SPECIAL COMMUNICATION

A Formal Model of Family Medicine

Walton Sumner II, MD, Miroslaw Truszczynski, PhD, Victor W. Marek, PhD, DSc

Background: **The American Board of Family Practice is developing a computer-based recertification process. An optimal implementation requires a formal model of family medicine, which will become the basis for a knowledge base.**

Design: **The proposed model of family medicine contains six entities: Population, Record, Agents of Change, Health States, Findings, and Courses of Action. The model illustrates 15 important relations between entities. For instance: Health States Lead to Health States, and Findings Associate with Health States. These two relations describe natural history, manifestations of disease, and the effects of medical interventions and risk factors. Because time is such an important aspect of primary care, nearly all numeric data are represented as graphs of possible values over time, called Patterns, which include details about periodicity. Patterns and other aspects of the model provide a means of describing covariance between observations, such as the influence of height on weight.**

Results: **The model reflects many family practice activities and suggests some formal descriptions of family practice. For instance, diagnostic activities focus largely on classifying early or short segments of Patterns in Findings. Most medical interventions attempt to alter either the probability distributions in a Lead-to relation or the impact of a Finding.**

Conclusion: **The proposed model of family medicine could find uses in many applications, including computer-based tests, medical records, reference systems, and decision support tools. (J Am Board Fam Pract 1996; 9:41–52.)**

The American medical specialty boards certify physicians to practice medicine in their respective fields. In theory, board certification indicates at least minimal competence in an area of advanced training. Minimal competence is not precisely defined.

The American Board of Family Practice (ABFP) currently administers paper-based certification and recertification examinations. These examinations contain conventional item formats, case descriptions, and still photographs. Recertification examinations allow physicians to select portions of the test according to their interests. Both examinations intend to engage and test problem-solving skills, disease recognition, and fact retention.

The ABFP would like to expand its testing process to (1) allow testing at remote sites and convenient times; (2) uniformly test an expanded range of important family practice activities, with fewer questions on exotic problems; (3) adapt tests to examinees' responses or needs; and (4) create reasonable questions at test sites to simplify administrative, economic, and especially security issues. In 1992 the ABFP initiated a project to develop a computer-based recertification examination as a means of achieving these goals.[1] We were asked to propose an approach to this task. Because of concerns regarding the cost and feasibility of computer-based recertification, the ABFP and an independent review panel reevaluate this project annually.

A Vision for Recertification

The recertification process might unfold as follows. After initial certification, examinees would install recertification software from the ABFP on their own computer systems. The examinee could begin recertifying at any convenient time and could suspend the examination at the conclusion of any simulated patient encounter. The software would present a patient by using text, illustrations, still pictures, and video. The examinee

Submitted, revised, 13 September 1995.

From the Departments of Family Practice (WS) and Computer Science (MT, VWM), University of Kentucky. Address reprint requests to Walton Sumner II, MD, Department of Family Practice, University of Kentucky, Lexington, KY 40536-0284.

This work is sponsored by the American Board of Family Practice.

would question and examine the simulated patient, reach conclusions about the situation, and suggest treatment options. The simulated patient might express preferences about these options. After receiving a treatment plan, the patient would leave, maybe follow the plan, and perhaps later return for follow-up. In the meantime, the examinee would see other simulated patients. To discourage cheating, the software could offer so many cases that a diplomate observing another examinee recertify would gain little advantage with regard to test content.

The software would maintain records of the information gathered, the hypotheses entertained, and the recommendations made for each patient. After monitoring performance on several similar cases (for instance, cases involving diagnosis and management of adult-onset diabetes mellitus), the program would draw conclusions about the physician's ability to handle this class of problems. If competence has been demonstrated, the class of problems could be removed from further consideration for several years. Until competence has been demonstrated, the physician would receive feedback on specific areas for improvement and would continue to see cases from this class of problems.

Recertification could eventually become a continuous learning experience at the office or home. Some recertification activities might qualify as continuing medical education, partially offsetting the time needed to recertify. Examinees could anticipate failure to recertify and take corrective measures years before actually failing. The ABFP might be able to document examinees' understanding of cost-effective management strategies.

The Challenges

An obvious disadvantage of this approach is that it requires some sophisticated software. We reviewed similar software development efforts, such as CBX,[2] and became aware of substantial expenditures associated with these efforts. The ABFP needs an approach that does not incur high maintenance costs if it is to maintain its tradition of efficient and affordable examinations.

A less obvious problem is the lack of a formal model of family medicine. A formal model of family medicine is an absolute prerequisite to a relevant and realistic implementation of this kind of computer-based examination.

A model describes the kinds of information that could be collected regarding a topic. For instance, a model of a mailing address should include at least a name, street address, apartment number, city, state, and ZIP code. A database built upon this model could list these items for each entry. Not every item in the model should be described for every entry in the database: many addresses have no apartment number. Incomplete database entries still provide useful information: even if a street address is missing, the city to search can be found. Finally, the model limits what the database could do; it could not easily list first names.

Models of diagnostic medicine typically include diseases, historical and examination data, and links between diseases and data. These models represent knowledge that physicians apply to uncertain or imprecise cases. The address example suggests a list of simple observations, called a database. A diagnostic program uses a collection of more abstract information, such as a statistical summary of a database, to draw inferences about a single case. The program and its information are often called a knowledge base. Again, broad concepts in a medical model might not apply to particular examples (information should be missing), an incomplete knowledge base could still provide insights (some information can be missing), and the design of a medical model determines how it can be used (the structure of information is important).

