayesian and heuristic) of TiMeDDx, using the clinical problem of diarrhea as a case study, and contrast TiMeDDx with models of existing diagnostic DSSs.

Keywords: diagnostic decision-support systems, decision-support systems, clinical diagnosis, problem-oriented diagnosis, TiMeDDx

1 Introduction

The diagnosis process is a complex cognitive process comprising a variety of different types of problem solving tasks that are involved in the clinical reasoning process. In addition, physicians must follow progress in clinical research and incorporate ever growing new knowledge regarding diagnosis of clinical problems and diseases. Clinical Decision Support Systems (DSSs) have been recognized as important tools to aid clinicians in gathering relevant knowledge and data, making clinical decisions, managing medical actions more effectively, and thus achieving reduced practice errors, a higher standard of care, and reduced costs [1]. Clinical DSSs can provide tools for information management (e.g., retrieval and storage), for focusing attention (e.g., alerts and reminders), and for providing patient-specific recommendations [2]. Diagnostic DSSs [3] assist a clinician with one or more component steps of the clinical diagnostic process. Diagnostic DSSs have been developed since the 1970's for different ranges of diseases and symptoms [4-13]; some systems aim to cover broad medical domains while other systems focus on specific problems (e.g., de Dombal's DSS for the differential diagnosis of abdominal pain [14]). As described in the next section, developing diagnostic DSS poses difficult challenges. Despite these challenges, several evaluation studies have shown diagnostic DSSs to be reliable and accurate [10, 12, 15-19]. However, relatively few are being used at present [7, 9-13] and the rate of usage in routine clinical practice is low. Part of the difficulty experienced in incorporating them may be associated with the lack of integration into the clinical reasoning process involved in clinical diagnosis.

In essentially all of the diagnostic DSSs, the user enters data about symptoms, signs, and laboratory test results, and the DSS produces a list of possible diagnoses, ranked in order of

likelihood. These DSSs help the user in selecting controlled vocabulary terms to describe the findings. Some systems [15] guide the user with the diagnostic process, incorporating rule-in and rule-out diagnostic processes depending on the scores of the hypotheses in the differential diagnosis set (DD-set). Other systems [12, 13, 17] offer decision-support services that the user can invoke, such as providing a disease profile, focusing the diagnosis on important features, viewing evidence for a diagnosis, and obtaining explanations of findings. Most of the existing diagnostic DSSs are used by novice clinicians, or by experienced clinicians to aid them in diagnosing difficult cases.

We have noted that the available probabilistic diagnostic DSSs are not specially tailored toward assisting non-expert physicians like primary care physicians (PCPs), interns, and residents, in the proper and efficient workup of a clinical manifestation which may be a symptom, sign, abnormal laboratory or imaging test results, or a combination of these. We refer to this kind of clinical diagnosis process as main-clinical manifestation (MCM)-oriented diagnosis. MCM-oriented diagnosis is a problem-oriented process that starts with a chief clinical problem, reasons about possible diagnoses that would be manifested as the MCM, and suggests the clinical data items, laboratory, and imaging tests that should be collected in order to differentiate among alternative diagnoses.

The MCM-oriented reasoning process is often conducted in phases. At the initial phases, the differential is sometimes between diagnosis groups that are meaningful in the context of the diagnostic process, for example, chronic vs. acute diarrhea or infectious vs. non-infectious diarrhea. We refer to such diagnosis groups as *abstract diagnoses*. As the diagnostic process advances, the differential is between actual diagnoses. Such phased problem-oriented process [20] is preferred for many clinical problems (e.g., syncope, jaundice), as seen in the classical

clinical textbooks and in some specific medical books and evidence-based clinical practice guidelines [20, 21].

The process of clinical investigation of clinical problems is complex and requires using and analyzing a wide relevant set of clinical data items in a systematic organized way. Physicians are expected to properly handle a wide range of clinical problem investigations. Yet incomplete workup was found to be a major source of quality of care problems [22]. DSS can aid physicians to manage the investigations, avoiding unnecessary referrals, unnecessary costly tests, or diagnostic errors, by empowering them with updated knowledge, evidence-based when possible. Representing and delivering such knowledge could potentially help overcome diagnostic errors that are due to cognitive biases, such as 'confirmation bias', 'outcome bias', or 'overconfidence bias' [23]. Motivated by the need for diagnostic DSSs that would support the investigation process of clinical *problems*, we developed a prototype diagnostic system called TiMeDDx - a diagnostic decision-support model that is MCM-oriented and supports a diagnostic process that is conducted in phases of decreasing abstraction. TiMeDDx is based on an information model that integrates several new notions, discussed in Section 3, with relational or Bayesian representations used in old diagnostic DSSs (Quick Medical Reference [16], DXplain [12], and QMR-DT [8]). In Section 2, we discuss background material related to existing diagnostic DSSs and MCM-oriented diagnosis. Section 3 introduces the TiMeDDx model and discusses its knowledge-elicitation process. In Section 4 we present our preliminary evaluation studies of TiMeDDx, including its empirical contrast with some of the existing diagnostic DSSs (QMR, DXplain, and GIDEON). We conclude with a Discussion followed by a Conclusion section.

2 Related Work

In this section, we discuss diagnostic DSSs in more detail, focusing on the challenges involved in development of diagnostic DSSs and comparing existing diagnostic DSSs. We also expand our discussion on main-clinical-manifestation-oriented diagnosis and discuss the hypothetico-deductive clinical reasoning process.

2.1 Challenges in developing diagnostic DSSs

Developing diagnostic DSSs that cover large domains poses great challenges [3] including:

(1) Acquiring the clinical knowledge and keeping it up to date. Knowledge can be acquired by eliciting it from domain experts [24, 25] or it can be gathered from the literature or by compiling data found in electronic medical record systems [8];

(2) Representing and reasoning with the clinical knowledge. The main decision-support models are quantitative (e.g., statistical models including Bayesian networks, machine learning approaches) or qualitative (e.g., heuristic knowledge represented as rules, ontologies, or decision tables) [26, 27];

(3) Supporting the sequence of reasoning used in the diagnosis process;

(4) Integrating with controlled vocabularies and clinical information systems;

(5) Supporting system evolution, including evaluation, testing, and quality control;

(6) Addressing legal and ethical issues;

In this paper, we address the first three challenges.

2.2 Available Diagnostic Decision-Support Systems for broad medical domains

The “Leeds abdominal pain system” [14] was the first diagnostic DSS, published in 1972. Since then, a number of computer-based systems with diagnostic capabilities have been developed for

broad ranges of diseases. Examples include Dxplain [12], Iliad [10], Meditel [19], Quick Medical Reference (QMR) [6], Problem Knowledge Coupler (PKC) [13], Isabel [7], Physician Assistant Artificial Intelligence Reference System (PAIRS) [9] (previously known as QMR-DT [8]), and Global Infectious Diseases and Epidemiology Network (GIDEON) [11]. These systems differ in the data used to determine their probability estimates, the extent to which diseases and related clinical data are addressed in their knowledge bases, the particular vocabulary they require to describe clinical data, and the computational model they use to combine and analyze data, as shown in Table I and described in detail below.

In terms of the computational model, Iliad, Meditel, PAIRS, and GIDEON are based on Bayes' theorem; for example, the differential diagnosis list in GIDEON [11] is based on a Bayesian formula that compares the mathematical product of disease incidence times the rate of symptom occurrence for all relevant infectious diseases within a given country. In addition to using Bayes theorem, Iliad [10] also uses decision rules for reasoning with clusters of conditionally-independent findings. This is meant to solve the problem of over confident, unreliable diagnostic results that occurred because findings were not completely independent.

Isabel [7] uses pattern-matching algorithms to compare findings entered by a user to terms used in a selected reference library that includes text from medical books and journals. By collating text related to one specific diagnosis under a single diagnostic label within a pre-designed diagnostic tree, it was possible for the software to generate a unique signature of key concepts for each diagnosis.

DXplain [12] and QMR [6] (and QMR's predecessor, Internist-1 [15]) use non-Bayesian algorithms that focus on a relational model describing relationships between case findings (symptoms, signs, laboratory data) and individual diseases to derive a weighted assessment of a

patient's clinical presentation [17]. In Internist-1 [15] as well as in QMR and DXplain, one type of disease-finding relationship represents the frequency with which the finding occurs in the disease, and the other the degree to which the presence of the finding suggests consideration of the disease (evoking strength). Other tables store the importance of explaining findings, disease frequencies (prevalence) and disease importance (impact of not considering the disease if it is present). The DXplain algorithm also considers the number of diseases in the differential-diagnosis set (DD-set).

Problem Knowledge Coupler (PKC) [13] takes a philosophical stand that the clinician assessing a patient should understand the pattern of findings (and test results) occurring for her patient rather than rely solely on probabilities taken from the general population. Therefore, the relationships between diseases and findings/tests are not weighted and no algorithm is used to rank the possible diagnoses that explain the finding; instead, the number of present findings associated with each disease hypothesis is shown next to the disease hypothesis, along with the total number of findings which may be associated with that disease, according to the PKC knowledge base. PKC also allows evaluating the expected prognosis of a disease with treatment and without treatment. In addition to the relationships between diseases and findings, predecessor/successor relationships between entities (diseases or findings) can flexibly be defined by knowledge modelers to define knowledge networks. However, the semantics of the specific modeler-defined relationships is not formally defined and therefore it cannot be used to reason in a way that is specific to the relationship type.

As shown in Table I, all the diagnostic DSSs that we considered, except for PKC, focus on broad domains where knowledge regarding the weight of relationships between diseases and findings

do not depend on the MCM. The knowledge sources for different diagnostic DSSs that are in use today come from the literature, but except for Isabel, where machine learning algorithms are used to gather relevant knowledge, in the other systems the knowledge from the literature is incorporated into the knowledge-base manually.

All of the diagnostic DSSs can aid a physician during his clinical reasoning process. They all allow the user to start the diagnosis process with patient findings, but out of the systems surveyed in Table 1, only Iliad, DXplain, and PKC provide advice on data that should be collected and laboratory tests that should be employed. In PKC, the sequence of data that should be collected is represented in the system's model ahead of time. In all of the diagnostic DSSs, the DD-set is shown (or can be displayed upon request) at every stage of the iterative diagnostic process.

GIDEON is the only system that explicitly represents temporal relationships between data items. All the systems that use a Bayesian computational model can support synergistic effects between findings, i.e., findings that together suggest a diagnosis with a higher probability. GIDEON, and to some extent also DXplain, consider the geographical location of the patient as a factor in the diagnostic process. When findings are entered, all possible diagnoses that cover those findings are considered in all of the DSS, but disease prevalence is taken into account for ranking the possible diagnoses. In DXplain, rare diseases are displayed separately.

