Abstract Articles about medical diagnostic decision support (MDDS) systems often begin with a disclaimer such as, "despite many years of research and millions of dollars of expenditures on medical diagnostic systems, none is in widespread use at the present time." While this statement remains true in the sense that no single diagnostic system is in widespread use, it is misleading with regard to the state of the art of these systems. Diagnostic systems, many simple and some complex, are now ubiquitous, and research on MDSS systems is growing. The nature of MDSS systems has diversified over time. The prospects for adoption of large-scale diagnostic systems are better now than ever before, due to enthusiasm for implementation of the electronic medical record in academic, commercial, and primary care settings. Diagnostic decision support systems have become an established component of medical technology. This paper provides a review and a threaded bibliography for some of the important work on MDSS systems over the years from 1954 to 1993.


Since primeval times, man has attempted to explain natural phenomena using models. In the past four decades, a new kind of modeler, the medical informatician, has developed and proliferated a new kind of model, the medical diagnostic decision support (MDSS) system. Modeling remains an inexact science. Ptolemy, in the Almagest, placed the earth at the center of the universe, and still could explain why the sun would rise in the east each morning. Past and present MDSS systems incorporate inexact models of the incompletely understood and exceptionally complex process of medical diagnosis. Yet mankind, using imperfect models, has built machines that fly, and cured many diseases. Because MDSS systems augment the natural capabilities of human diagnosticians, it is likely they will be employed productively.

By providing a representative bibliography, this paper presents an overview of the state of the art of MDSS systems and the principles that underlie such systems. The bibliography is based on 1) MEDLINE searches, whose results were edited to remove papers whose themes were already represented in the collection, and 2) the author's own collection of reprints gathered over the years. Figure 1 indicates that 1,665 references from 1954-1992 were identified by the author as being relevant to MDSS systems. The references in Figure 1 represent a lower bound on the actual number of peer-reviewed publications relevant to MDSS systems. MEDLINE does not reference many relevant publications in the engineering and computer science literature, and in the past it did not index relevant conference proceedings such as MEDINFO or SCAMC (prior to 1991). Inconsistencies in the author's MEDLINE searching techniques, possible variability in MEDLINE indexing of MDSS articles (especially in the years prior to 1975), and the underrepresentation of non-English-language pub-
lications in both the author's collection and MEDLINE also limited the number of publications in Figure 1. Many important books and monographs were not identified through the process followed by the author. As a result, the totals in Figure 1 for individual years underestimate the true numbers of publications on MDDS systems.

The references cited in the bibliography of this paper were selected from the larger bibliography of Figure 1. The larger bibliography of Figure 1 will be made available electronically by JAMIA.* In the preparation of this article, an attempt was made to select representative projects and ideas, but it is not possible to be comprehensive. In general, whenever a researcher, a project, a laboratory, or a topic relevant to MDDS systems is cited, an attempt has been made to cite subsequent publications from the same source in order to provide a threaded history of relevant work.† Only a small number of systems are discussed in this commentary, and many of the discussions are abbreviated and superficial. Sometimes only a bibliographic trail is provided. Constraints regarding the length of the article also precluded specific mention of many important projects.

Work on Human Diagnostic Problem Solving

The goal of diagnosis is to place a nosologic label on a process that manifests itself in a patient over time. However, diagnosis is a complex procedure more involved than producing a nosologic label for a set of patient descriptors [1980, 1990]. Efficient and ethical diagnostic evaluation requires a broad knowledge of people and of disease states. The nosologic labels used in diagnosis reflect the current level of scientific understanding of pathophysiology and disease, and may change over time without the patient or the patient's illness per se changing [1976, 1972]. For example, changes occur in how a label is applied when the "gold standard" for making a diagnosis shifts from a pathology biopsy result to an abnormal serologic test—patients with earlier, previously unrecognized forms of the illness may be labeled as having it. The labels used by the diagnostic process often represent different levels of description. Some diagnoses label observations or findings, such as congenital clubbing of the fingers; others label intermediate states, such as hypokalemia or metabolic acidosis; others label syndromes, such as nephrotic syndrome or left-sided congestive heart failure; and others label anatomically defined conditions, such as calcific aortic stenosis or focal glomerulosclerosis.

The utility of making specific diagnoses lies in selection of effective therapies, in making accurate prognoses, and in providing detailed explanations. In some situations, it is not necessary to arrive at an exact diagnosis in order to fulfill one or more of these objectives. Treatment is often initiated before an exact diagnosis is made. Furthermore, the utility of making certain diagnoses is debatable. Labeling a patient as having "obesity" does not flatter the patient, and even worse, may cause the physician to do more harm than good.

Newell and Simon, in their classic text on human problem solving [1972, 1974], cite psychological studies that provide insight into how humans make diagnoses. They discuss the manners in which novices and experts solve chess problems, and summarize earlier work by de Groot [1966, 1971] and Jongman...
Experienced chess players are capable of "chunking" a chess game into logical subcomponents, and as a result were far better than novices at memorizing the positions of pieces on the chessboard from actual chess games. The greater the skill of the chess player, the greater was the ability to rapidly and accurately reproduce the positions on the board. By analogy, in familiar settings, human diagnosticians learn to recognize common disease states, by "compiling" their experiences, just as people can recognize their friends or relatives in a crowd. In contrast, experts were no better than novices in memorizing positions of pieces on the chessboard when the pieces were placed randomly.

In medical diagnostic reasoning, there are also cases where recognition from compiled knowledge does not pertain. Some cases present an overwhelming array of seemingly contradictory information; others present with common conditions in unexpected or unusual manners, some patients manifest rare findings or disorders. Unlike expert chess players who are no better than novices in reproducing random board positions from memory, medical experts have different modes of reasoning that can be invoked when simple pattern recognition based on experience fails. Medical diagnosticians in such settings attempt to reason from first principles, using their detailed knowledge of pathophysiologic processes, to construct scenarios under which an illness similar to the patient's might occur. Feinstein, in a series of papers that appeared in the early 1970s, constructed a complex theoretical framework in an attempt to model physicians' pathophysiologic reasoning [1970; 1973].