A well-designed model might support automatically created case simulations, reducing the long-term cost of writing cases by hand and improving security. Simulations have been created from medical knowledge bases before, but often without an appropriate level of realism.[3-6] Researchers have described several problems, often related to simplified models of medical knowledge, which limit the capabilities of these systems.

First, medicine is full of diagnostic complexities. Diseases interact, and no one has attempted to capture such subtleties in a 600-disease knowledge base. Thus, most knowledge bases do not "know" that diabetes could change the severity of pain experienced during an acute myocardial infarction. Without this information, the knowledge base is unable to support a realistic simulation process—a simulated diabetic having an acute myocardial infarction will experience the same discomfort that anyone else would. The current project can attempt to define carefully interactions for the much smaller number of health states that constitute the bulk of family medicine.

Second, diagnosis and patient management are inextricably linked to time. Time receives relatively little attention in many knowledge bases and is often summarized very succinctly. For instance, a knowledge base might describe "chest pain lasting more than 30 minutes" as a symptom of acute myocardial infarction. This knowledge base could misinterpret 29 minutes of chest pain as evidence against acute myocardial infarction, and 2 years of chest pain as an indicator of acute myocardial infarction. The current project must also support the related concepts of continuity of care and observation.

In addition to these problems, family physicians deal with a host of issues not routinely modeled in diagnostic software. Most of these issues reflect the overwhelming importance of patient management in family medicine.

First, family medicine occurs in a social context that is often ignored in computer-generated simulations. Knowledge bases might allow a family history of illnesses or an exposure to tuberculosis but do not model social interactions or family structure.

Second, family practice patients arrive with attitudes shaped by experience, and physicians must adjust their strategies to cope with those attitudes. Adjustments range from changing interview style to altering treatments. Variability in patient attitudes limits the likelihood that there exists one best answer for groups of patients with similar medical conditions.[7]

Third, family physicians are not so much engaged in diagnosis as in helping patients improve the length and quality of their lives. Family physicians spend considerable time reassuring worried patients, alleviating symptoms, and preventing the onset or progression of disease.

The final recertification problem, evaluating the responses of diplomates, also requires a model of what family physicians do. After all, they do not serve society by answering multiple-choice questions. They try to help people. Performance data gathered by the envisioned process will not (and should not) meet the assumptions underlying current test theory. A meaningful evaluation requires development and validation of new psychometric methods, which in turn require some understanding of how family physicians help people.

All dichotomous evaluations, especially pass-fail tests, use arbitrary standards. The challenge is to set standards using generally agreeable and meaningful criteria. The ABFP will determine to whom the criteria should be agreeable—certainly to diplomates, but perhaps also to patients, insurers, or others. Specifying these "customers" will help the ABFP establish meaningful criteria for certification decisions. For instance, diplomates have an interest in maintaining respected credentials, patients want effective care, insurers desire low costs, and public health advocates have an interest in clinical guidelines. It is not at all clear how to respond to these diverse interests. We hope to deliver flexible models to describe the consequences of family practice activities, as seen by various parties, so that board certification remains a pertinent process regardless of changes in the health care system.

The State of the Project

We began our development effort by designing a formal model of family medicine. The model attempts to describe the scope of family medicine in epidemiologic terms, while including the information about individual variation that differentiates individualized patient care from public health.

The model will be the foundation of a family practice knowledge base storing data about family medicine. The family practice knowledge base should include slots for information that the ABFP might consider important to family medicine, regardless of whether those data have been gathered previously. It should permit graceful growth in complexity and volume of data and facilitate periodic updates. It should allow automated data-acquisition tools to supply as much information as possible from electronic medical records or other data repositories.

The formal model of family medicine is largely complete. Although the model requires occasional modifications to support maturing patient-simulation processes, the general structure has remained stable for 2 years. We have discovered a number of interesting features that generally confirm or elaborate broad family medicine concepts. Furthermore, the model appears to support many kinds of tests, from traditional multiple-choice and true-or-false questions through its intended goal of supporting interactive, real-time patient simulation.

The model might also support other applications of benefit to family physicians. Specific software applications might involve medical records, structured vocabularies, medical reference tools, decision support systems, and continuing education programs. We describe here an overview of the formal model of family medicine and briefly illustrate how the model might influence other tools for family practice. More detailed descriptions of the model will follow in publications presenting algorithms.

Methods

The current test development process, test questions, test content, and future goals of the ABFP were reviewed with attention to enhancements that a computer-based test could deliver. Data structures to describe the activities of family physicians were developed as a series of entity-relation diagrams, a standard approach to database development.[8-10] In an entity-relation diagram, entities usually represent things (nouns). The relations (verbs) illustrate how the entities interact. For instance, an entity-relation diagram of an address list might have an entity called "person," and an entity called "place," connected by a relation called "is at." One could read this diagram, "person is at place." The person entity would store people's names, the place entity would store addresses, and the "is at" relation would describe when and why this person is at that place. Thus, a person could now live at one place, previously live at another place, and continuously work at the first place. One person, two places, and three "is at" relations describe this address history. This address model is more flexible and realistic than the model described in "The Challenges" section above.