The different diagnostic DSSs all provide explanations for why each of these diseases might be considered. DXplain also lists the clinical manifestations, if any, which would be unusual or atypical for each of the specific diseases and GIDEON also explains why other diagnoses are not considered. In Isabel, the explanations are in the form of linking with up to date knowledge from

textbooks and journals. However, none of the diagnostic DSSs currently in use offer pathophysiological reasoning that create models of a specific patient's illness [28-30].

2.3 Hypothetico-deductive Reasoning and Main Clinical Manifestation-oriented Diagnosis

Several cognitive models of clinical diagnostic reasoning processes have been developed. Some of the highly-accepted models view the diagnostic process as either hypothesis formulation or pattern recognition [31]. Our formulation of the clinical problem-solving process is aligned with the hypothesis-formulation approach, or hypothetico-deductive reasoning model [32]. Hypothetico-deductive reasoning is an iterative process, which involves staged data collection followed by data interpretation and the generation of a set of hypotheses (which in the case of clinical diagnosis is known as the DD-set), leading to hypothesis-directed selection of the next most appropriate data to be collected. The data collected at each stage are used to reformulate or refine the active hypotheses. The reasoning process is iterated until one hypothesis reaches a threshold level of certainty. The staged-process helps to focus the reasoning process. When physicians have collected initial data from the patients' history and physical examination, they can generate an initial DD-set. By that time, physicians have expectations of what they will find on further examination or may have specific tests in mind that will help them to distinguish among still active hypotheses.

A clinical investigation usually starts from some clinical anchor finding. Many times, this clinical anchor is the reason for patients to seek medical care as well as for physicians to initiate an investigation. We refer to this as a Main Clinical Manifestation (MCM), which may consist of a single clinical problem such as, diarrhea, syncope, or jaundice, laboratory test result (e.g., hyponatermia), or combinations of several linked findings, such as fever and rash (which is a

common clinical manifestation in pediatrics). The MCM plays an important role in focusing the diagnostic process. This is in concert with the findings of Eddy and Clanton [33] who showed that identification of a pivotal finding is often used to simplify the diagnostic problem and to narrow the focus to a limited set of hypotheses. During the clinical reasoning process, when doctors consider the various possible diagnoses that explain the MCM they take into account the probability of each diagnosis to be manifested as the MCM, ranking diagnoses that are more likely to be manifested as the MCM higher.

During the diagnostic process, physicians collect and analyze several types of data types, including subjective information acquired by questioning the patient (i.e., symptoms or medical history), objective findings obtained by performing physical examination (i.e., signs) and all sorts of laboratorial and imaging data. At any point in this process, there are several diagnoses that might fit the data collected (i.e., differential diagnosis). Their number should decrease as the diagnostic process progresses. As has already been shown years ago, expert clinicians can make a diagnosis in the majority of patients using the history and physical examination data alone [34, 35].

MCM-oriented diagnosis is a well-accepted approach in clinical diagnosis. It is evident in medical books [36, 37] (e.g., diagnosing fever and rash in children). Traditionally, text books often did not report evidence-based statistics regarding disease prevalence per clinical problem or frequency of clinical data items given a disease. Medical books based on the principles of evidence-based MCM-oriented diagnosis that report such data are becoming more prevalent [20]. As our intention was to create a MCM-based diagnostic DSS that relies on evidence-based clinical knowledge when possible, we wanted to assess the availability of evidence-based (EB) sources, such as clinical practice guidelines, for aiding MCM-oriented clinical diagnosis utilizing

primarily data types found during history and physical examination. To determine the extent at which clinical guidelines follow MCM-oriented diagnosis and report EB statistics, we conducted a study [21] of diagnostic guidelines that were archived in the National Guideline Clearinghouse (NGC) web site (www.ngc.gov) - a public resource for evidence-based clinical practice guidelines, initiated and maintained by the Agency for Healthcare Research and Quality and the US Department of Health and Human Services. We employed filtering features provided by NGC's website to consider only the potential diagnostic guidelines (1182 guidelines, at the time of the study). We then manually inspected each guideline and found that 146 of them indeed addressed diagnosis that starts with a MCM. After characterizing 25% of these guidelines [21], we found little use of quantitative statistical data, such as frequency of manifestation of findings in given diseases and disease prevalence, for determining diagnosis. That trend found in the study [21], which was done in 2007, was also observed in an updated study that we are currently summarizing. In addition, we found that many of the guidelines make use of disease categories i.e., abstract diagnoses rather than just individual diagnoses. Some guidelines reported temporal and synergistic relationships between patient findings, which serve as important knowledge for diagnosis.

These findings suggest that although MCM-oriented diagnosis is a well accepted diagnostic approach, MCM-oriented guidelines that report evidence-based statistical data are not very common, necessitating the elicitation of such data from other sources, such as experts or from statistical clinical databases.

3 TiMeDDx

We developed a prototype diagnostic DSS called TiMeDDx to assist physicians in the process of MCM-oriented diagnosis. TiMeDDx emphasizes proper workup of a presenting symptom, sign, abnormal test result or a combination of these and supports a hypothetico-deductive diagnostic process. Toward this goal, TiMeDDx integrates several notions in a novel way resulting in a multi-phase, anchor-based information model that uses abstract diagnosis groups. This multi-phase approach, which revolves upon an anchor finding per each phase, enables efficiency in conducting the diagnostic process using a minimal effective set of clinical data items (CDIs) in each phase. In section 3.1 we explain the notions of the MCM-oriented approach of TiMeDDx. Section 3.2 details how we integrated these notions and structured them in the information model of TiMeDDx. Sections 3.3 and 3.4 discuss two possible computational models for scoring disease hypotheses, which can be used with TiMeDDx' information model. Finally, section 3.5 discusses the development process of a TiMeDDx knowledge base.

3. 1 The notions used in TiMeDDx' MCM-oriented approach

TiMeDDx's MCM-oriented approach uses the following six notions.

3.1.1 MCM-oriented diagnosis

TiMeDDx enables MCM-oriented diagnosis and emphasizes the use of clinical data items from the history and physical examination; a MCM is any CDI or a combination of CDIs that can be a symptom, sign, laboratory or other test, which is the starting point of the diagnosis process. MCM-orientation was first introduced in the Present Illness program [39] and in Problem Knowledge Coupler [13]. Unlike other DSSs, in TiMeDDx the MCM is treated as being more important than other findings and plays a crucial role at focusing the diagnostic process.

TiMeDDx does not associate every disease with every finding, as done in diagnostic DSS for broad domains, nor does it represent its knowledge as a network of interconnected frames of diseases and clinical states from which disease hypotheses are selected, based on findings exhibited by the patient, as done in the Present Illness program. TiMeDDx considers and ranks for each MCM only the set of diagnoses whose main clinical manifestation is the MCM, as reported in EB sources or medical literature that discuss problem-oriented diagnosis. This is expected to enable more efficiency and accuracy in scoring the different diagnosis hypotheses that are evoked by the MCM. For example, if the MCM is hyponatremia (low level of sodium), then in TiMeDDx, pneumonia will not be part of the DD-set as it is in some other models. This is because although hyponateremia can be found in pneumonia, it will never be the main clinical manifestation of pneumonia; a patient with pneumonia will exhibit other findings (e.g., fever, cough, rapid breathing, etc.) that will focus the clinician on pneumonia.

3.1.2 Phases

TiMeDDx is novel in supporting a diagnostic process that is carried out in *predetermined phases*. TiMeDDx supports phases by structuring the process of DD reduction as a predetermined tree of hierarchical DDs, which we refer to as DD-tree. Each layer of the tree corresponds to a diagnostic phase (e.g., acute diarrhea, infectious diarrhea), mimicking the clinician's hypothetico-deductive reasoning process of diagnosis that is used during problem-oriented diagnosis. At the beginning of the diagnostic process, the focus (anchor) of the diagnosis is the MCM (e.g., diarrhea, jaundice, syncope) that triggered the diagnostic process, and serves as the root of the tree. For this anchor, a set of diagnoses that can be abstract and relevant clinical data items for making a diagnosis are defined. As shown in Table 2(a) for the diarrhea anchor, a set of two abstract diagnoses are provided in phase 1: acute vs. chronic diarrhea. To differentiate between

them, TiMeDDx uses the CDIs: "duration of diarrhea ≤ 14 days" and "duration >14 days", which, per definition, are considered pathognomonic (i.e., unambiguously characteristic of a particular disease) for discriminating between these two alternatives. As the diagnostic process advances through the levels of the diagnostic tree, the DD-set becomes more and more specific. For example, if in phase 1, the selected alternative was acute diarrhea, then, as shown in Table 2(b), acute diarrhea serves as the anchor for phase 2. Phase 2 includes five alternatives, including infectious diarrhea, medication change, inflammatory bowel disease, intermittent bowel obstruction, and colonic ischemia. To differentiate among these hypotheses TiMeDDx uses a collection of CDIs that are relevant for that phase, as shown in the second row of Table 2(b) and the user enters values to indicate whether these CDIs are present in the patient. The strengths/probabilities of relationships between disease hypotheses and CDIs in each phase are indicated as numbers in the table, as explained later in this section. Based on the CDIs values for the patient, an algorithm (heuristic or Bayesian) ranks the diseases in the DD-set and sets the highest ranking disease as the new anchor for the next phase.

3.1.3 Abstractions

As explained above, the multi-phase diagnostic process of TiMeDDx often starts with abstract concepts, is refined in each phase, and ends in specific diagnoses. This diagnostic process that uses abstractions is valuable not only because of the efficiency of CDIs considered in each phase, but also because abstractions are often used by clinical experts during problem-solving. As discussed by Newell and Simon [38], studies examining constrained problem spaces such as chess-playing have documented that experts recognize patterns of activity within a domain at an integrated, higher level ("chunking") than novices. Abstractions have been used in diagnostic DSSs before. Abstractions have been used in the Internist-1 [30] knowledge base, which contains

a hierarchy of disease categories, organized primarily around the concept of organ systems, where positive findings can evoke either individual disease nodes or higher-level nodes in the disease hierarchy. Pople [30] suggested a reasoning model where any given disease can be classified in as many descriptive categories of the hierarchy as are appropriate.