Experts, because they possess greater stores of compiled knowledge, more seasoned knowledge of pathophysiology, and broader arrays of strategic approaches, reason more efficiently than novices in diagnostic settings [1978; 1988; 1989; 1990]. Patel and her colleagues have conducted detailed analyses of the differences between experts and novices in solving clinical problems [1989; 1990].

Elstein and colleagues showed that experts form diagnostic hypotheses early in case evaluation [1978; 1983; 1987]. Eddy and Clanton noted that hypothesis formation is based on recognition of key or "pivotal" findings [1982]. Kassirer and colleagues, through protocol analysis studying medical diagnostic reasoning, have indicated how diagnosticians refine their initial hypotheses as more information becomes available [1978; 1988; 1989; 1990]. It is clear from such studies that human diagnosticians use prevalence and probabilistic concepts in their reasoning. However, they do not directly remember or utilize mathematically exact probabilities in performing mental arithmetic (i.e. they do not perform Bayesian calculations subconsciously). Studies have documented how poorly humans carry out Bayesian reasoning [1974; 1989].

Physicians are capable of reasoning with incomplete and imprecise information, and often make clinical judgments at times when they have unfulfilled information needs. Covell, Uman, and Manning [1985] demonstrated that physicians were unable to prospectively identify the information sources they used during clinical practice accurately, and that a reasonable number of information needs (clinical questions based on patient care) go unmet in the setting of a busy outpatient group practice. The key question is whether the unmet information needs substantially alter the quality of care delivered; it may be acceptable to give the second-best therapy if the patient is likely to respond anyway. Data generated by Williamson et al. [1989] suggest that in the area of recent innovations, unmet information needs may compromise care.

Osheroff, Forsythe, and colleagues [1991; 1992] used participant observation, a standard anthropological technique, to identify and classify information needs during the practice of medicine in an academic health center. They identified three components of "comprehensive information needs": currently satisfied information needs (information recognized as relevant and already known to the clinician), consciously recognized information needs (information recognized as important to know by the clinician but not known by the clinician), and unrecognized information needs (information that is important for the clinician to know to solve a problem at hand, but not recognized as being important by the clinician). Osheroff and Forsythe noted the difficulty men and machines have in tailoring general medical knowledge to specific clinical cases. There may be a wealth of information in a patient's inpatient and outpatient records, and also a large medical literature describing causes of the patient's problems. The problem is to quickly and efficiently reconcile one body of information with the other. Timpka and his colleagues [1990] analyzed information needs arising in the clinical setting by videotaping doctor-patient encounters and subsequently debriefing the physicians to determine the reasons for their actions recorded by the camera. They verified that a number of information needs go unmet.

Pauker and a number of colleagues have developed
medical decision analysis as a method for overcoming arbitrary clinical reasoning [1975; 1980; 1986]. The method of decision analysis is more often applied to therapy selection than to diagnosis. However, the common dilemma of “test versus treat” (i.e., whether to confirm a suspected diagnosis by ordering additional specific tests or to begin therapy without ordering more tests) is amenable to decision-analytic approaches. The major clinical value of decision analysis is the method itself, which requires decision makers to make explicit all components of a particular clinical problem. Sensitivity analysis allows users of decision analysis to determine which factors involved in a decision are more critical and to identify tolerance levels for those critical values, beyond which the preferred strategies of the decision analysis are altered.

Early MDDS System Research as the Foundation for Present MDDS System Development

The majority of important concepts related to current MDDS systems were developed and presented in the literature prior to 1976. In a comprehensive 1979 review of reasoning strategies employed by early MDDS systems, Shortliffe, Buchanan, and Feigenbaum [1979] identified the following classes of MDDS systems: clinical algorithms; clinical databases that include analytic functions; mathematical pathophysiologic models; pattern-recognition systems; Bayesian statistical systems; decision-analytical systems; and symbolic reasoning or “expert” systems. This section, without being comprehensive, describes how some of the early pioneering efforts led to many classes of systems present today.

Just as computer-based implementation of many complex algorithms involves making tradeoffs between space (memory) and time (CPU cycles), development of real-world diagnostic systems involves a constant balancing of theory (model complexity) and practicality (ability to construct and maintain adequate medical databases or knowledge bases, and ability to create systems that respond to users’ needs in acceptably short time intervals). We may understand, in theory, how to develop systems that take into account gradations of symptoms, the degrees of uncertainty in patients and/or physician—users regarding a finding, the severity of each illness under consideration, the pathophysiologic mechanisms of disease, and/or the time courses of illnesses. However, it is not yet practical to build such broad-based systems for patient care. Early system developers faced such constraints, yet made far-reaching discoveries.

In their classical paper published in 1959, Ledley and Lusted [1959] observed that physicians have an imperfect knowledge of how they solve diagnostic problems. Ledley and Lusted detailed the principles underlying work on Bayesian and decision-analytic diagnostic systems that has been carried out over subsequent decades. They stated that both logic (as embodied in set theory and Boolean algebra) and probabilistic reasoning (as embodied in Bayes’ rules) were essential components of medical reasoning. Ledley and Lusted mentioned the importance of protocol analysis in understanding human diagnostic reasoning. They stated that they had reviewed how physicians solve New England Journal of Medicine CPC (clinico pathologic conference) cases as the foundation for their work on diagnostic computer systems. Both for practical reasons and for philosophical reasons, much work on MDDS systems has focused on the differences between logical deductive systems and probabilistic systems.