Diagrams were tested for conceptual completeness by comparison with (1) teaching and clinical activities at the Department of Family Practice at the University of Kentucky, (2) questions from previous ABFP examinations, and (3) anticipated ABFP examination goals. Primary care physicians with related experience reviewed the resulting diagrams and suggested further changes and areas for clarification. Entities and relations were added or reconfigured as necessary.

During this process an important class of events became apparent, which we called "modifying relations," or modifiers. In database terms, modifiers are relations between traditional relations. Modifiers extend the conventional entity-relation diagram and provide a means of managing statistically dependent events.

Model Structure

The family medicine model includes the major entities, relations, and modifying relations shown in Figure 1. The following discussion illustrates the model in more detail, with examples of important features. Formal concepts in the model are capitalized throughout the text. The model emphasizes diagnostic and management issues, variability in populations, and time. It describes consequences of anatomic and physiologic processes, but largely omits anatomic and physiologic reasoning as such. It can describe interpersonal relationships, but does not currently include an explicit representation of families or communities.

Major Entities

The following major entities appear in the design: Populations, Records, Health States, Findings, Courses of Action, and Agents of Change.

Populations represent real humans; their relations should precisely describe all data that physicians consider. Populations can be large groups with a shared characteristic, such as white males or single-parent families. An individual patient is a Population of 1; a pregnant woman is a Population of 2; a nuclear family with 2 children and 2 parents is a Population of 4.

Records model beliefs about people; a Record's relations summarize inferences about a Population. If a parent brings an infant to the office, this design represents the infant as a Population, the parent as another Population, and the parent's description of the infant as a Record. The physician can obtain historical information about the infant from two sources: the physician's medical Record of the infant, and the parent's Record of the infant. The physician can obtain current objective information by examining the infant as a Population. The data linked to Populations are absolutely precise, but can be observed, if at all, only during medical encounters. Records summarize the history of those real data imprecisely and potentially inaccurately.

Populations have Records of themselves, modeling a patient's self-image and memories. As with

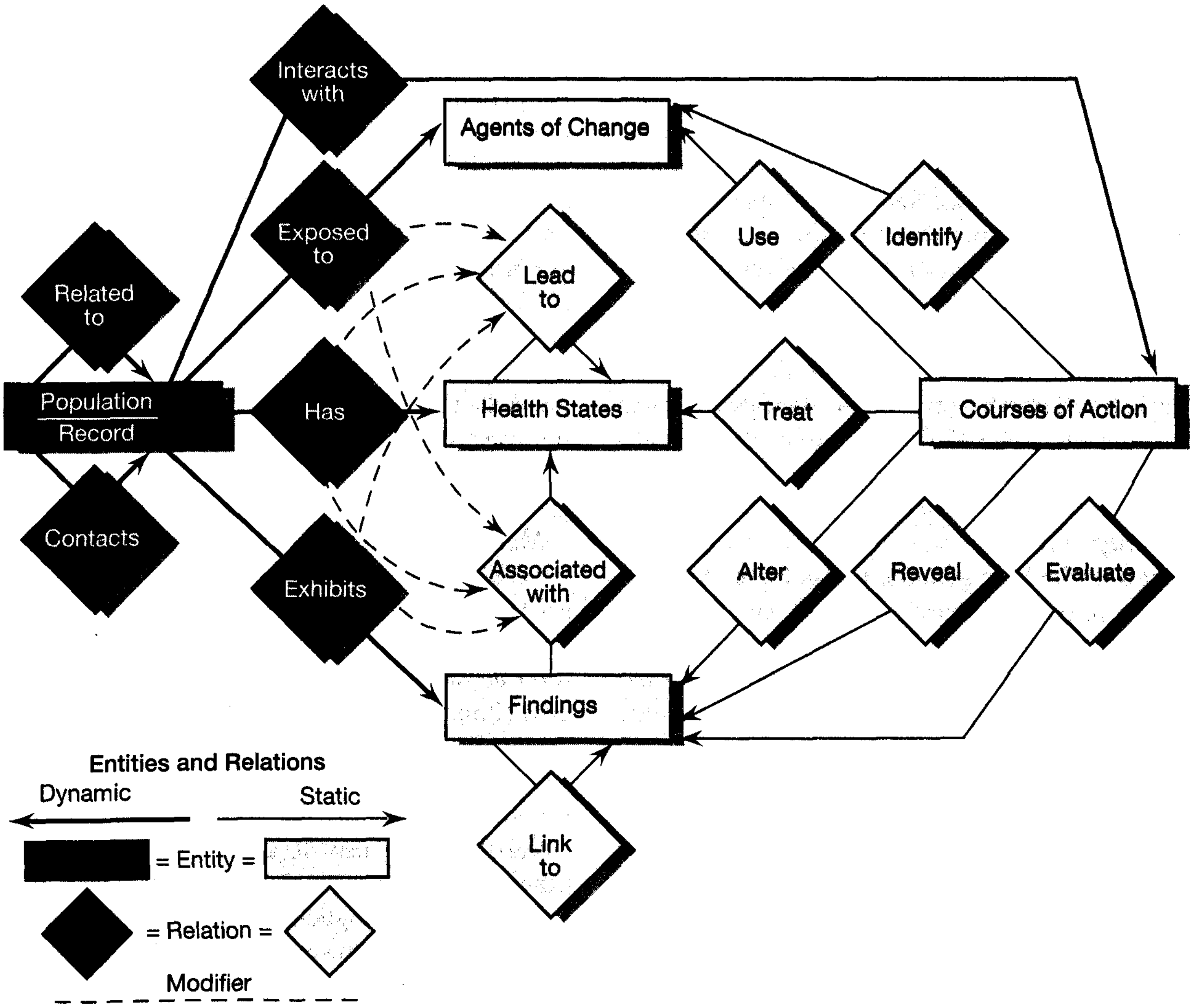

Figure 1. An entity-relation diagram of a formal model of family medicine. Modifiers are relations that might change values in other relations. Dynamic entities and relations, in black, contain information relevant to patient simulations. Dynamic information for an individual patient will be derived from data in other dynamic and static entities and relations. The dynamic Record entity has relations mirroring the Population's relations. Static entities and relations contain the best available medical knowledge, similar to data in medical literature.