3.1.4 Anchor-specific disease-finding relationships

TiMeDDx provides weighted relationships between disease and findings that are specific to the given anchor and to a given geographical location, including (a) the DD-set that is relevant and probable for the given MCM (first column in Table 2); and (b) the set of relevant findings that can distinguish among the diagnoses in the DD set (the top row in each table of the Table 2 table set). While these two relationships are present in PKC, in PKC, they are not weighted; (c) the likelihood of a diagnosis to be manifested as the anchor (second column in Table 2) – a notion that is unique to TiMeDDx. This is used to rank higher diagnoses that are usually manifested as the anchor finding. For instance, for an anchor of syncope, the DD-set includes cardiac arrhythmias and pulmonary embolism, yet cardiac arrhythmias are more likely to be manifested as syncope than pulmonary embolism; (d) the evoking strength with which a finding suggests a diagnosis in the DD set (numbers in the cells of Table 2). Unlike the use of this feature in Internist-1 [15] (where it was first introduced), QMR, DXplain, and the Present Illness program, the evoking strength in TiMeDDx considers just the disease hypotheses that are relevant for the anchor finding (which, at the beginning of the process is the MCM); (e) the penalty that a disease hypothesis should receive in the absence of a finding (numbers given in parentheses in the cells of Table 2). Penalties are also used in other diagnostic DSSs, such as Internist-1 [15] (where they were first introduced), QMR, DXplain, and the Present Illness program. However, in TiMeDDx, the size of this penalty is proportional to the frequency at which the disease exhibits

the finding; and (f) synergistic effect between findings that together suggest a diagnosis with greater certainty than the sum of the two is explicitly modeled (Table 3); these relationships may be temporal (e.g., fever before rash, Jaundice after fever) or not temporal, just two findings that together strengthen a diagnosis.

3.1.5 Computational model

TiMeDDx' information model can be combined with different computational models (a heuristic model and a Bayesian model) for scoring disease hypotheses. This feature was guided by the Bayesian formulation [8] of the heuristic algorithm of QMR [6]. Uniquely, in TiMeDDx the Bayesian formulation is structured according to the phases of anchors used in the MCM-oriented diagnostic process. The last three relationships between disease and findings discussed above are used with the heuristic scoring algorithm, discussed in Section 3.4. The Bayesian approach, discussed in Section 3.3, uses prior probabilities for each disease hypothesis and conditional probabilities for each combination of finding and disease (finding frequencies).

3.1.6 User control

TiMeDDx' philosophy allows the user to follow the diagnostic process with any diagnosis in the DD-set, even if it is not the highest-ranking one. This feature, which also exists in PKC, is in accordance with the modern view of DSSs [40] as providing assistance to a user who is in charge of the clinical process rather than being a Greek oracle who solves the clinical task for the user. This property is important because, as pointed out by Miller [41], no computer program can know all that needs to be known about the patient case, no matter how much time or effort is spent on data input into the computer system, and therefore the clinician user who directly evaluated the patient must be considered to be the definitive source of information about the patient during the entire course of any computer based consultation.

3.2 The TiMeDDx information model

The main class in the TiMeDDx model is *Anchor*, which represents the MCM in the first phase of the diagnostic process. Anchor is defined using three structural slots and three relationship slots. The relationship slots store knowledge that is used by the TiMeDDx algorithm to score diagnoses in the DD-set. Figure 1 shows an instance of the Anchor class, as modeled using the Protege-2000 [42] modeling tool.

3.2.1 Structural Slots

anchor_concepts – the concept (or concept combination, such as fever and rash) on which the DD is focused at the current diagnosis phase. In Figure 1, the anchor concept is Acute Diarrhea.

relevant_diagnoses_or_abstractions – the relevant diagnoses for the current DD phase. At initial DD phases, we often use abstractions instead of final diagnoses.

relevant_CDIs – the CDIs that should be collected in order to select the most probable diagnosis from the DD set. For each CDI we specify the medical concept and whether it is a required value for a certain phase of the DD process given a certain anchor, as shown in Figure 2.

3.2.2 Relationship slots

The following types of relationships relate CDIs to diagnoses. They are used by the algorithm that scores the diagnoses in the DD-set. Figure 3 shows the information model of these three relationships, using an Entity-Relationship (ER) Diagram [43] – where each information class is represented by an entity type (depicted as a rectangle) and relationships between information classes are represented by a relationship type (depicted as a diamond). We use the ER notation because it is widely familiar and simple and often is used to represent information models.

diagnoses_manifested_as_anchor – bonus points are given to diagnoses that are usually manifested as the anchor concept[1](s). For example, cardiac arrhythmias are often manifested as syncope (anchor), but pulmonary embolism, which is in the DD-set of the syncope anchor, is usually not manifested as syncope.

CDI_evokes_Diagnosis – Like the Internist-1/QMR system [15], TiMeDDx considers CDIs that suggest a diagnosis, with a certain evoking strength (1-10). As in Internist-1, frequency_penalty stores the frequency at which a finding is found in a disease; when the patient does not exhibit a CDI that is frequent in a disease hypothesis, points can be deducted from that hypothesis.

Synergistic_CDI_Relationship_Given_Diagnosis_For_Anchor – in some cases, when combinations of two (or more) CDIs occur together or in a certain temporal pattern, this suggests a certain diagnosis more probable than the additive contribution of each one of the CDIs alone. For example, if it is known that diarrhea developed less than six hours after ingesting suspicious food, it suggests the diagnosis of toxin-borne Diarrhea, based on the combination of the diarrhea and ingestion of suspicious food as CDIs, as shown in Figure 4.

3.3 A Bayesian model for scoring disease hypotheses

We have attempted to reformulate TiMeDDx model in Bayesian terms with certain simplifying assumptions. As was done for the Bayesian formulation of the QMR knowledge base [8], we assumed that findings are conditionally independent given any disease hypothesis (therefore the probability of having multiple findings given a hypothesis is the product of probability of having one finding given the hypothesis, for all findings in the set of findings). Therefore, we convert temporal synergistic relationships between two findings and a disease hypothesis into one

[1] Often, the anchor concept holds a MCM that is a patient's CDI. However, the anchor may alternatively be an abstraction used in the DD (part of the DD-set), such as in the case of infectious diarrhea, which is an abstract anchor concept and not the MCM (the MCM is diarrhea)

finding. As in QMR-DT, we model the influence of multiple diseases on a finding assuming causal independence (i.e., the probability of a finding given only one disease is present P(F|D) instead of given combinations of diseases. In this way, all the conditional probabilities that we use in the Bayesian Networks (BNs) are of the form P(F|D), representing frequency data. We also assume that only one of the alternative hypotheses would be present in a patient.

To take advantage of the phased model of TiMeDDx, we arrange the knowledge in sets of small BNs, where each network corresponds to one phase of TiMeDDx's knowledge-base (KB) and includes the relevant findings and disease hypotheses for that phase, the prior probabilities of the disease hypotheses and the conditional probabilities P(F|D). Parts a, c, and d of Figure 7 show BNs that correspond to the example discussed in Section 3.1.

Because the TiMeDDx model is arranged in diagnostic phases, as we move from a BN of one phase to the BN of the next phase, we simply refine our diagnosis without the need to combine the numerical results from previous phases. For example, in phase 1, setting the value of duration to <=14 and updating the model will result in a posterior probability of 1 for acute diarrhea. Given that abstract diagnosis, we proceed to the phase 2 BN whose anchor is acute diarrhea.

To distinguish the main clinical manifestation from other manifestations of a disease, we use the information contained in the disease_manifested_as_anchor relationships of TiMeDDx to update the values of the posterior probabilities of the disease hypotheses. To do so, we construct another BN consisting of the diseases (or abstractions) used in the DD-set of a phase and the phase's anchor (see Figure 7, part b). The probabilities used in this network would be the prior probabilities of the diseases. The conditional probabilities relating the anchor to the individual disease hypotheses P(anchor|D) are derived from the disease_manifested_as_anchor relationships. From this small network, we can derive the posterior probability for a disease

hypothesis given that the anchor is present. These probabilities can now serve as the prior probabilities for the diseases in the BN composed of the diseases in the DD-set for the phase and all the other findings apart from the MCM (part c of Figure 7).

Using the above Bayesian approach requires having prior probabilities for each disease hypothesis and conditional probabilities for each combination of finding and disease (finding frequencies). In QMR-DT, the prior probabilities were assembled from data compiled by the National Center for Health Statistics on inpatients discharged from short-stay nonfederal hospitals and the conditional probabilities P(F|D) were derived from the QMR frequency data. We tried to base our data on EB studies. However, as explained in Section 3.5, many of the required statistical values are not found in the EB sources. Therefore, for simplicity of implementation, we chose a heuristic approach for scoring diagnoses that requires eliciting fewer probabilities from experts, eliminating the need to specify all prior probabilities of diseases and conditional probabilities of finding given disease. This approach is described next.

3.4 TiMeDDx' diagnosis-scoring algorithm

The heuristic scoring algorithm of TiMeDDx requires the input of fewer probabilities than the Bayesian methods discussed in the previous subsection. The algorithm, shown in Figure 5, considers CDIs that suggest a diagnosis with a certain evoking strength. If the CDI is present, then points are awarded according to the evoking strength. If the CDI is considered to be pathogneumonic in evoking the diagnosis, as is sometimes the case for laboratory or imaging findings and very rarely for history and physical examination findings, then that diagnosis would be concluded. If, however, the CDI is not present penalty can be used to deduct points for certain

disease hypotheses. The size of the frequency_penalty indicates the likelihood of a hypothesis being inappropriate when a certain CDI is absent.

The last component used for scoring diagnoses is Synergistic_CDI_Relationship_Given_Diagnosis_For_Anchor. When such a relationship is defined and both CDIs are present, points are added to the relevant diagnosis, according to the weight defined for that relationship. If not both CDIs are present, we use selective penalty to deduct points from the diagnosis.

After all the diagnoses in the DD-set were scored, we calculate the cutoff – diagnoses that are below the cutoff will not continue to future steps of the algorithm. The cutoff is calculated as a value that is 10% lower than the score of the highest ranking diagnosis. The algorithm will suggest the diagnoses that are above the cutoff. If a disease hypothesis was selected via a pathogneumonic finding, then no cutoff value is necessary – only that hypothesis is suggested. However, the user may choose to override the recommended diagnosis and select a different diagnosis from the DD-set for the current anchor. The system can then set this diagnosis as the anchor for the next diagnosis phase, taking the information for the appropriate phase from the respective node in the DD-tree. We do not allow the user to jump to any node in the DD-tree; if the user has already reached a certain anchor we interpret it to mean that he has accepted all the abstractions leading to that node. If this is not the case, the user can start another session.