Logical systems, based on “discriminating questions” to distinguish among mutually exclusive alternatives, have played an important role since the pioneering work by Bleich and his colleagues [1969] on acid base and electrolytes. To this day, such systems are applicable to narrow domains, especially those where it is fairly certain that only one disorder is present. Ideal application areas are those where detailed knowledge of pathophysiology or extensive epidemiologic data make it possible to identify parameters useful for dividing diagnostic sets into non-intersecting subsets based on specific characteristics. When users of a branching logic system incorrectly answer one of the questions posed by the system, they may find themselves “out on a limb,” with no way to recover except by starting over from the beginning; the likelihood of such problems increases when multiple independent disease processes interact in the patient. The problem with branching logic systems is that the world of medical decision making often contains many shades of gray, rather than simple black and white issues that can be decided through answering critical questions. A hospitalized patient may experience nausea and vomiting for a variety of interrelated reasons, including anxiety, underlying illnesses, and side effects of medications. The search for a single answer in many situations may be futile.

Warner and colleagues in 1960–61 demonstrated that applicability of Bayes’ rule to diagnostic problems was of more than theoretical interest. They developed one of the first medical application systems based on
Bayes’ rule. In their original contribution [1961], they discussed the independence assumption required among diagnoses and among findings by the most commonly employed Bayesian applications, and proposed a method for eliminating the influence of redundant findings. They obtained the probabilities used in the diagnosis of congenital heart diseases from literature review, from their own series of over 1,000 cases, and from experts’ estimates based on knowledge of pathophysiology. Warner et al. observed how diagnostic systems can be very sensitive to false-positive findings and to errors in the system’s database. They emphasized the importance of obtaining accurate data from the user. In their evaluation of their system’s performance, they pointed out the need for an independent “gold standard” against which the performance of the system could be judged. In the evaluation of their system, they used cardiac catheterization data and/or anatomic (postmortem) data to confirm the actual patient diagnoses. Warner et al. have continued to develop and refine models for Bayesian diagnosis over the years [1972, 1987, 1990, 1991].


Utilitarian considerations (ability to implement programs on existing computers) probably influenced the initial development of two dichotomous approaches to computer-based medical diagnostic systems—branching logic and probabilistic systems. Yet two decades after the original contribution by Ledley and Lusted, Szolovits and Paulcoer provided a detailed discussion of deeper philosophical issues related to categorical and probabilistic reasoning [1978]. In physics, it is advantageous to treat light sometimes as a wave and sometimes as a particle. In medical diagnosis, it is sometimes advantageous to reason categorically (causally), and other times to reason probabilistically. The dichotomy between logical (categorical) and probabilistic (actuarial) styles of reasoning continues to be enigmatic for current system developers.

A third alternative to categorical and probabilistic reasoning combines features of both but retains a fundamental difference. That alternative is heuristic reasoning. The HEME program for diagnosis of hematologic disorders was one of the earliest systems to employ heuristics, and also one of the first systems to use, in effect, criterion tables for diagnosis of disease states. It was developed initially by Lipkin, Hardy, Engle, and their colleagues in the late 1950s [1957, 1958, 1961, 1975, 1976, 1987, 1992].

Programs that heuristically match terminology from stored descriptions of disease states to lexical descriptions of patient cases are similar conceptually to HEME. The CONSIDER program developed by Lindberg et al. [1966, 1968], and the RECONSIDER program developed by Blois and his colleagues [1981, 1983] used heuristic lexical matching techniques to identify diseases in CMIT, a manual of diseases compiled and previously maintained by the American Medical Association. More recently, the EXPERT system shell developed by Weiss and Kulkowski [1979] has been used extensively in developing systems that utilize criterion tables, including Al/Rheum [1980, 1982, 1983, 1985, 1986, 1987, 1988].

G. Anthony Gorry was an enlightened pioneer in the development of heuristic diagnostic systems that employ symbolic reasoning. In a classical paper in 1968, Gorry outlined the general principles underlying expert-system approaches to medical diagnosis that were subsequently developed in the 1970s and 1980s. Gorry proposed a formal definition of the diagnostic problem. In a visionary manner, he analyzed the relationships among a generic inference function (used to generate diagnoses from observed findings), a generic test-selection function that dynamically selects the best test to order (in terms of cost and information content), and a pattern-sorting function that is capable of determining whether competing diagnoses are members of the same “problem area” (i.e., whether diagnostic hypotheses should be considered together because they are related to pathology in the same organ system). He pointed out the difference between the information value, the economic cost, and the morbidity or mortality risk of performing tests; discussed the cost of misdiagnosis of serious, life-threatening, or disabling disorders; noted the po-
tential influence of "red herring" findings on diagnostic systems; described the "multiple-diagnosis" problem faced by systems when patients have more than one disease; and suggested that the knowledge bases underlying diagnostic systems could be used to generate simulated cases to test the diagnostic systems.

Gorry's schemata represent the intellectual ancestors of a diverse group of medical diagnostic systems, including the PIP (Present Illness Program) developed by Pauker et al. [197667]; MEDITEL, developed by Waxman et al. [199051]; INTERNIST-I, developed by Poppe, Myers, and Miller [197543; 198225,126; 198545]; QMR, developed by Miller, Masarie, and Myers [1985142; 198676,150; 1989100,102,103; 1983128; 1986163]; MYCIN used backward chaining through its rule base to collect information to identify the organism(s) causing bacteremia or meningitis in patients. A large number of rule-based MDDS systems have been developed over the years, but most rule-based MDDS systems have been devoted to narrow application areas, due to the extreme complexity of maintaining rule-based systems with more than a few thousand rules. An example of a recently developed rule-based system in a focused domain is TRAUMAID, a system for diagnosis and treatment of penetrating injuries to the chest or abdomen developed by Clarke and Webber [19868]. A more general rule-based diagnostic system is the SEEK-I system (and its successor, SEEK-2), developed by Politakis and Weiss at Rutgers University [1984138]. Many of the data-driven warning and reminder systems incorporated into medical record systems use, in effect, rules to "diagnose" conditions that trigger the reminders. Examples include the Regenstrief Clinic System (CARE) developed by McDonald et al. [197668] and the HELP system developed by Warner, Pryor, Gardner and colleagues [1991294].