other Records, a patient's self-Record summarizes historical information with variable accuracy and might be the physician's only source of some historical information.

A Population is primarily a list of relations with other entities. A Record not only lists relations with other entities, but also defines encounters during which these relations were discovered. A Record can contain conflicting data acquired at different encounters.

Health States

Health States include all normal health states; classic disease presentations; early, subtle, or late disease presentations; and some disease combinations. Health States also include groups of Health States with shared characteristics, such as cardiovascular diseases and diseases of glucose intolerance. The SysteMetrics corporation publishes *Disease Staging Clinical Criteria*, which define numerous stages in the development of diseases.[11] Each of these stages would represent a distinct Health State entity in this design. The SysteMetrics staging of diabetes mellitus defines stage 1.1 as asymptomatic diabetes, stage 1.2 as symptomatic diabetes, stage 1.3 as type I diabetes mellitus, and stage 2.1 as diabetes with end-organ damage. Each of these stages defines at least one Health State by the presence of specific objective criteria. Stage 2.1 might be divided into a group of Health States representing each damaged end organ. To represent multiple end organ damage, one might simply superimpose these states. When combinations of Health States greatly influence patients' clinical appearance, treatment choices, or prognosis, the combination should be treated as a distinct Health State. For instance, type II diabetes mellitus and hypertension commonly occur concurrently, and the presence of each disease alters treatment selections and the prognosis of the other disease. Initially we expect to define Health States using the SysteMetrics staging data. The ABFP will modify that classification as necessary to meet testing goals.

Findings

Findings include genetic, physiologic, symptomatic, physical, and test-generated data, and clusters of such data. For instance, musculoskeletal chest pain could be a Finding. This Finding could be an example of a Finding called chest pains, which would represent all kinds of chest pains. Chest pains could be an example of a still larger Finding called symptoms. Findings are defined by a collection of one or more Features, whose current value can be described by a number on a scale. One Feature pertinent to pain is severity, which might be described on a 10-point scale.

Structures called Patterns describe the possible values of each Feature over time. A Pattern typically lists a series of values and corresponding percentiles at several points in time. Pediatric growth charts are the most widely used real example of Patterns. A blank growth chart illustrates at least the following observations: (1) Normal birth weights vary within a narrow range. (2) Weight increases relatively rapidly in the first few months and years. (3) The absolute variation in weight (eg, the difference between 90th and 10th percentile weights) increases after birth. (4) Most people reach a fairly constant weight by early adulthood. A Pattern listing 10th and 90th percentile weights for people at age 0, 1 year, 2 years, and so on, illustrates the same concepts.

Growth charts also predict future values from past information. A child at the 50th percentile for weight now is expected to stay near the 50th percentile. If this child later reaches the 5th percentile of weight, the expected Pattern is absent. The ensuing diagnostic evaluation is an effort to account for the deviation by finding a weight Pattern that explains all observations. These concepts extend easily to many other values, such as temperature. People have an average temperature of about 37°C, but some are a little cooler and some a little warmer. Normal temperature fluctuates within a narrow range during a lifetime, and most deviations from that range are considered abnormal. A more interesting example would be ST segments on an electrocardiogram. Following an acute myocardial infarction, ST segments usually rise by varying amounts, fall, and return to normal. The ST segment deviation from base line varies with time and can be described by a Pattern, similar to the variation in weights of growing children.

Many values change in predictable ways. Patterns might have cycles, sub-Patterns, and sub-sub-Patterns to describe these changes: the average value of a variable often changes during a lifetime, while the instantaneous value depends on a combination of annual, lunar, and circadian cycles. For instance, a nonpregnant 20-year-old woman should experience predictable lunar and circadian temperature fluctuations.

Sub-Patterns also describe consequences of other events, such as taking a drug. For instance, a dose of acetaminophen might lower a fever for 4 hours. A fever responsive to acetaminophen could be modeled by a high-temperature Pattern with a sub-Pattern indicating 4 hours of normal temperatures following acetaminophen doses. A person experiencing this fever and taking acetaminophen every 4 hours maintains a normal temperature. A physician observing this temperature Pattern would need to halt the acetaminophen to distinguish between a normal temperature and a fever responsive to acetaminophen.

Sub-Patterns very precisely characterize Features and therefore Findings. For instance, one of the chest pain Findings might be "crushing substernal chest pain relieved by rest or nitroglycerin and exacerbated by exertion." This description implies a Finding with a designated location, a "crushing" Feature with some Pattern, and 3 sub-Patterns describing the effect of rest, nitroglycerin, and exercise. The clinical appearance of simulated patients with this Finding might still vary, depending on the allowed variation in sub-Patterns. For instance, pain might be more quickly relieved by nitroglycerin than rest or vice versa.