Figure 6 provides an example of one phase of the algorithm, based on the knowledge shown in Figure 1. That knowledge was adapted from a medical book based on the principles of evidence-based MCM-oriented diagnosis [20] and from the infectious diarrhea guideline [44], as explained in the next sub-section. The full run of the algorithm can be found in Appendix A. The patient for which the algorithm was executed has been having high fever for 7 days, nausea, vomiting, bloody stools, abdominal pain, and tenesmus. The symptoms appeared abruptly and we do not

know whether other people had the disease. However, we know that the following findings are not present: ingestion of suspicious food, mucus, antibiotics, flu-like symptoms, and arthritis. Examining the score of the first diagnosis, Infectious Diarrhea, we can see that 16 points (out of 20) were awarded based on the fact that this diagnosis is usually manifested as diarrhea (the anchor), and 9, 8, and 8 points (out of 10 maximum points per finding) were awarded based on the CDI_evokes_diagnosis links between this diagnosis and the following CDIs: fever (high), abrupt presentation, and nausea/vomiting. Note that the relationship of arthritis evokes inflammatory bowel disease with a frequency_penalty of 5 was used to deduct points from that disease hypothesis because arthritis was not present.

3.5 Developing a TiMeDDx knowledge base

In the TiMeDDx approach, which is centered on a MCM, the set of disease hypotheses and CDIs considered at each diagnosis phase as well as the relationships between findings and disease hypotheses depend on the MCM. Therefore, knowledge added to the TiMeDDx knowledge base to support diagnosis of a new MCM is independent of the knowledge that already exists in the knowledge base for existing MCMs. This has several consequences. First, phase-specific knowledge usually cannot be reused for different MCMs. For example, while arrhythmias are considered as disease hypothesis for a syncope MCM and for a palpitations MCM, different arrhythmias are considered for each MCM, and with different likelihoods. However, as we advance toward more specific phases in a DD-tree (e.g., bradyarrhythmia), it is more likely that these phases could be reused for the DD-trees of different MCMs. A second consequence of the independence of MCM-oriented diagnostic knowledge is that the addition of knowledge for a new MCM will not affect system performance, because only the DD-tree for that MCM would

need to be considered. A third consequence is that the process of developing the TiMeDDx knowledge for different MCMs can be done independently and in parallel. In the rest of this section we discuss the steps involved in developing the TiMeDDx knowledge needed to support diagnosis of a MCM, addressing the level of effort needed. This is based on our experience in developing the knowledge for the diarrhea MCM and for our ongoing development of the syncope MCM.

Whenever possible, we tried to elicit disease hypotheses, relevant CDIs, evoking strengths, frequencies of manifestations, weights of diagnosis manifested as anchor, and weights of synergistic relationships based on EB studies. The disease hypotheses that we considered for the diarrhea case and the CDIs used to distinguish between them were based on EB sources [20, 44]; Table 2 shows the disease hypotheses (first column) that we considered and the CDIs (top row) used to differentiate among them.

Some EB studies report probabilities of manifestation of a finding given a disease (frequency). In TiMeDDx, the important probabilities for the heuristic algorithm are the evoking strengths, i.e., the probabilities of disease given a finding P(D|F) for a given anchor. However, these probabilities are usually not available in evidence-based studies. Bayes law can be used to convert the frequency data into P(D|F) based on the prior probabilities of diseases and of findings per a given anchor. However, the prior probabilities of a finding (per anchor) are difficult to find, and the prevalence of common etiologies of diarrhea in a primary care setting is reported to be unknown in the EB source that we used [20, p. 282]. Nevertheless, since for a given anchor, a small set of relevant diagnoses (or abstract diagnosis groups) are considered as the DD-set, we used the frequency numbers P(F|D) to select the most relevant diagnosis in the limited DD-set, assuming uniform prior probabilities of diseases (prevalence). In this way, the disease in which

the CDIs exhibited by the patient are most frequent is the disease that should be evoked. We converted the frequency numbers to a scale of 0..10. When ranges were reported, we used the average. Frequency data were provided in the clinical guideline [44] that we used for the different diseases belonging to the infectious diarrhea abstraction (Figure 2c). Evoking strength data for the non-infectious acute diarrhea (Figure 2b) were supplied by our experts.

In order to determine the frequency_penalty which we use to subtract points in a selective way from a hypothesis when the patient does not exhibit a finding that is manifested in high frequency in a disease, we relied on judgment of two clinical experts, who consulted the frequency values reported in the EB studies, but used their expert opinion to decide about the selective penalty (see Section 3.4). These experts also provided numbers for other relationships for which no data were reported in the EB studies: the synergistic effects (scale of 1..10) and Diagnosis_Manifested_As_Anchor (scale of 1..20) in the context of diarrhea.

To elicit from experts the frequencies, penalties, and weights for diagnosis manifested as anchor for our preliminary study, we prepared Excel tables such as those shown in tables 2 and 3. The structure of the tables was set based on evidence-based sources and the numbers were supplied by two experts by consensus formulation.

When eliciting such data from experts for a more comprehensive evaluation study, we suggest following the methodology for knowledge base construction based on expert opinion that was proposed by van Ast et al [25]. That methodology suggests starting with a group of experts and calculating the inter-rater interclass correlation coefficient; if it is not large enough, the Spearman-Brown prophecy can be used to predict the number of additional experts.

The effort required to develop the TiMeDDx knowledge for a given MCM is considerable. Based on our experience, gathering information from evidence-based sources and arranging it in

phases of disease hypotheses and CDIs used to distinguish among them required less effort than acquiring the numbers (which include frequencies, penalties, and weights for diagnosis manifested as anchor) that were not available in the EB studies. Working with the experts requires several iterations; in the first iteration, which spans several sessions, the experts supply all the requested numbers using the tables that we prepared. Then a statistical examination of expert agreement is conducted to see if the numbers could be averaged. As noted above, establishing agreement between experts may require using additional experts. After agreement is established the numbers are entered into the knowledge base and the system's performance on test cases is evaluated, as described in the next section. Fine tuning the knowledge base to support the initial set of test cases requires further iterations with the experts.

4 Preliminary evaluation studies

We tested and refined the TiMeDDx model by examining the MCM-oriented diagnostic process of diarrhea. As EB sources of medical knowledge, we used a medical book of problem-oriented diagnosis [20] and a guideline [44] for diagnosing infectious diarrhea. Screenshots from the TiMeDDx model of that guideline were presented in Section 3. We validated our encoding using the diarrhea test case. An example of the TiMeDDx heuristic algorithm run on the diarrhea test case is provided in Appendix A. We used 8 case vignettes to develop and fine-tune the diarrhea knowledge base and 10 additional test cases to validate it, using the heuristic algorithm. All the test cases are presented in Appendix B. The test cases and two of the training set cases were developed by a clinician who was not involved in the development of TiMeDDx and had no knowledge of it. TiMeDDx produced the expected results for all the test cases. In one test case, TiMeDDx could not differentiate between two diagnoses. The higher-ranking diagnosis (Bacterial diarrhea, non-Shigella) was the correct one, but it received a score that was just one

point higher than the diagnosis of Shigellosis. However, even experts find it hard to differentiate these two diagnoses from the presenting clinical data items.

When designing a more comprehensive evaluation study, we intend to train the system with about five patient cases for each diagnosis and to test it with five test cases for each diagnosis. The patient cases should have different characteristics and should be prepared by 2-3 clinicians who are not the developers of the KB. As was done in [45], we plan to use real patient cases as an additional evaluation.

We executed the same test case shown in Appendix A on several diagnostic DSS for broad domains: QMR, DXplain, and GIDEON. Our aim was to see how probabilistic diagnostic DSSs for broad domains perform in supporting the process of an investigation of a clinical problem (e.g., diarrhea). If they would perform well in MCM-oriented diagnosis —a task for which they were not designed— there would not be a need for special-purpose diagnostic DSSs, such as TiMeDDx. The results are shown in appendices C, D, and E, respectively. As can be seen, entering a single clinical manifestation (acute diarrhea into QMR and bloody diarrhea into DXplain) produced a DD-set that does not use abstractions but contains concrete diagnoses. In the DD-set, the correct diagnosis (Shigellosis) was not one of the top diagnoses (above 35%) in QMR. In DXplain, it was the fourth diagnosis in the rare disease list. QMR does not guide the user as to what additional data should be collected to distinguish among diagnoses, so we entered the case findings unaided, to refine the DD-set. Once again, Shigellosis was not the top scoring diagnosis. Moreover, the first score in QMR, Toxin-borne diarrhea, ranked extremely low in TiMeDDx (because it is known that the patient did not ingest suspicious food) and was eliminated by it. Similar results were obtained with DXplain. Although DXplain asked the user

about additional findings that may be present, most of them were not relevant to differentiate Shigellosis from the other diagnoses in the DD-set. This strengthens the advantage of TiMeDDx in supporting efficient investigations of clinical problems.

Running the diarrhea case in GIDEON produced better results than the QMR and DXplain runs. Upon entering the single problem "diarrhea", the correct diagnosis (Shigellosis) was ranked first. Note that, GIDEON normally prompts the user to input a few other parameters (not just one finding): disease onset time and geographical location. Entering the case's values for these parameters changed the DD; Shigellosis was no longer the top diagnosis. GIDEON did not guide us as to what other finding we should be looking for. After entering the other findings in the test case, GIDEON correctly identified Shigellosis as the top diagnosis, well separating it from the other diagnoses in the DD-set. However, GIDEON contains knowledge just for infectious diseases. Thus, naturally, GIDEON will not help in diagnosing inflammatory (non infections) or medication-change related diarrhea.

5 Discussion

Several diagnostic DSSs have already established themselves as valuable instruments for supporting clinicians in diagnosis of broad areas of medicine [7, 9, 10, 12, 13, 16, 19] and infectious diseases [11]. We focused on a niche that has been much less addressed – decision support for MCM-oriented diagnosis meant to support non-expert physicians in the process of investigating clinical problems in all fields of medicine. We present TiMedDDx —a novel model for multi phase MCM-oriented anchor based diagnosis that is conducted in phases of decreasing abstraction— and its preliminary evaluation. The information model of TiMeDDx allows it to center the diagnostic process around an anchor finding, considering all diseases that may be manifested as the MCM. Unlike other approaches that are not based on a MCM and try to relate

all findings to all diagnoses, TiMeDDx considers only diagnoses that are manifested as the anchor, enabling an efficient way of managing the diagnostic process.