To complete discussion of early MDDS systems, important work by Reichertz [19659; 196711; 196819; 196923; 197250,51; 197352; 197854; 1980112] and Mohr [197253; 197354; 197855] and their colleagues in Germany; by Greym and Salamon [196713; 197556; 197657,58; 1980111] and their colleagues in France; and by Furukawa [197541; 197642; 197743; 197844; 197945,46; 1982140,147; 1990288] and Kathara [197448; 197551; 197997; 1981117] and their colleagues in Japan, among others, should be mentioned.

**Trends in Current and Future Work on MDDS Systems**

The conceptual basis for MDDS system construction developed during the 1950s and 1960s, leading to exploratory and innovative implementations in the 1970s. Evolutions in MDDS systems during the 1980s and 1990s have been motivated by changes in hardware platforms and user interfaces; by philosophical changes in the perceived role of MDDS systems; by new models for diagnostic decision support; and through expanded understanding of how to evaluate MDDS systems.

**Changes in Hardware Platforms and User Interfaces**

One of the most important developments during the 1980s was the invention, evolution, and ubiquitous proliferation of the microcomputer. The microcomputer made it possible for system developers to distribute MDDS systems in a cost-effective manner to a large user community. Desktop microcomputers in the office and home replaced limited and cumbersome access to mainframe computers via dial-up modems or via direct lines. The microcomputer platform also encouraged development of new, sophisticated graphic user interfaces for MDDS systems. Better evaluations from a broad-based audience allowed developers to evolve systems in response to users' feedback. Availability of standard microcomputer platforms allowed software developers to target common machines and computing environments, rather than developing exotic hardware and software prototypes on one-of-a-kind development platforms that were less likely to appeal to health care workers in the field.

The 1980s also heralded a new era of connectivity via local and national networks. Models for shared and distributed processing developed. A vision
emerged for interaction among modular components of an overall health care information system [199021], which might include various forms of MDDS systems. Modular design of software packages that share a common graphic interface has been encouraged. The Apple Macintosh™ interface, Microsoft Windows™, and Unix-based X-windows are commonly available environments that promote such developments. Medical end-users can collect a series of programs that run on a single machine—helping users to overcome the inertia of not wanting to buy a machine to run only a single software program, however useful the single package.

In order to facilitate data exchange among local and remote programs, it is mandatory to have a lexicon or interlingua that facilitates accurate and reliable transfer of information among systems that have different internal vocabularies (data dictionaries). The United States National Library of Medicine Unified Medical Language System project, which started in 1987 and continues through the present time, represents one such effort [199227].

Philosophical Changes in the Perceived Role for MDDS Systems

With the advent of the microcomputer came a change in philosophy regarding the development of MDDS systems. By the later 1980s and early 1990s, developers abandoned the “Greek oracle” model of diagnostic decision support [1990242]. For example, the style of diagnostic consultation in the original 1974 INTERNIST-I program treated the physician as unable to solve a diagnostic problem. The model assumed that the physician would transfer all historical information, physical examination findings, and laboratory data to the INTERNIST-I expert diagnostic consultant program. The physician’s subsequent role was that of a passive observer, answering yes or no to questions generated by INTERNIST-I. Ultimately, the omniscient Greek oracle (consultant program) was supposed to provide the correct diagnoses and explain its reasoning.

There were fatal flaws in the Greek oracle model. A physician cannot convey his or her complete understanding of an involved patient case to a computer program. One can never assume that a computer program “knows” all that needs to be known about the patient case, no matter how much time and effort goes into data input into the computer system. As a result, the physician-user who has directly evaluated the patient must be considered the definitive source of information about the patient during the entire course of any computer-based consultation. In addition, the highly skilled health care practitioner who understands the patient as a person possesses the most important intellect to be employed during a consultation. That user should intellectually control the process of computer-based consultation in the manner that a pilot controls a complex aircraft in going from point A to point B.

Encouraged by the critiquing model developed by Perry Miller and his colleagues [1984135,136, 1990246, 1987172], recent MDDS system developers have had as an objective to create a mutually beneficial system that takes advantage of the strengths of both the user’s knowledge and the system’s abilities. The goal is to improve the performances of both the user and the machine over their native (unassisted) states. The unit of intervention for evaluation studies becomes more complicated for this reason—it must be viewed as man plus machine, not simply the machine analyzing cases in isolation.

When researchers in medical informatics encounter the term “medical diagnostic decision support systems,” many think primarily of general-purpose, broad-spectrum consultation systems. However, a key distinction must be made in reviewing and analyzing MDDS systems. There exist systems for general diagnosis (no matter how broad or narrow their application domains) and systems for diagnosis in specialized domains, such as interpretation of ECG tracings. The general notion of MDDS systems conveyed in the medical literature sometimes overlooks specialized, focused, yet highly successful, systems. Early work on ECG analysis was carried out by Pipberger [19657, 1990248], Macfarlane [197666, 1990238], Willems [197778, 1990241], and Shiina [197999], among many others. In 1990, an entire issue of Methods of Information in Medicine summarized the state of the art in computerized ECG analysis [223,225,236,238,239,241,246,249,252,253,255]. Willems and colleagues presented the results of a collaborative, blinded evaluation of ten modern ECG analysis programs in 1991 [310]. Commercial ECG interpretation programs are now used ubiquitously. Computerized programs for interpretation of arterial blood gas results and interpretation of pulmonary function tests are now used commonly. Similarly, MDDS systems for cytologic recognition and classification, such as those pioneered by Bartels and colleagues [197026, 197444, 197557, 197670, 1977477, 197880, 1988190, 1990230, 1992318], have found successful application in devices such as automated differential blood count analyzers. Small, focused MDDS systems are one of the most widely used forms of diagnostic decision support programs, and their use will grow as they are coupled with other automated medical devices.