Finally, Patterns include Shape Selectors that help maintain consistency between variables. Shape Selectors are an example of Reasoning Elements, small programs loosely based on the structure of Arden syntax medical logic modules.[12] Reasoning Elements define variables; assign their values from data about the simulation; use loops, "if . . . then" statements, equations, and random numbers to reach conclusions; and finally produce some output. In Findings, the Shape Selector produces one percentile curve to represent the values of a Feature in an individual patient. For instance, although pediatric growth charts allow considerable variation in normal height and weight, one child will exhibit a precise series of values for both height and weight. Height will closely track one percentile curve, as will weight. The percentile of the height curve often limits the possible percentiles of the weight curve: healthy children at 95th percentile height rarely exhibit 5th percentile weight. Most children follow a weight percentile equal to the height percentile ± 20. The weight Shape Selector can use this equation to restate the familiar height-weight growth chart.

Patterns represent our major effort to model time and one approach to interrelated medical observations. Time affects most numeric values in the model. Consequently, Patterns appear in nearly every entity and relation. Patterns describe the incidence of diseases at different ages, the likelihood of diseases progressing with time, and concentrations of drugs.

Courses of Action

Courses of Action represent people's activities. Not only can these activities be medical, such as taking a blood pressure or performing a coronary artery bypass graft, but they can also include attending school, working, asking and answering questions, and following advice.

Populations invoke Courses of Action to decide when to visit a physician, how to answer questions, and whether to follow advice; therefore, Courses of Action could be written to miss appointments, lie to physicians, and ignore their advice. These actions could even depend on aspects of the physician's conduct, such as how the physician chooses to obtain information. Although the ABFP did not request and does not currently plan to use this capability, it developed naturally in the course of modeling physician-patient interactions.

Courses of Action have complex internal structures. A Course of Action organizes Steps, which gather, process, and modify information about Populations or Records. For example, a Step might be to obtain a blood pressure from a person. Each Step uses a Reasoning Element to accomplish its tasks. In the case of obtaining a blood pressure, the Reasoning Element would determine and report the simulated patient's systolic and diastolic blood pressure. A group of Steps that can occur in any sequence is called a Batch. For example, when checking both right and left arm blood pressures, the order in which the arms are checked is probably unimportant, so these can be distinct Steps within a Batch. The Course of Action lists a series of Batches that must be executed in sequence, and describes any mandatory delays between Batches. For example, to check orthostatic blood pressures, recumbent pressures would be obtained in one Batch. The patient would sit or stand in a second Batch. After a short mandatory delay, sitting or standing pressures would be obtained in a third Batch.

Courses of Action also describe possible earnings, costs, pleasure, and discomfort that motivate people to seek or avoid activities.

Agents of Change

Agents include physical, chemical, biological, behavioral, and social events capable of influencing Health States or Findings. These Agents can be therapeutic, injurious, or both. Agent descriptions include data about intake, metabolism, and excretion, as applicable. For instance, a long-acting steroid is a chemical agent. Following intramuscular injection, the steroid will have predictable local and systemic concentration Patterns as the chemical dissipates from the injection site. The steroid might be metabolized to other compounds and excreted. Exposure to Agents normally occurs during a Course of Action, as this example illustrates.

The model of Agents describes their recognition, their presence, and the presence of metabolites or byproducts. Other parts of the model, such as the sub-Patterns of Findings, describe the effects of Agents.

Major Relations

Table 1 lists relations shown in Figure 1. The Health States Lead to Health States relation de-

Table 1. Relations Between Entities.

Population Contacts Population
Population Related to Population
Population Interacts with Courses of Action
Population Exposed to Agents of Change
Population Has Health States
Population Exhibits Findings
Agents of Change Cause Health States *
Health States Lead to Health States
Findings Associated with Health States
Findings Link to Findings
Courses of Action Use Agents of Change
Courses of Action Identify Agents of Change
Courses of Action Treat Health States
Courses of Action Alter Findings
Courses of Action Reveal Findings
Courses of Action Evaluate Findings

Note: These relations link entities in the model together. *This relation can be inferred from other information in the model.

scribes how diseases evolve and is therefore critical for simulations. Preventive medicine scenarios might use this relation to generate patients who would benefit from screening. Case management problems can use this relation to model both the past and evolving history of a patient.

The Findings Associated with Health States relation is analogous to similar relations in diagnostic knowledge bases, where sets of Findings essentially define Health States.[3-5] Unlike traditional knowledge bases, this relation links Findings (with their Patterns) to a Health State, rather than linking a range of Finding values to a Health State. Sensitivity and specificity are represented as age-dependent Patterns, rather than constants. The sensitivity of a Finding will be lower and the specificity higher in this model than in traditional knowledge bases.

The Findings Link to Findings relation describes causal associations between Finding Patterns, such as "severe cough causes abdominal muscle pain." This relation contains data about causality, mechanisms, and temporal constraints. This relation should facilitate reasoning about Findings, but will not likely play a role in generating simulations.

The Courses of Action Treat Health States relation illustrates means of curing Health States or preventing their progression. Treatments therefore modify probabilities in a Lead to relation.

Courses of Action have three relations with Findings. The first, Alter, implies changing a Feature Pattern by invoking a sub-Pattern. For example, giving acetaminophen could alter a fever. The second relation, Reveal, links examining Courses of Action to the Findings they produce. For instance, a procedure called "taking a blood pressure" reveals systolic blood pressure. The third relation, Evaluate, links a Finding to a Course of Action that might be used to investigate its cause. This relation would link a Finding of systolic hypertension to a Course of Action describing its work-up.