In an early paper [29], Szolovits et al. suggested that any diagnostic DSS should have a model of disease and an algorithm performing clinical reasoning. Such an algorithm has the following properties, also observed in TiMeDDx: (1) it matches what is known about the patient with represented diagnostic knowledge; (2) it directs elicitation of useful information about additional findings that may be present; (3) it limits the number of hypotheses considered and performs sophisticated evaluation of a small number of hypotheses; (4) it narrows the focus of the diagnostic process using abstractions of disease hierarchies; and (5) it concludes the most probable diagnosis. Unlike the reasoning process described in [29], TiMeDDx does not have a causal disease model that supports pathophysiologic reasoning, like all other models discussed. TiMeDDx leverages the work done in early diagnostic DSSs, such as QMR and DXplain, utilizing relationships between diseases and findings that indicate evoking strength (used to reward points for diseases that explain findings present) and frequency_penalty (used to penalize a disease for findings absent). However, unlike earlier systems, TiMeDDx' heuristic algorithm does not penalize disease hypotheses for findings present that are not explained by the disease (finding_importance, used in Internist-1[15]); since the set of disease hypotheses per anchor is relatively small, such penalty need not be used in order to differentiate between the disease hypotheses per a given anchor. Since finding importance is not used in TiMeDDx its heuristic algorithm uses fewer variables to compute the score for each disease. While this saves effort in eliciting the knowledge, the ability to fine-tune the knowledge is restricted, and therefore the knowledge base needs to be validated carefully to confirm that the correct decision-support is provided for different patient cases.

Any diagnostic DSS needs to contain knowledge about relationships of findings and diseases. The difficulties of eliciting probabilities of manifestation from the literature and from experts makes it extremely hard to gather and maintain massive databases of probabilities that copes with all possible interactions [29]. This task is much smaller in TiMeDDx, for which the interactions considered are for a given MCM, and not for any possible finding and disease. The probability elicitation task is even more limited because modeling a MCM-oriented diagnostic process in TiMeDDx considers only the findings and diagnoses that are relevant for a particular anchor. Our Bayesian formulation of the phased anchor-oriented diagnostic process aligned the Bayesian model with hypothetico-deductive problem-solving process and decreased the combinations of diseases and findings for which probabilities should be defined. However, the number of required probabilities was still high. Such probabilistic data are rarely supplied by EB studies or diagnostic clinical guidelines and it is difficult to extract such data from experts. Therefore we utilized a heuristic scoring algorithm that requires eliciting fewer numbers. After training, that algorithm performed very well on the test cases of the preliminary evaluation.

To elicit data about probabilities of manifestations of findings in diseases, which is required to formulate the TiMeDDx knowledge base, we wanted to rely on numbers reported in EB studies. However, such data was not always available. When it was, we often found that EB studies reported a large range of probabilities. For example, the diarrhea guideline [44] reports that 48-100% of patients infected by *Campylobacter* species exhibit bloody stools, based on two studies that produced very different frequencies. To determine a number to use for TiMeDDx, we simply took the average, although averaging is problematic in cases where the numbers do not agree. In that case, there is an advantage of having local adaptation features in TiMeDDx; the probabilities differ according to geographical and even institutional factors.

For some of the manifestation probabilities and synergistic effects no numbers were provided in EB-studies. Therefore, we had to rely on experts' judgment. It is possible that the multi-phase model of TiMeDDx that uses abstractions may facilitate knowledge extraction from clinician experts. It has been demonstrated that expert problem-solving techniques rely on using appropriate abstractions [46] and that their reasoning processes depend on making distinctions between cases that they encounter (e.g., patients presenting with MCMs) and classification of such cases [47], as in the diagnosis process supported by TiMeDDx. Expert knowledge acquisition tools focusing on the problem-solving method have been successfully used in a variety of real-world applications [48]. While some studies point out that humans do not always provide accurate sources of probabilistic information [49], it has been shown that human experts can provide reliable information about the frequency of occurrence of manifestations given a disease [24]. After eliciting the knowledge for the diarrhea MCM and validating the resulting knowledge base, we used the diarrhea test case to contrast TiMeDDx with different probabilistic diagnostic DSSs, starting with diarrhea as a single finding. Our results showed that TiMeDDx produced very good results in the preliminary evaluation, identifying the correct diagnosis. QMR and DXplain contained the correct diagnosis in the DD-set of the rare diseases when a more refined MCM was entered (bloody diarrhea or acute diarrhea rather than diarrhea), and DXplain's suggestion about further relevant findings did not improve the diagnosis (QMR does not make such suggestions). On the other hand, the GIDEON, which was specifically built for infectious diseases, DSS included the correct diagnosis as the top-most diagnosis in the DD set even when diarrhea was the only finding entered. Although we cannot draw conclusions based on this preliminary comparison, it illustrates the differences in the decision-support between a DSS whose aim is to suggest all possible diagnoses in which some of the patient findings are present,

and rank them in order of fit with the patient's findings, and a DSS whose aim is to support the clinical investigation of a clinical problem. The test case's execution in the different systems points to the value of a system, such as TiMeDDx for supporting the investigation workup of clinical problems.

Limitations and Future Work

A limitation of our study is that TiMeDDx has not yet been evaluated in practice. The potential success of TiMeDDx depends on its performance as well as on the possibility of eliciting the knowledge required to build a useful knowledge-base. So far, only preliminary evaluation has been carried out. We are currently conducting an elaborate study to evaluate the TiMeDDx model with the test case of syncope as a MCM. In this study we are eliciting the medical knowledge from expert cardiologists in Israel.

TiMeDDx follows a problem-oriented diagnostic process that is often used by physicians. However, a second limitation is that it is possible to use TiMeDDx only for diagnosing the MCMs contained in its knowledge base. This is in contrast to other diagnostic DSSs that associate findings and diseases regardless of an MCM, thus enabling the use of their knowledge for a diagnostic process that is not necessarily based on a defined set of MCMs.
Another limitation is that while TiMeDDx' model fits many clinical problems, such as syncope or diarrhea, for which there is practically one explanation, the model might be inappropriate for a problem like anemia in an elderly patient, where more than one explanation may exist.

A further limitation is that in case the clinician observed several findings in her patient (e.g., fever, cough, rapid breathing, hyponatremia) and chose as the anchor a finding that is not the most important one (e.g., hyponatremia), then the system will not help her in reaching the correct

diagnosis (e.g., pneumonia). However, we assume that clinicians are skilled in selecting the appropriate anchor, which can be a finding or a combination of findings.

The diagnostic process supported by TiMeDDx works in phases, where a minimal effective set of clinical data items (CDIs) is considered in each phase; not all the CDIs for all phases of the diagnostic process are considered at once. If the knowledge base is built in a hierarchically-correct way, then each final diagnosis would belong to an abstraction that fits with the data collected at the initial phase. The knowledge base should be tested to validate this. Difficult cases may occur if the score of two abstractions is not very far apart (yet the higher-scoring abstraction is separated from the next one by over 10%).

TiMeDDx is still under development. Our future plans are to combine principles learned from the GLIF3 guideline modeling language [50]. While GLIF3 is not a highly suitable model for diagnosis, as its model is deterministic and does not reason under uncertainty, as done in probabilistic models, we may still leverage from it. As in GLIF3, we would like to use controlled vocabularies to specify clinical terms (findings, diseases) and to incorporate a patient information model, enabling retrieval of data from electronic medical records.
Since we are aware of the complexity that sometimes exists in real-world diagnosis, we would like to develop a user-interface that would enable the user to select any diagnosis/abstraction in the current DD-set and continue with it even if it is not the one that has the highest calculated probability, based on the knowledge in the TiMeDDx knowledge base.

Finally, TiMeDDx can potentially be relevant for other domains, such as business process management. Diagnostic business processes, such as trouble shooting and 'helpdesk' functionality start with a main presenting problem (e.g., some computer failure manifestation), span several phases of diagnosis that start from abstract to specific, and include instructions

about eliciting symptoms from the customers presenting with the problem. Such translation from a clinical to a non-clinical domain was made by the developers of the MOLE [5] system, who used their system to elicit expert knowledge and develop a DSS in domains such as clinical diagnosis and diagnosis of car problems. We plan to research the applicability of TiMeDDx to the business domain.

6 Conclusion

TiMeDDx is a prototype diagnostic system that is MCM-oriented, supports a diagnostic process that is conducted in phases of decreasing abstraction, and follows the staged reasoning process of differential diagnosis, guiding the user in a phase-by-phase manner regarding data that should be collected during the clinical investigation process. Due to the new notions modeled in TiMeDDx, and the encouraging preliminary evaluation results, we believe that there is a justification to examine the model further in the domains of clinical as well as non-clinical diagnosis.

Acknowledgement

We thank Dr. Noa Lavi for providing her expert opinions for determining quantitative values used in the diarrhea DSS, Dr. David Goldstein for preparing the test cases, and Osama Nasreldeen for his analysis of MCM-oriented guidelines found in ngc.gov. We thank Samson Tu for his insightful suggestions and comments. We thank the anonymous reviewers for their helpful comments.

Bibliography

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academy Press; 2001.

2. Musen MA, Shahar Y, Shortliffe EH. Clinical Decision-Support Systems. In: Shortliffe EH, Perreault LE, Wiederhold G, Fagan LM, editors. Medical Informatics: Computer Applications in Health Care and Biomedicine. Second ed. New York: Springer; 2001. p. 573-609.

3. Miller RA, Geissbuhler AG. Ch. 5 Diagnostic Decision Support systems. In: Berner ES, editor. Clinical Decision Support systems: Theory and Practice second edition: Springer; 2nd edition; 2007. p. 99-125.

4. Shortliffe EH, Axline SG, Buchanan BG, Merigan TC, Cohen SN. An artificial intelligence program to advise physicians regarding antimicrobial therapy. Computers and Biomedical Research 1973;6:544-560.

5. Eshelman L, Ehret D, McDermott J, Tan M. MOLE: A tenacious knowledge-acquisition tool. Intl J of Man-Machine Studies 1987;26(1):41-54.

6. Miller R, Masarie FE, JD JDM. Quick Medical Reference (QMR) for diagnostic assistance. MD Comput 1986;3(5):34-48.

7. Ramnarayan P, Kulkarni G, Tomlinson A, Britto J. ISABEL: a novel Internet-delivered clinical decision support system. In: Bryant J, editor. Current Perspectives in Healthcare Computing. Harrogate, UK; 2004. p. 245-254.

8. Shwe M, Cooper G. An empirical analysis of likelihood-weighting simulation on a large, multiply-connected medical belief network. Comput Biomed Res 1991;24(5):453-75.