The diversity of MDDS systems continues to increase.
Table 1

<table>
<thead>
<tr>
<th>General MDSS Systems with Bread Application Domains</th>
<th>MDSS Systems with Intermediate-scope Application Domains</th>
<th>Focused MDSS Systems with Narrow Application Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis in general internal medicine</td>
<td>ECG interpretation</td>
<td>Acute MI detection</td>
</tr>
<tr>
<td>Forensic diagnosis</td>
<td>Cardiac arrhythmias</td>
<td>Malignant melanoma</td>
</tr>
<tr>
<td>Diagnosis in veterinary medicine</td>
<td>Analysis of EEG tracings</td>
<td>Diagnosis of hypertension</td>
</tr>
<tr>
<td>Diagnosis in general pathology</td>
<td>Diagnosis of white-matter lesions on MRI</td>
<td>Thyroid disorders</td>
</tr>
<tr>
<td>Radiologic diagnosis</td>
<td>Disorders of lumbar spine</td>
<td>Speech disorders</td>
</tr>
<tr>
<td>Geriatric assessment</td>
<td>Assessment of risk factors for ischemic heart disease</td>
<td>Optokinetic testing</td>
</tr>
<tr>
<td>Psychiatric diagnosis</td>
<td>Detection of cancer</td>
<td>Perimetry testing</td>
</tr>
<tr>
<td>Detection and evaluation of adverse drug effects</td>
<td>Localization of acute neurologic deficits</td>
<td>Sperm motility analysis</td>
</tr>
<tr>
<td>Orthodontic diagnosis</td>
<td>Diagnosis in the ICU</td>
<td>TB via sputum gas chromatography</td>
</tr>
<tr>
<td>Diagnosis of rheumatologic disorders</td>
<td>Diagnosis of acute abdominal pain</td>
<td>HIV risk assessment</td>
</tr>
<tr>
<td></td>
<td>Diagnosis of peripheral nerve lesions</td>
<td>Blood aggregometry</td>
</tr>
<tr>
<td></td>
<td>Analysis of gait disorders</td>
<td>Diagnosis of appendicitis</td>
</tr>
<tr>
<td></td>
<td>Classification of congenital heart disorders</td>
<td>Helminthic infestations</td>
</tr>
<tr>
<td></td>
<td>Differential diagnosis of jaundice and of viral hepatitis</td>
<td>Analysis of breath sounds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cushing's syndrome</td>
</tr>
</tbody>
</table>

*See references 256–316 in the reference list.

The bibliography of this manuscript contains a representative sampling of articles describing MDSS projects from 1991.256-316 Table 1 indicates examples from three general categories of diagnostic systems described in 1991: broad applications (ten examples); systems with intermediate scope (14 examples); and narrowly focused systems (16 examples).

Changes in Approaches to MDSS System Construction—Evolution of New Models

Several innovative techniques have been added in the 1980s and 1990s to previous models for computer-assisted medical diagnosis. The trend has been to develop more formal models that add mathematical rigor to the successful but more arbitrary heuristic explorations of the 1970s and early 1980s. Systems based on fuzzy set theory and Bayesian belief networks were developed to overcome limitations of heuristic and simple Bayesian models. Reggia and Peng [1983,131; 1985,145; 1987,177,178] developed set covering models as a formalization of ad hoc problem-area formation (partitioning) schemes, such as that developed by Pople for INTERNIST-I [1982,140; 1991]. Neural networks represent an entirely new approach to medical diagnosis, although the weights learned by simple one-layer networks may be analogous or identical to Bayesian probabilities.

Fuzzy set theory was developed by Zadeh [1965] and others in the 1960s. It includes formal methods for addressing the incompleteness, inaccuracies, and inconsistencies that are often found in medical data and medical knowledge. Adlassnig [1980,104; 1985,139; 1986,140,149,154] and others have applied fuzzy set theory to diagnosis of medical conditions such as rheumatologic disorders and pancreatic diseases. MDSS systems based on fuzzy set theory embody representation schemes for the degree to which a given patient exhibits a set of findings, and represent confidence or certainty of a given diagnosis on a continuum from 0 to 1. By formally tracking upper and lower bounds on patient parameters; by representing symptom-disease relationships, symptom-symptom (intermediate state) relationships, symptom-symptom relationships, and disease-disease relationships, using, in effect, sensitivity (frequency) and predictive value (strength of confirmation) fuzzy measures; and by using basic operators of conjunction, disjunction, negation, and compositional inference, it is possible to derive bounded certainty values for possible disease states.

Bayesian belief networks, also referred to as probabilistic causal networks or Bayesian networks, represent a mathematical formalism, consistent with the axioms of probability theory, that was developed to overcome the difficulties with data acquisition and reasoning associated with earlier, simple Bayesian approaches (especially the independence assumption). Bayesian belief networks in many ways represent a merger of symbolic reasoning (AI) approaches and Bayesian approaches developed in the 1960s and 1970s [1984,133; 1987]. Belief networks
provide a method for representing probabilistic dependencies and independence. Relationships among observations, intermediate states, and diagnoses can be expressed on a continuum from full independence to full causal dependency. Belief networks consist of a directed, acyclic graph containing nodes whose link strengths are represented by probabilities. The only determinants of the probability distribution of a node are the values of its parents, of its children, and of its children's parents in the graph. Once values for some nodes in a belief network are known, they can be propagated in a forward or backward direction to parents or children.