The Population Contacts Population relation traces transmission of communicable Agents and potentially beliefs. Population Is Related to Population describes biological and social relations and the history of those relations, and traces transmission of genetic Agents. These two relations should allow descriptions of arbitrarily defined families, with arbitrarily harmonious interactions.

The Population Interacts with Courses of Action relation describes why the Population began the Course of Action, what the Courses of Action cost interested parties, and how comfortable the Population was during the Courses of Action. This model allows a patient to remember an unpleasant experience and resist having it repeated. Because Courses of Action can include negative (buying a therapy) or positive (receiving a paycheck) change in wealth, this relation could help model patients' economic inability to follow medical advice.

The Population Exposed to Agents of Change relation describes perceptions about the exposure, knowledge of exposure, and the Course of Action responsible for the exposure. This relation can describe exactly how an Agent was distributed in, metabolized by, and excreted from this Population.

The Population Has Health States relation includes the preceding Health State, a list of Findings attributable to the Health State, and the age at onset, diagnosis, and resolution of the Health State. Health States affect different individuals in different ways, and treatment often depends on the patient's impairments and perceptions. Consequently, a patient's beliefs about disease progression and perceptions of a Health State belong in the Has relation.

The Population Exhibits Findings relation has similar perception attributes. Perceptions can be divided into disutility and concern. Disutility indicates a trade-off a patient would accept to return to normal. Concern indicates a trade-off a patient would accept for full reassurance that a Finding or Health State does not portend future disutility. For instance, a patient with a minor left-sided chest pain might rate its current disutility as $5 ("I would spend $5 to relieve this pain for today."), and the concern as $100 ("I would spend $100 for assurance that nothing serious caused this pain."). If the pain persists unchanged, both of these values might decline as the patient learns to cope with the discomfort and becomes confident that the symptom has no prognostic importance. Thus, patients can have changing attitudes about stable conditions. Patients would typically seek medical care when provoked to do so by a disutility or concern.

Records have the same relations as Populations, except that the details are always more ambiguous, inaccurate, or both. For instance, a patient might have influenza starting December 15, while his Record of himself indicates that he developed influenza between December 10 and December 13. The patient's Record of himself is both ambiguous (there are 4 possible days of onset) and incorrect (none of the days is December 15).

Modifying Relations

The data described in the Lead to, Associated with, and Link to relations often change with medical interventions or other events. Modifiers describe events that cause a permanent variation in the expected history of these relations. For instance, an event might make evolution to another Health State more or less likely (regular low-dose aspirin reduces the risk of acute myocardial infarction), or could permanently alter the likelihood of exhibiting a finding (cardiac transplant prohibits myocardial ischemic pain). The dashed lines in Figure 1 show possible Modifiers. The following examples illustrate some modifiers.

Population Interacts with Courses of Action modifies Health States Lead to Health States. An appendectomy alters the progression of acute appendicitis to appendiceal rupture. Life-span-altering interventions always modify a Lead to relation.

Population Exhibits Findings can modify Health States Lead to Health States. Being overweight increases chances of developing a deep vein thrombosis or pulmonary embolism.

Population Has Health States can modify Health States Lead to Health States. Diabetes accelerates the onset of cardiovascular disease.

Population Has Health States can modify Findings Associated with Health States. Diabetic neuropathies diminish pain associated with myocardial infarction or extremity injuries.

Modifications of these relations account for many benefits ascribed to receiving medical care. Other benefits can occur when medical interventions temporarily decrease the severity of Findings.

Discussion

The model is intended to be a highly structured and realistic representation of family medicine that will guide the design of the family practice knowledge base and eventually support the generation and evaluation of recertification examinations. The complexity of this model derives from our desire to limit the number of unrealistic assumptions accepted as inherent features of the design. All models make some simplifying assumptions about their real life counterparts, however, and it is important to recognize these approximations. In the case of this model, the following are strong assumptions (ie, those that seem most likely to cause trouble): (1) Health States are discrete and distinguishable on the basis of associated Findings, which are also discrete and distinguishable on the basis of the Patterns of their Features. (2) After choosing a percentile curve in a Pattern to represent some value, the percentile does not change. (3) Changes in Patterns (eg, the probability of one Health State evolving to another) can be described for important combinations of risk factors, interventions, and time of occurrence. (4) Transitions from one Pattern to another can be estimated by simple means. (5) Modifying relations do not have important interactions with one another. (6) Highly developed anatomic and physiologic models are not necessary, because associations between Findings provide the same information.

Furthermore, all models are constrained by the accuracy of available data. Many of the data that would be needed to "fill" accurately the family practice knowledge base have never been collected. Therefore, although the design attempts to avoid strong simplifying assumptions about life, data collection could involve numerous conjectures about values. For instance, the family practice knowledge base allows most values to fluctuate with time, but very few data exist regarding changes in the sensitivity and specificity of a test for patients of different ages. These values will most likely be estimated to be constant with age, until more accurate data become available. Experiments will reveal which data assumptions result in unacceptable simulations.