9. Logic Medical Systems. PAIRS (Physician assistant Artificial Intelligence Reference System). 2007. http://cyberdoc.freewebspace.com/, last accessed June 25, 2009

10. Lincoln MJ, Turner CW, Haug PJ, Warner HR, Williamson JW, Bouhaddou O, et al. Iliad training enhances medical students' diagnostic skills. J Med Syst 1991;15(1):93-110.

11. Edberg SC. Global Infectious Diseases and Epidemiology Network (GIDEON): a world wide Web-based program for diagnosis and informatics in infectious diseases. Clin Infect Dis 2005;40(1):123-6.

12. Barnett GO, Famiglietti KT, Kim RJ, Hoffer EP, MJ MJF. DXplain on the Internet. Proc AMIA Symp. 1998. p. 607-11.

13. Weed LL, Hertzberg RY. The use and construction of problem-knowledge couplers, the knowledge coupler editor, knowledge networks, and the problem-oriented medical record for the microcomputer. Proc Symp Comp Appl Med Care; 1983. p. 831 - 836.

14. deDombal FT. Computer-aided diagnosis and decision-making in the acute abdomen. J R Coll Physicians Lond 1975;9(3):211-8.

15. Miller RA, Pople HE, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982;307(8):468-76.

16. Aliferis CF, Cooper GF, Bankowitz R. A temporal analysis of QMR: abstracted temporal representation and reasoning and initial assessment of diagnostic performance trade-offs. Proc Symp Comp Appl Med Care; 1994. p. 709-15.

17. Lemaire JB, Schaefer JP, Martin LA, Faris P, Ainslie MD, Hull RD. Effectiveness of the Quick Medical Reference as a diagnostic tool. CMAJ 1999;161(6):725-8.

18. Friedman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. JAMA 1999;282(19):1851-6. Erratum in: JAMA 2001 285(23):2979.

19. Waxman HS, Worley WE. Computer-assisted adult medical diagnosis: subject review and evaluation of a new microcomputer-based system. Medicine (Baltimore) 1990;69(3):125-36.

20. Tierney LM, Henderson M. The Patient History: Evidence-Based Approach: McGraw-Hill Medical; 2005.

21. Denekamp Y, Nasreldeen O, Peleg M. Characterization of the Knowledge Contained in Diagnostic Problem Oriented Clinical Practice Guidelines. Proc AMIA Symp. 2007, p. 929.

22. Balla U, Malnick S, Schattner A. Early readmissions to the department of medicine as a screening tool for monitoring quality of care problems. Medicine (Baltimore) 2008;87(5):294-300.

23. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med 2003;78(8):775-80.

24. Giuse DA, Miller RA, Giuse NB. Chapter 17: Strategies for medical knowledge acquisition. In: van Bemmel JH, Musen MA, editors. Handbook of Medical Informatics: Springer; 1997. p. 277-291.

25. van Ast J, Talmon JL, Renier WO, Hasman A. An approach to knowledge base construction based on expert opinions. Methods Inf Med 2004;43(4):427-32.

26. van Bemmel JH, Musen MA. Handbook of medical Informatics: Springer; 1997.

27. Peleg M, Tu SW. Decision Support, Knowledge Representation and Management in Medicine. In: Kulikowski C and Haux R, editors. International Medical Informatics Association (IMIA) Yearbook of Medical Informatics: Assessing Information Technologies for Health: Schattauer GmbH, Germany; 2006.

28. Patil RS, Szolovits P, Schwartz WB. Causal understanding of patient illness in medical diagnosis. Proc Seventh Intl Joint Conf Artif Intel; 1981. p. 893-899.

29. Szolovits P, Patil RS, Schwartz WB. Artificial Intelligence in Medical Diagnosis. Annals of Internal Medicine 1988;108(1):80-7.

30. Pople HE Jr. Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics. In: Szolovits P, editor. Artificial Intelligence in Medicine. Boulder, Colorado: Westview Press; 1982.

31. Elstein AS, Schwartz A. Chapter 10: Clinical Reasoning in Medicine. In: Higgs J, Jones MA, editors. Clinical Reasoning in the Health Professions, 2nd edition: Butterworth-Heinemann; 2000.

32. Shortliffe EH, Barnett GO. Chapter 2: Biomedical Data: Their Acquisition, Storage, and Use. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine, Third Edition: Springer; 2006.

33. Eddy DM, Clanton CH. The art of diagnosis: solving the clininocopathological conference. N Engl J Med 1982;306(21):1263-9.

34. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975;31(2(5969)):486-9.

35. Peterson M, Holbrook JH, Hales DV, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med 1992;156(2):163-5.

36. Behrman RE, Kliegman RM, Jenson HB. Nelson Textbook of Pediatrics, 17th Edition: Elsevier; 2004.

37. Fauci AS, Braunwald E, Kasper DL, Hauser SL. Harrison's Principles of Internal Medicine, 17th Edition: McGraw-Hill Professional; 2008.

38. Newell A, Simon HA. Human problem solving. Englewood Cliffs, NJ: Prentice Hall; 1972.

39. Pauker SG, Gorry GA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition: Taking a present illness by computer. Am J Med 1976;60(7):981-996.

40. Miller RA, Masarie FE. The demise of the "Greek Oracle" model for medical diagnosis systems. Methods Inf Med 1990;29(1):1-2.

41. Miller RA. Why the standard view is standard: people, not machines, understand patient's problems. J Med Philos 1990;15:581-91.

42. Gennari J, Musen MA, Fergerson RW, Grosso WE, Crubezy M, Eriksson H, et al. The Evolution of Protege: An Environment for Knowledge-Based Systems Development. Intl J Human-Computer Interaction 2002;58(1):89-123.

43. Hoffer JA, George JF, Valacich JS. Modern System Analysis and Design, Third Edition: Prentice-Hall; 2002.

44. Guerrant RL, Gilder TV, Steiner TS, Thielman NM, Slutsker L, Tauxe RV, et al. Practice guidelines for the management of infectious diarrhea. Clin Infect Dis 2001;32(3):331-51.

45. Peleg M, Wang D, Fodor A, Keren S, Karnieli E. Lessons learned from adapting a generic narrative diabetic-foot guideline to an institutional decision-support system. In: A. ten Teije, Miksch S, Lucas P,

editors. Computer-based Medical Guidelines and Protocols: A Primer and Current Trends, Studies in Health Technology and Informatics; 2008. p. 243-52.

46. Patel VL, Ramoni M. Cognitive models of directional inference in expert medical reasoning. In: K. Ford PF, & R. Hoffman, editor. Human & Machine Cognition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1997. p. 67-99.

47. Gaines BR, Shaw MLG. Integrated Knowledge Acquisition Architectures. J Intel Info Sys 1992;1(1):9-34.

48. Steels L. Components of Expertise. AI Magazine 1990;11(2):28-49.

49. Baroni P, Guida G, Zanella M. Managing uncertainty in diagnosis of acute coronaric ischemia. Artif Intel Med 2001;23(2):129-47.

50. Peleg M, Boxwala AA, Tu S, Zeng Q, Ogunyemi O, Wang D, et al. The InterMed Approach to Sharable Computer-interpretable Guidelines:A Review. J Am Med Inform Assoc 2004;11(1):1-10.

Figure and Table Legends

Figure 1. An instance of the Anchor class.

Figure 2. Specification of an anchor: Fever. Fever is a required CDI for the anchor of acute diarrhea, which is collected during the physical exam. The insert on the right shows the allowed value type of this CDI: temperature. Temperature is an instance of Categorical_Value. The categories of temperature are low or high.

Figure 3. An Entity-Relationship Diagram showing the information model of the three relationships used to define an Anchor. Concepts are shown as rectangles and relationships as diamonds. Multi-cardinality relationships are marked with n, m, p, and r. For example, part (a) represents the following statement: "Anchor is manifested as n (many) Diagnoses or Abstractions". Properties of relationships are written below the diamond symbols. Examples of entity relationships instances are provided in parentheses. The arrows mark the directionality in which the relationship should be read. (a) Diagnosis_Manifested_ As_Anchor – the example shows that a diagnosis of Hepatitis B is manifested as jaundice with a probability of 17 (out of 20); (b) CDI_Evokes_Diagnosis_In_Anchor – the CDI fever with a value of high or low, in an anchor of infectious diarrhea evokes the diagnosis Bacterial (non-Shigella) diarrhea with an evoking strength of 7 (out of 10) and a frequency_penalty of 5; (c) Temporal_CDI_Relationship_ Given_Diagnosis_For_Anchor – If the time of ingesting suspicious food is 3 +/- 3 hours before the time of onset of diarrhea, in an anchor of Infectious Diarrhea, the diagnosis of Toxin-borne diarrhea is evoked with weight of 10 (out of 10) and 9 points are penalized from that diagnosis hypothesis for the absence of ingestion of suspicious food (frequency_penalty).

Figure 4. Definition of a temporal synergistic relationship between a pair of CDIs that evokes a diagnosis in a given anchor. The evoking strength is specified by the slot "weight", whose range is 1..10. The figure shows that when the anchor is infections diarrhea, then the occurrence of ingestion of specific food 3 hours (+/- 3 hours) before the onset of diarrhea suggests, with weight 10, the diagnosis of Toxin-borne diarrhea. But, when ingestion of suspicious food is not present in the patient, 9 points are penalized from the diagnosis of Toxin-borne diarrhea.

Figure 5. The diagnoses scoring algorithm

Figure 6. The scores of the hypotheses defined in the DD-set of the current phase of the algorithm, shown in Figure 1. Numbers appearing in italics (the first number in each row) are based on the Diagnosis_manifested_as_anchor relationship. Numbers appearing in regular font are based on the CDI_evokes_Diagnosis relationships (evoking strengths and penalties). The scores of the Dxs that are above the cutoff are shown in bold.

Figure 7. Bayesian Networks (BN) for the three phases of diagnosis of diarrhea derived from the knowledge represented in Table 2 with uniform prior probabilities of disease hypotheses. (a) phase 1 BN. The top insert shows the prior probability of the two abstractions: acute and chronic diarrhea. The uniform prior probabilities are shown for illustrative purposes. The bottom insert shows the conditional probabilities of P(F|D); (b) BN for computing prior probabilities of disease hypotheses of phase 2 BN, relying on the "diagnosis manifested as anchor" relationships of Table 2(b); (c) BN corresponding to phase 2. The prior probabilities are those derived by computing the posterior probability based on the BN shown in part (b). The conditional probabilities for P(high fever |D) are shown for illustrative purposes; (d) BN corresponding to phase 3. The BN were created using the GeNIe tool (http://genie.sis.pitt.edu/) and reproduced in the figure.