Belief networks add a formalism that makes dependencies explicit. Belief networks function in a manner consistent with probability theory. Yet, there are several shortcomings associated with their use. Values for link weights can be assigned by an expert (subjective probabilities) or found in the literature (objective ones). When an expert guesses values for link weights, the formal system can be only as good as the expert's guesses. The belief network formalism, just like the decision-analytic model, can at least indicate where it is critical to have precise values. Recent research suggests that it is possible to directly construct (induce) belief networks from observational data [1991270]. A key problem with belief networks is that inference in belief nets with a topology that includes more than one path between any two nodes is NP-hard in the general case (i.e., potentially intractable computationally). Just as early Bayesians had to live with the independence assumption, it is necessary to introduce heuristics and approximations in order to make some large-scale belief networks tractable [1991290,304]. Researchers have tried to address this problem by trying to transform a complex belief network to a singly connected one, or by stochastically estimating the probabilities through simulations and sampling.

Artificial neural networks have been promoted as MDDS systems for focused diagnostic problems by a large number of groups [1981111, 1988173, 1990224,226,232,233,241, 1991277,279,286, 1992223,230]. Development of a neural network for a specific application involves selection of topology (number of input units, number of output units, number of hidden layers, number of units in each layer, connections among units—including feedback loops in some cases), selection of a training rule (the overall feedback mechanism used to adjust weights when network performance for a sample case is suboptimal; in some cases, this may include manual as well as automatic adjustments), selection of training cases or examples, and determination of how far training is to proceed (criteria for determining when a network is "trained"). Problems with neural networks include selecting the best topology, preventing overtraining and undertraining, and determining what cases to use for training. The more complex a neural network is (number of input and output nodes, number of hidden layers), the greater the need for a large number of appropriate training cases. Often, large epidemiologically controlled patient data sets are not available. There is a tendency among some developers to resort to simulation techniques to generate training cases. Use of "artificial" cases to train neural networks may lead to suboptimal performance on real cases.

The late M. Scott Blois, an eminent philosopher—informerican, pointed out in 1980 that computer-based medical consultant systems were most advantageous when applied at the narrow end of a funnel representing how focused medical decision-making problems are [1980107]. A general criticism of all of the newer, more formal models is that in their present form, they are most useful for narrow application areas where data are known with a reasonable degree of certainty. However, when formal mathematical models are applied to general diagnostic problems in broad application domains, they represent, to some extent, reincarnations of the Greek oracle model. The rigidity imposed by a single formalism is often not suited to a flexible and multifaceted analysis of a complex patient problem. The physician—user, who possesses both a detailed knowledge of the patient case and common sense, should be included as an integral part of the decision-making process. Flexible heuristic systems may be better suited to such tasks, although they must be proven to be effective before they are adapted for use. At least for the present, human perception can take into account many more parameters than existing formal models can handle efficiently. In the author's opinion, decision support systems should augment reasoning by physicians and other health care practitioners. Any model that in effect replaces a physician's reasoning, or does not allow the physician to modulate the performance of the system in a patient-specific manner, should be viewed with caution.

**Work on Evaluation and Validation of MDDS Systems**

A critical area relevant to all MDDS systems is that of validation, evaluation, and ongoing quality assurance. The medical informatics and clinical communities have not yet fully determined what a proper evaluation of a diagnostic system should entail, although much past and present work has been de-
voted to this topic. In 1980–81, Hilden and Habhmna presented original work on the theory and practice of system evaluation. Additional effort related to evaluation has been carried out by a large number of investigators. The performance of the system may be exemplary under the circumstances for which it was designed. For example, one of the most thoroughly evaluated and successfully rated MDDS programs is de Dombal’s program for differential diagnosis of acute abdominal pain. That program was developed to support triage of patients in Great Britain who presented to emergency rooms. The major concern was surgical versus nonsurgical therapy. The system was built to handle a very limited number of diagnoses (fewer than 20), and most diagnoses were surgical disorders such as acute appendicitis, acute cholecystitis, and acute diverticulitis; all nonsurgical causes of acute abdominal pain were lumped together as “nonspecific abdominal pain.” The performance of the system may be exemplary under the circumstances for which it was designed. Evaluations of MDDS systems must be considered in clinical context. For example, one of the most thoroughly evaluated and successfully rated MDDS programs is de Dombal’s program for differential diagnosis of acute abdominal pain. That program was developed to support triage of patients in Great Britain who presented to emergency rooms. The major concern was surgical versus nonsurgical therapy. The system was built to handle a very limited number of diagnoses (fewer than 20), and most diagnoses were surgical disorders such as acute appendicitis, acute cholecystitis, and acute diverticulitis; all nonsurgical causes of acute abdominal pain were lumped together as “nonspecific abdominal pain.” The performance of the system may be exemplary under the circumstances for which it was designed. Yet, consider how a male patient, employed as a painter, who comes to an emergency department with severe, colicky periumbilical abdominal pain and foot drop due to lead poisoning might be handled by de Dombal’s system. Ideally, the patient would be correctly assigned to the category “nonspecific abdominal pain” (and not to a surgical diagnosis such as appendicitis). This would be of little consolation to the patient, who is suffering from a potentially life-threatening, yet treatable, condition related to occupational exposure. Other important medical conditions, such as trichinosis or acute intermittent porphyria, might present in a similar fashion and also be labeled “nonspecific abdominal pain.” The possibility exists that a computer system could steer treating physicians away from consideration of further diagnostic and/or therapeutic interventions through reassurance that a patient’s condition was nonsurgical. Thus, the scope of an MDDS system and its medical content are important components in considering the results of an otherwise sound evaluation.