Although the model should have clear places to store nearly all interesting facts about family practice, test generation does not require a comprehensive description of all facts used in family practice. The proposed test should generate plausible problems from a set of data intentionally skewed to generate interesting (ie, discriminating) cases. The ABFP can avoid controversial questions by controlling skewed data. For instance, if the management of borderline diabetes is controversial, the ABFP can edit the family practice knowledge base so that diabetics' fasting blood glucose levels are always markedly elevated. The family practice knowledge base would then be incapable of creating a borderline diabetic. Neither exhaustive nor completely accurate data are required to create a useful examination. Of course, the test requires factual data on all concepts used to critique examinees' inferences. The accessible facts define defensible evaluation questions.

In spite of these challenges for the recertification project, this model of family medicine should provoke interest from developers of other family medicine tools. The model discards many assumptions that limit the applicability of other medical models to teaching, medical records, decision support, structured vocabularies, or reference systems.

Implications for Teaching Family Medicine

The diagram of the model reflects many family medicine concepts and might therefore help students understand the processes at work in family medicine. For instance, the diagram illustrates that Populations have biological and social relations. Populations exist in Health States, which evolve into new, sometimes undesired Health States. A major goal of family medicine is to retard or stop undesirable evolutions and promote desirable evolutions. Stopping one undesirable evolution could, however, result in a different undesirable evolution. In addition, physicians who treat symptoms will Alter Findings, but do not necessarily Treat Health States. Altering Findings usually changes current quality of life, whereas treating Health States usually changes future quality and quantity of life. Because Findings occur in the context of Health States, physicians should contemplate what Health States might be responsible for Findings, rather than Alter the Finding without considering future quality of life. The only tools available for these causes are Courses of Action. Physicians prescribe Courses of Action, but only patients Interact with Courses of Action. The prescription does not guarantee that the patient follows the correct Course of Action. Agents (eg, drugs) make a difference only when used in the context of a Course of Action.

The model's details could provide further insights for students. First, time is an extremely important element of primary care. Patterns become more distinctive as time passes, simplifying diagnosis. The total risk of going from one Health State to another increases with time, increasing the value of early interventions. Second, patients have variable and evolving attitudes about Health States, Findings, and Courses of Action. The goal of medicine might not be to adhere to an endorsed Course of Action, but to optimize each patient's perception of his or her quality of life. To reach this goal, physicians adjust Courses of Action to accommodate individuals' attitudes. Third, the importance of time and attitude in optimizing the quality of a patient's lifetime suggests that continuity of care might help some patients. Many family physicians believe strongly in continuity of care. The scant evidence supporting this belief generally involves chronic and severe disease rather than primary care settings. Trials have shown that continuity improves some aspects of patient satisfaction or reduces costs of managing chronic diseases without dramatic changes in patient outcomes.[13-15] This model might concretely illustrate continuity to students pending more illuminating studies of continuity in primary care.

Implications for Clinical Record Systems

The Record entity, with its information-storing relations to other entities, models a computerized clinical record system. These relations link the Record to specific Health States, Agent exposures, and Courses of Action. They also link the Record to long Patterns of Feature values describing lifelong changes in Findings. Medical record systems constructed to support these data structures would become an important source of data for the ABFP.

Implications for Diagnostic Decision Support Tools

The scope of family practice and the importance of protocols, time, individual variations and attitudes, and rationales will distinguish the content of the family practice knowledge base from that of available diagnostic knowledge bases.[3-5] Although this project does not share the ambitious requirement that it accurately diagnose externally generated cases, some differential diagnosis of internally generated cases might be possible. Algorithms developed for this purpose could have implications for other diagnostic tools.

In this model, differential diagnosis largely depends on establishing the presence of Findings, which in turn depends on establishing the presence of Patterns and sub-Patterns of Features. Except in rare cases of pathognomonic values, confidence in the presence of a Pattern will increase with the passage of time. The simplifying assumption that Pattern percentiles do not change, while useful in the ABFP project, would cause difficulty for diagnostic decision support tools. Diagnostic tools would need to recognize how random variations in a real patient's Patterns affect the likelihood that a particular Finding is present. Furthermore, the ABFP can ignore borderline Patterns by not entering them in the family practice knowledge base, whereas a decision support tool ought to provide some meaningful response to all Patterns.

Use of Structured Medical Vocabularies

The SNOMED International structured vocabulary is a versatile nomenclature for describing medical ideas.[16] This nomenclature allows one to make inferences from the codes used to represent each idea. For instance, the code F-37022 represents "retrosternal chest pain." The first character, "F," indicates that the code is from a broad class of ideas called functions. The next two digits, "37," indicate that the code involves a refinement of the code F-37000, "chest pain, not otherwise specified." Similarly, code F-37020 specifies "precordial chest pain." The code F-37022 implies that retrosternal chest pain is a kind of precordial chest pain, which is a kind of chest pain, which is a kind of function.

Another code, F-37050, represents "exertional chest pain." To describe a patient having exertional retrosternal chest pain appears to require two "cousin" codes, F-37022 and F-37050. The code F-01390 indicates "Exertion, NOS," but has no structural relation to F-37050.

The model suggests a means of using such codes or even an alternative code structure. The F-37000 code fits well with the concept of a broad Finding called "chest pain," and an example of chest pain is a more specific Finding called "precordial chest pain" code F-37020. The idea that the pain increases with exertion would suggest a sub-Pattern triggered by any Course of Action involving exertion, which might be coded, using SNOMED codes, as F-37020(+F-01390). SNOMED does not appear to have a code for a Course of Action called "rest," but the additional concept that rest relieves this pain could be coded as F-37020(+F-01390)(-REST). This example demonstrates some of the power of structured nomenclatures. After encoding "precordial chest pain increased by exertion decreased by rest," it is not necessary to describe how every form of exertion affects the chest pain. As long as the codes for exercise stress test and shoveling snow indicate exertional Courses of Action, both can be predicted to cause exertional chest pain.