Table 1. Characteristics of diagnostic DSSs currently available

Table 2. A set of three tables showing a path in the DD-tree. Each table represents a different layer (phase) in the tree. The second column shows the disease manifested as anchor relationship data. The other cells in the tables show evoking strengths and penalties (penalties are shown in parentheses) for diagnoses given an anchor. The anchor is *diarrhea* in phase 1, *acute diarrhea* in phase 2, and *infectious diarrhea* in phase 3. The numbers in table *b* were elicited from experts and the numbers from tables *a* and *c* were elicited from EB sources.

Table 3. A table for eliciting synergistic temporal relationships between CDIs and diagnoses, for the anchor of *infectious diarrhea*.

	QMR	DXplain	Iliad	GIDEON	QMR-DT (PAIRS)	Isabel	Problem Knowledge Coupler	TiMeDDx
Sources of knowledge	Literature and experts	Literature and experts	Literature	Literature	Literature	Literature	Literature	Literature and experts
MCM-oriented	-	-	-	-	-	-	+	+
Starting point can be a single clinical data item	+[1]	+[1]	+	+	+	+	+	+
DD shown at each stage	+	+	+	+	+	+	Can be requested	+
Advice on data/tests that should be collected	+/- Advice is not actively offered. Users can invoke menu option for ruling-in and ruling-out diagnoses	+	+	-	-	- Collection of text material indexed with the disease concepts is accessible	+	+
Representation of temporal relationships between data items	-	-	-	+	-	-	-	+
Synergism between data items	-	-	+	+	+	-	-	+
Consideration of doctor's clinical reasoning , in which clinical investigation is done in stages of decreasing abstraction	-	-	-	-	-	-	-	+
Local adaptation of system	-	+some support	-	+	-	-	possible	+
Ability of user to override system recommendations, selecting a different DD set	+	+	-	-	-	-	+	+
Consideration of all hypotheses	+	+ Rare disease displayed separately	+	+	+	+	Only relevant diagnoses are listed	Only relevant diagnoses are listed
Scope of knowledge	Diseases and findings for internal medicine	Diseases and findings for internal medicine	Diseases and findings for internal medicine	Diseases and findings for Infectious diseases	Diseases and findings for internal medicine	Diseases and findings for internal medicine, pediatrics, geriatrics	Knowledge relevant for MCM	Knowledge relevant for MCM
Explanations provided	+ what findings support a	+ what findings support a	+	+	-	-	References to literature may be	Theoretically-possible but not

	diagnosis	diagnosis					provided	implemented
Computational model	Disease/finding Relationships	Disease/finding Relationships	Bayes + decision rules	Bayes	Bayes	Pattern-matching algorithms	Disease/finding Relationships; predecessor/successor relations between entities. No ranking of disease hypotheses	Multi phase, anchor based, relational or Bayesian

[1]focusing the diagnosis on selected findings is possible

Table 2

(a) Phase 1 anchor: diarrhea		CDI	
Diagnosis	Dx manifested as diarrhea	Duration ≤14 days	Duration >14 days
Acute Diarrhea		pathognomonic	
Chronic Diarrhea			pathognomonic

(b) Phase 2 anchor: acute diarrhea		CDI					
Diagnosis	Dx manifested as acute diarrhea	High fever	Abrupt presentation	Nausea/ vomiting	Mucus	Arthritis	More people developed
Infectious Diarrhea	16	9	9	8	1		9
Medication change	8		7	3			
Inflammatory Bowel Disease	14	4	2	2	8	9 (-5)	
Intermittent Bowel Obstruction	10		2	2			
Colonic Ischemia	8		2	2			

(c) Phase 3 anchor: infectious diar		CDI							
Diagnosis	Manifested as infectious diar.	Fever	Abdominal pain	Tenes mus	Nausea /vomit	Watery diarrhea	Flu-like symptom	Bloody stools	Recent antibio.
Shigella		8(-7)	8	7(-7)	8			4	
Non-Shigella bacterial		7(-5)	7		4			3	
Clostridium		3	2						8(-6)
Parasitic		1	3						
Food-borne*		0							
Viral		8	2		8	9	7(-8)		

* disease-finding relationships for the Food-borne diagnosis are provided in Table 3.

Diagnosis	CDI-1	CDI-2	Time between two CDIs	Weight (penalty)
Food-borne diarrhea	Time of onset of diarrhea	Time of ingestion of suspicious food	0-6 hours	10 (-9)

Table 3

Figure 1

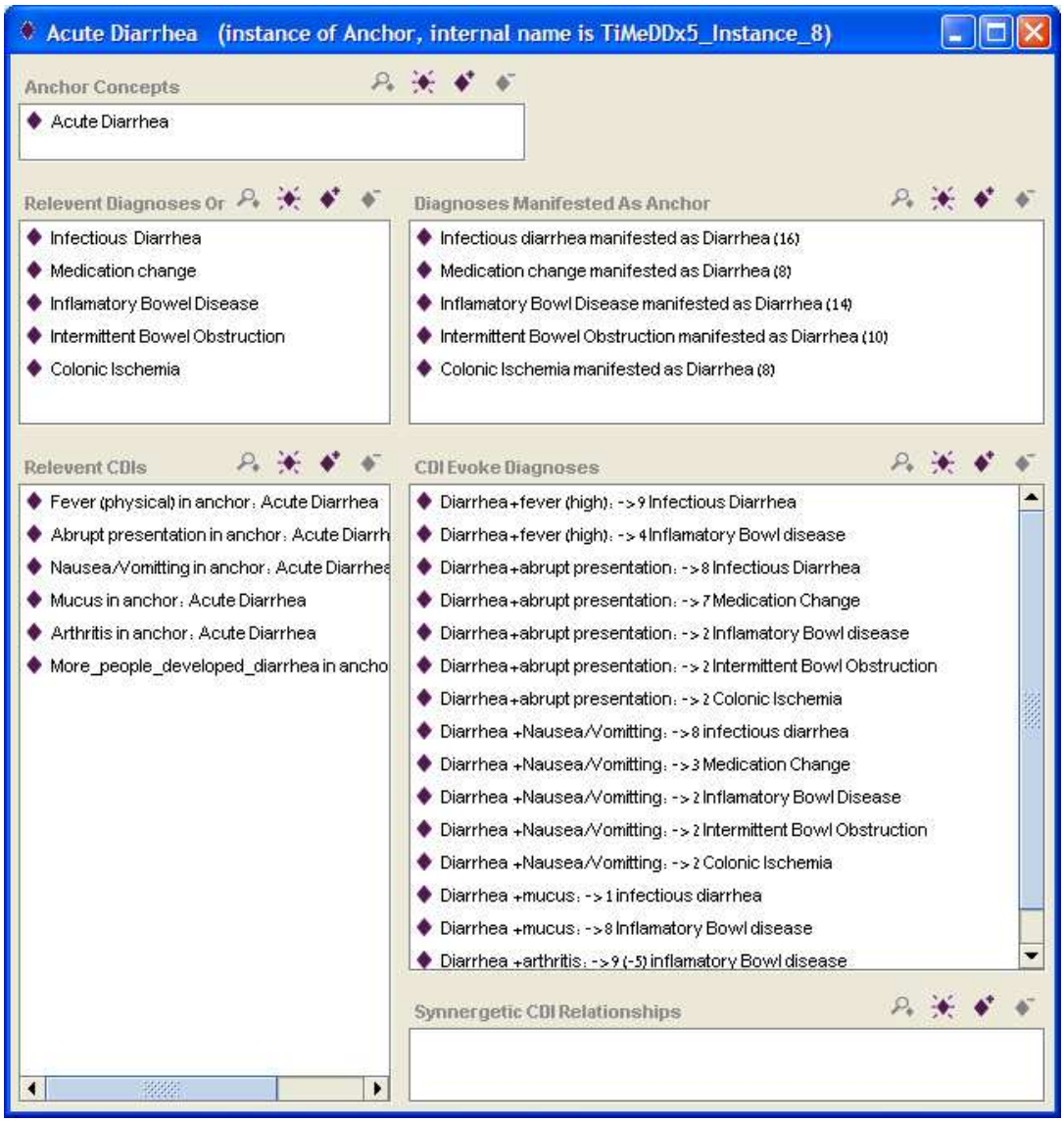
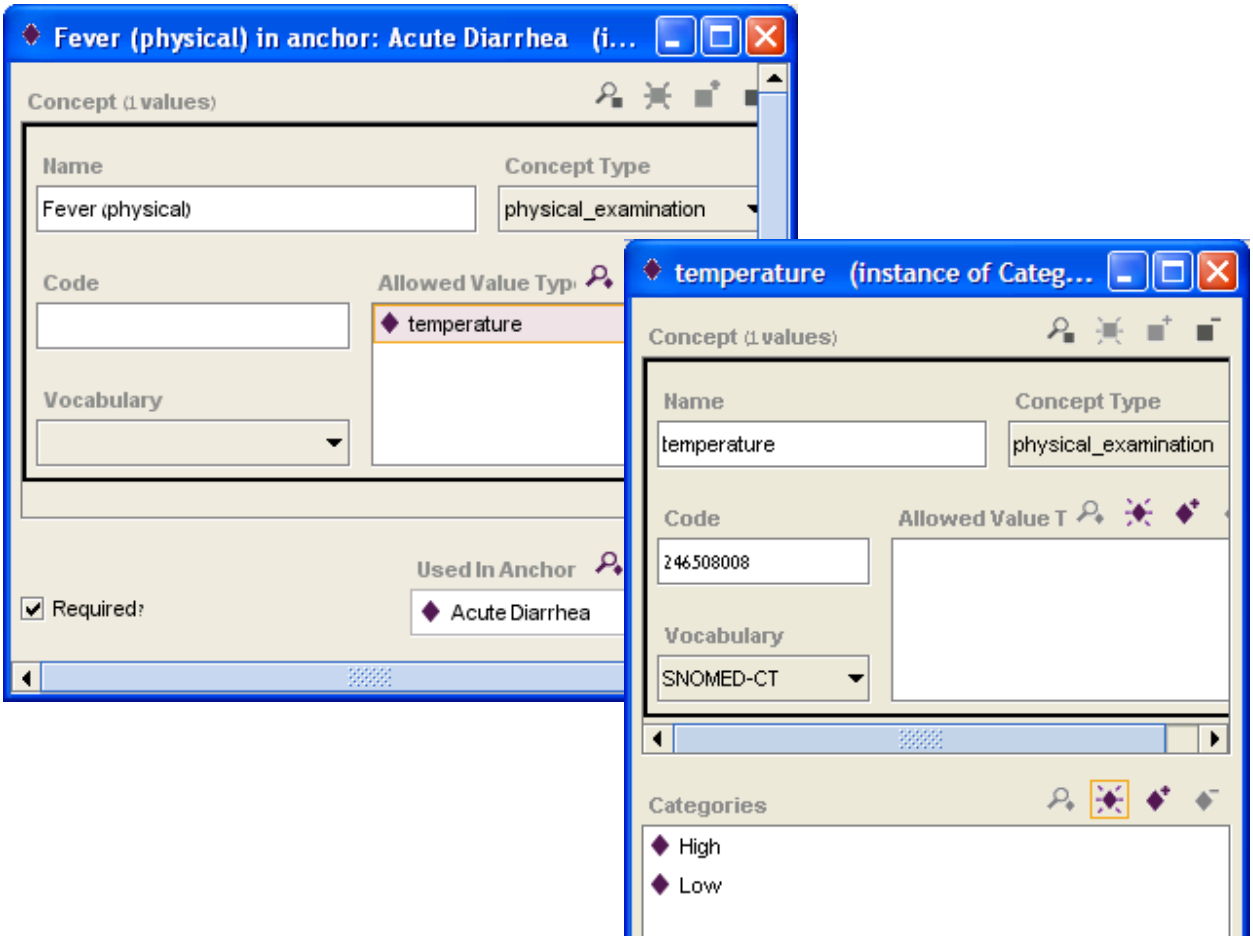