Another important area related to the scope and context of system evaluations is health care practitioners’ attitudes toward computers and diagnostic systems, and introspection into the proper clinical role for computers in medicine. Evaluations of MDDS systems must be considered in clinical context. For example, one of the most thoroughly evaluated and successfully rated MDDS programs is de Dombal’s program for differential diagnosis of acute abdominal pain. That program was developed to support triage of patients in Great Britain who presented to emergency rooms. The major concern was surgical versus nonsurgical therapy. The system was built to handle a very limited number of diagnoses (fewer than 20), and most diagnoses were surgical disorders such as acute appendicitis, acute cholecystitis, and acute diverticulitis; all nonsurgical causes of acute abdominal pain were lumped together as “nonspecific abdominal pain.” The performance of the system may be exemplary under the circumstances for which it was designed. Yet, consider how a male patient, employed as a painter, who comes to an emergency department with severe, colicky periumbilical abdominal pain and foot drop due to lead poisoning might be handled by de Dombal’s system. Ideally, the patient would be correctly assigned to the category “nonspecific abdominal pain” (and not to a surgical diagnosis such as appendicitis). This would be of little consolation to the patient, who is suffering from a potentially life-threatening, yet treatable, condition related to occupational exposure. Other important medical conditions, such as trichinosis or acute intermittent porphyria, might present in a similar fashion and also be labeled “nonspecific abdominal pain.” The possibility exists that a computer system could steer treating physicians away from consideration of further diagnostic and/or therapeutic interventions through reassurance that a patient’s condition was nonsurgical. Thus, the scope of an MDDS system and its medical content are important components in considering the results of an otherwise sound evaluation.

Another important area related to the scope and context of system evaluations is health care practitioners’ attitudes toward computers and diagnostic systems, and introspection into the proper clinical role for computers in medicine. Evaluations of MDDS systems must be considered in clinical context. For example, one of the most thoroughly evaluated and successfully rated MDDS programs is de Dombal’s program for differential diagnosis of acute abdominal pain. That program was developed to support triage of patients in Great Britain who presented to emergency rooms. The major concern was surgical versus nonsurgical therapy. The system was built to handle a very limited number of diagnoses (fewer than 20), and most diagnoses were surgical disorders such as acute appendicitis, acute cholecystitis, and acute diverticulitis; all nonsurgical causes of acute abdominal pain were lumped together as “nonspecific abdominal pain.” The performance of the system may be exemplary under the circumstances for which it was designed. Yet, consider how a male patient, employed as a painter, who comes to an emergency department with severe, colicky periumbilical abdominal pain and foot drop due to lead poisoning might be handled by de Dombal’s system. Ideally, the patient would be correctly assigned to the category “nonspecific abdominal pain” (and not to a surgical diagnosis such as appendicitis). This would be of little consolation to the patient, who is suffering from a potentially life-threatening, yet treatable, condition related to occupational exposure. Other important medical conditions, such as trichinosis or acute intermittent porphyria, might present in a similar fashion and also be labeled “nonspecific abdominal pain.” The possibility exists that a computer system could steer treating physicians away from consideration of further diagnostic and/or therapeutic interventions through reassurance that a patient’s condition was nonsurgical. Thus, the scope of an MDDS system and its medical content are important components in considering the results of an otherwise sound evaluation.

The Future of Diagnostic Decision Support Systems

It is relatively safe to predict that specialized, focused MDDS systems will proliferate, and a sizable number of them will find widespread application. As new medical devices are developed and older devices automated, MDDS software that enhances the performance of the device, or that helps users to interpret the output of the device, will become essential. Computerized ECG analysis, automated arterial blood gas interpretation, automated protein electrophoresis reports, and automated differential blood cell counters are but a few examples of such success at the present time. The future of large-scale, “generic” diagnostic systems is hopeful, although less certain. A number of major challenges remain to be solved before MDDS systems that address large medical problem domains can succeed over time. First and foremost of these challenges is medical knowledge base construction. It is relatively safe to predict that specialized, focused MDDS systems will proliferate, and a sizable number of them will find widespread application. As new medical devices are developed and older devices automated, MDDS software that enhances the performance of the device, or that helps users to interpret the output of the device, will become essential. Computerized ECG analysis, automated arterial blood gas interpretation, automated protein electrophoresis reports, and automated differential blood cell counters are but a few examples of such success at the present time.
tificate over time, yet these activities are required for academic survival.

Knowledge base maintenance is critical to the clinical validity of a MDDS system. Yet, it is hard to judge when new medical knowledge becomes an established “fact.” The first reports of new clinical discoveries in highly regarded medical journals must await confirmation by other groups over time before their content can be added to a medical knowledge base. Knowledge base construction must be a scientifically reproducible process that can be accomplished by qualified individuals at any site [1993][331]. If the process of knowledge base construction is highly dependent on a single individual, or can be carried out only at a single institution, then the survival of that system over time is in jeopardy. While much of the glamour of computer-based diagnostic systems lies in the computer algorithms and interfaces, the long-term value and viability of a system depends on the quality, accuracy, and timeliness of its knowledge base.

The use of lexical matching techniques and other straightforward methodologies to achieve impressive levels of diagnostic performance raises a philosophical issue of interest to all MDDS system developers. Given that no approach to computer-based MDDS is adequate for all situations, how much reasoning power and how detailed a representation of medical knowledge are enough? Is it adequate to perform lexical matching between loosely worded summaries of the important findings in a disease (e.g., CMIT) and the patient’s record (utilizing synonym mapping) simply to produce a list of diagnoses for the intelligent and knowledgeable physician to consider? Or is it necessary to develop extremely detailed and labor-consuming databases and knowledge bases that go far beyond the knowledge of the average clinician in order to provide assistance in the majority of challenging cases? While it is a tautology that common things are common, few, if any, detailed epidemiologic analyses have been carried out to examine the diagnostic dilemmas that generalists and specialists encounter, or of the level of sophistication required to address the majority of such problems. Important research on clinical information needs related to MDDS development and research on potential MDDS system effectiveness will continue.