The Generalized Architecture for Language Encyclopedias and Nomenclatures in medicine, or GALEN, project is a large effort to develop a similarly flexible medical concept model and attests to the importance of representing medical concepts.[17] The GALEN developers have focused primarily on understanding and representing radiographic reports. The family medicine model provides a complementary approach founded upon medical education and testable concepts, rather than medical records.

Implications for Reference Systems

The structure of and interface to medical reference systems might be enhanced using the model. Current reference systems use the structure of medical publications and lists of abstracted subject headings to facilitate searches through very large databases. These searches can yield large numbers of extraneous citations, especially for novice users.

The model suggests an alternative indexing strategy, as well as a graphical search interface. For instance, one could view a query interface similar to Figure 1. To request a query about the effect of insulin treatment on the development of retinopathy in diabetic patients, one would click on the Health State icon, and select diabetes from an unrestricted list of Health States. The query interface would reappear with diabetes in the Health State icon. Clicking on the Lead to icon would allow the user to select diabetic retinopathy from a list of diseases restricted to diabetic sequelae. One might then click a Modifier button to specify which Course of Action or Agent of Change to consider. The computer would deliver a list of references mentioning insulin in a diabetes Leads to diabetic retinopathy relation. Searching for a particular relation between two entities should improve the efficiency of searches usually performed by naming the entities.

Conclusions

Increasing computer involvement in recertification processes is probably inevitable because of the complexity and volume of data that must be integrated to generate meaningful and useful tests. Computers also facilitate presentation of sound, motion, and adaptively evolving cases. This database model could enhance recertification processes by supporting imitations of lifelike behavior in simulated patients.

Entity-relation diagrams are expressive tools for describing family medicine activities. This model directs attention to somewhat unconventional views of medical data. Particularly interesting data pertain to medical rationale, time, financial and emotional costs, and protocol definitions.

As part of this developmental effort, many concepts in family medicine must be described in a highly structured format. The resulting model of family medicine will facilitate the development of tools for learning and practicing family medicine.

We would like to acknowledge the helpful contributions of Mark Musen, of the Stanford Medical Informatics program.

References

1. Sumner W, Marek VW, Truszczynski M. Designing a knowledge base to support family practice certification examinations. Proc Ann Symp Comput Appl Med Care 1993;909.
2. Interim report on CBT phase II. Philadelphia: National Board of Medical Examiners, February 1992.
3. Miller RA, Masarie FE, Myers JD. Quick medical reference (QMR) for diagnostic assurance. MD Comput 1986;3(5):34-49.
4. Bergeron B. Iliad: a diagnostic consultant and patient simulator. MD Comput 1991;8(1):46-53.
5. Barnett GO, Hoffer EP, Packer MS, Famiglietti KT, Kim RJ, Cimino C, et al. DXplain—demonstration and discussion of a diagnostic decision support system. Proc Annu Symp Comput Appl Med Care 1992:822.
6. Sumner W 2nd. A review of Iliad and Quick Medical Reference for primary care providers. Arch Fam Med 1993;2:87-95.
7. Nease RF Jr, Kneeland T, O'Connor GT, Sumner W, Lumpkins C, Shaw L, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Ischemic Heart Disease Patient Outcomes Research Team. JAMA 1995;273:1185-90.
8. Chen PS. The entity-relationship model: toward a unified view of data. Cambridge Mass: MIT Alfred P. Sloan School of Management, 1976:9-36.
9. Elmasri R, Navathe SB: Fundamentals of database systems. Redwood City, Calif.: Benjamin Cummings, 1989.
10. Ullman JD. Principles of database and knowledge-base systems, volume 1, New York: Computer Science Press, 1988:32-95.
11. Gonella JS, Louis DZ, Gozum ME, editors. Disease staging clinical criteria, 4th ed. Ann Arbor, Mich: MEDSTAT Systems, 1994.
12. Johansson BG. Wigertz OB. An object oriented approach to interpret medical knowledge based on the Arden syntax. Proc Annu Symp Comput Appl Med Care; 1992:52-6.
13. Integrated care for asthma: a clinical, social, and economic evaluation. Grampian Asthma Study of Integrated Care (GRASSIC). BMJ 1994; 308:559-64.
14. Flint C, Poulengeris P, Grant A. The 'Know Your Midwife' scheme—a randomised trial of continuity of care by a team of midwives. Midwifery 1989; 5(1):11-6.
15. Davies AR, Ware JE Jr, Brook RH, Peterson JR, Newhouse JP. Consumer acceptance of prepaid and fee-for-service medical care: results from a randomized controlled trial. Health Serv Res 1986; 21: 429-52.
16. Côté RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, editors. SNOMED International: the systematized nomenclature of human and veterinary medicine. 3rd ed. Northfield, Ill: College of American Pathologists, 1993.
17. Rector AL, Glowinski AJ, Nowlan WA, Rossi-Mori A. Medical-concept models and medical records: an approach based on GALEN and PEN&PAD. J Am Med Informatics Assoc 1995; 2:19-35.