Figure 2


Fever (physical) in anchor: Acute Diarrhea (i...
Concept (1 values)
Name
Fever (physical)
Concept Type
physical_examination
Code
Allowed Value Typ
temperature
Vocabulary
Used In Anchor
Required?
Acute Diarrhea
temperature (instance of Categ...
Concept (1 values)
Name
temperature
Concept Type
physical_examination
Code
246508008
Allowed Value T
Vocabulary
SNOMED-CT
Categories
High
Low

Denekamp and Peleg Figure 4 diagnostic decision-support model

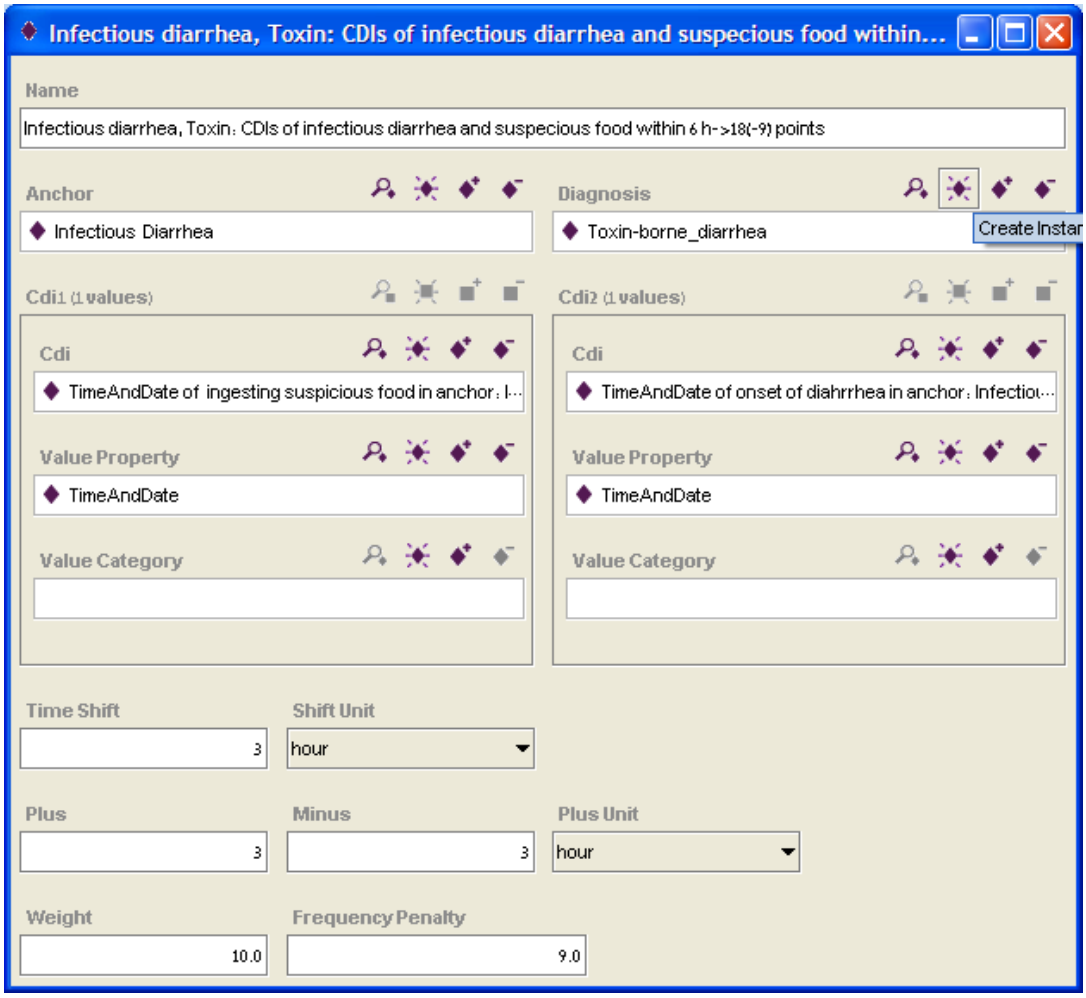


Figure 5

For a given anchor,

For each possible diagnosis Dxi calculate the following score:

+ (0..20) points if DXi is manifested as the anchor

For each CDIj for which (CDIj-evokes-Dxi) exists:

If the patient exhibits the CDI and the CDI is pathognomonic,

exit the algorithm, setting Dxi as the only Dxi that should be considered

+evokes (0..10) of CDI if the patient exhibits the CDI

-penalty (0..10) of CDI if the patient does not exhibit the CDI

For pairs of CDIs that have synergistic effects (CDI Relationship)

+weight (0..10) if both CDIs are present in the patient

-penalty (0..10) if not both CDIs are present in the patient

Figure 6

Infectious Diarrhea *16*+9+8+8 =**41**

Medication Change *8*+7+3 =18

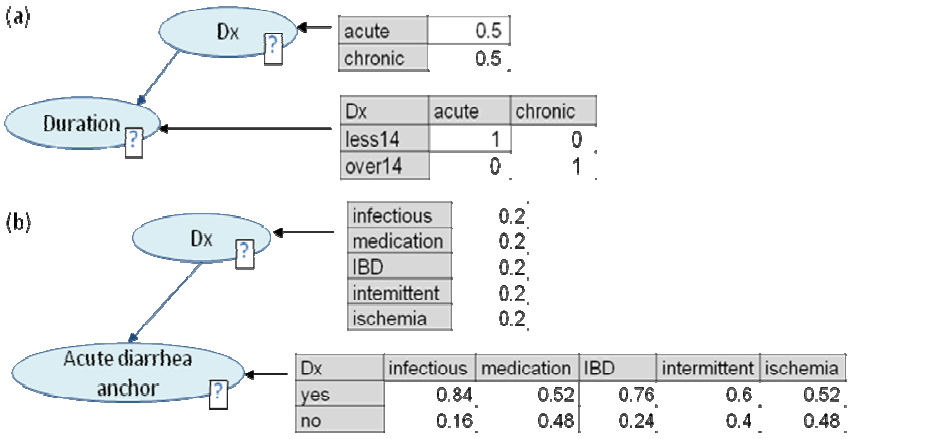
Inflammatory Bowl Disease *14*+4+2+2-5 =17

Intermittent Bowel Obstruction *10*+2+2 =14

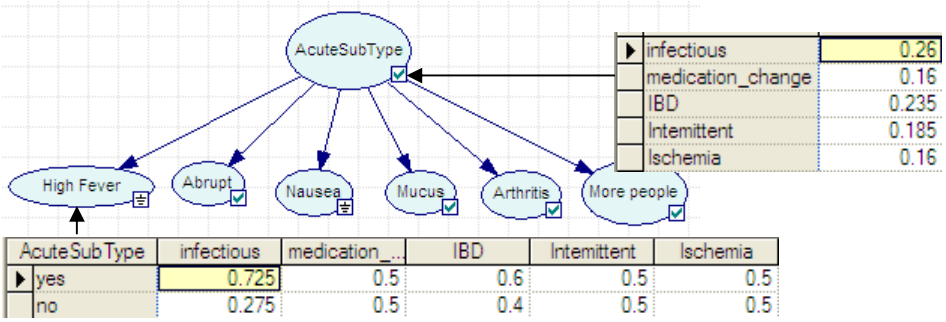
Colonic Ischemia *8*+2+2 =12

Cuttoff = 41 – (0.1*41) = 36.9

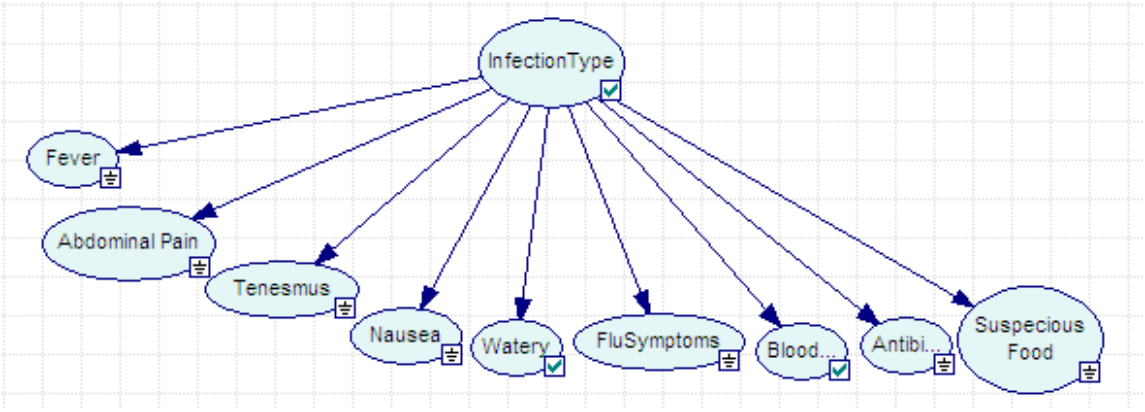
Figure 7



(c)



(d)


InfectionType
Fever
Abdominal Pain
Tenesmus
Nausea
Watery
FluSymptoms
Blood...
Antibi...
Suspecious Food