Another critical issue for the success of large-scale, generic MDDS systems is their environment. Paradoxically, small, limited, “niche” systems will be adopted and used by the focused community for which they are intended, while physicians in general medical practice, for whom the large-scale systems are intended, may not have need for diagnostic assistance on a frequent enough basis to justify purchase of one or more such systems. Therefore, it is common wisdom that MDDS systems are most likely to succeed when they can be integrated into a clinical environment so that patient data capture is already performed by automated laboratory and/or hospital information systems. In such an environment, the physician will not have to manually enter all of a patient’s data in order to obtain a diagnostic consultation. However, it is not straightforward to transfer the information about a patient from a hospital information system to a diagnostic consultation system. If 100 hematocrits were measured during a patient’s admission, which one(s) should be transferred to the consultation system—the mean, the extremes, or the value typical for a given time in a patient’s illness? Should all findings be transferred to the consultation system, or only those findings relevant to the patient’s current illness? These questions must be resolved by careful study before one can expect to obtain patient consultations routinely and automatically within the context of a hospital information system.

A key aspect of a system’s acceptability is its user interface. The graphic user interfaces (GUIs) that are now available facilitate system usage by physicians who have difficulty typing, but many physicians are as uncomfortable using pointing devices as they are typing. Systems must provide flexible environments that adapt to the user’s needs and problems, rather than providing an interface that is inflexible and penalizes the user for deviating from the normal order of system operation. It must be easy to move from one program function to another if it is common for the health care user to do so on his or her own mentally. Transitions must be facilitated when frequent patterns of usage emerge.

Interfaces between automated systems are at times as important as the man–machine interface. Fundamental questions, such as the definition of diseases and of findings, limit our ability to combine data from the literature, from clinical databanks, from hospital information systems, and from individual experts’ experiences in order to create MDDS systems. Similar problems exist when trying to match the records from a given case (collected manually or taken from an electronic medical record) with a computer-based diagnostic system. A diagnostic system may embody definitions for patient descriptors that differ from those of the physician who evaluated the patient, even though the words used may be identical [1991][329].

No matter what the level of use of large-scale, generic
MDDS systems in clinical practice, it is well established that such systems can play a valuable role in medical education [1989,216]. The process of knowledge base construction, utilization of such knowledge bases for medical education in the form of patient case simulators, and the use of MDDS systems have all been shown to be of educational value in a variety of institutional settings.

In summary, the future of MDDS systems appears to be bright. The number of researchers in the field is growing. The diversity of MDDS systems is increasing. The number of commercial enterprises interested in MDDS systems is expanding. Rapid improvements in computer technology continue to be made. A growing demand for cost-effective clinical information management and the desire for better health care are sweeping the United States. All these factors will ensure that new and productive MDDS applications be developed, evaluated, and used.

The author thanks Dr. Constantin Aliferis for his suggestions regarding the discussion of Bayesian belief networks.

References (Sorted Chronologically by Year and Alphabetically by First Author within Each Year)

1954

1957


1959

1961


1965


1966


1967


1968


1969


1970


1971


1972


1973


1974


1975


53. Popple HE, Myers JD, Miller RA. DIALOG: a model of diagnostic logic for internal medicine. In: Proceedings of the Fourth International Joint Conference on Artificial Intelligence. Cambridge, MA: MIT Artificial Intelligence Laboratory, 1975:348-55. [Note: this is the first paper describing the INTERNIST-1 system, which was initially called DIALOG.]


1976


69. Samson M, Salmon R, Bernardet M, Mihout B, Greny F. Terections on the experimental use of automatic data processing as an etiologic diagnostic aid in neurology. Educa-

1977

1978

1980


1981


1982


1983


1984


1985


1986


1987


1999


1990


234. Joseph GM, Patel VL. Domain knowledge and hypothesis
301. Schreiner W, Muller MR, Premauer W, Wolner E. Computerized acquisition and evaluation of whole blood aggre-

302. Schiya T, Watanabe A, Saito M. The use of modified con-
cellation graph method for computer-aided classification of

303. Shiffman RN, Greeneres RA. Use of augmented decision tables
to convert probabilistic data into clinical algorithms for the
1991;686–90.

diagnosis using a reformulation of the INTERNIST-1/QRMR
knowledge base. I. The probabilistic model and inference

305. Solokhin AA, Abdukarimov RKh. The forensic medical di-
gnosis of the types of vehicle-caused trauma by using mathemat-


307. Theodoropoulos O, Lominos V. A compiled computer pro-
gram for assisting the diagnosis of gastrointestinal helminths of

308. Thrope CW, Frighit WR, Toop LJ, Dawson KP. A microcom-
puter-based interactive cough sound analysis system. Com-

309. Tinkleman DG, Lutz C, Conner B. Analysis of breath sounds
in normal and asthmatic children and adults using computer

310. Turner CW, Lincoln MJ, Haug P, et al. lidar training effects:
a cognitive model and empirical findings. Proc Annu Symp

311. Vagnucci AH, Wang TP, Prutt V, Li OC. Classification of plasma
cortisol patterns in normal subjects and in Cushings's syn-

312. van der Knaap MS, Valk J, de Neeling N, Nauta JJ. Pattern
recognition in magnetic resonance imaging of white matter

313. van Ginneken AM, van de Lei J. Understanding differential
diagnostic disagreement in pathology. Proc Annu Symp


315. Westerman ST, Golz A, Gilbert L, Joachims HZ. An objective,
noninvasive method for the diagnosis of temporomandibular

performance of computer programs for the interpretation of

1992

the impact of a computer-based decision support system on

318. Bartels PH, Thompson D, Weber JE. Expert systems in his-
topathology. IV. The management of uncertainty. Anal Quant

319. Engle RJ Jr. Attempts to use computers as diagnostic aids
in medical decision making: a thirty-year experience. Per-


1993


1994

Medical Diagnostic Decision Support Systems—Past, Present, And Future: A Threaded Bibliography and Brief Commentary

Randolph A Miller

*J Am Med Inform Assoc* 1994 1: 8-27
doi: 10.1136/jamia.1994.95236141

Updated information and services can be found at:
http://jamia.bmj.com/content/1/1/8

These include:

**References**
Article cited in:
http://jamia.bmj.com/content/1/1/8#related-urls

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/