In a classic cartoon, a physician offers a second opinion from his computer. The patient looks horrified: How absurd to think that a computer could have better judgment than a human doctor! But computer tools can already provide valuable information to help human doctors make better decisions. And there is good reason to wish such tools were broadly available.

"If you want a second opinion, I'll ask my computer."
About half of the time, doctors fall short of providing quality medical care as defined by national guidelines, according to a 2003 paper in the New England Journal of Medicine. In addition, patients leave their doctors’ visits with an average of 1.6 unanswered questions. “That’s too many,” says Blackford Middleton, MD, assistant professor of medicine at the Harvard Medical School and corporate director of clinical and informatics research and development at Partners Healthcare System in Boston. And because medical professionals have incomplete knowledge or incomplete information about a patient, “we order too many tests, patients are called back, and sometimes bad things happen,” Middleton says. “It’s embarrassing. That’s why I get up every day and run to work.”

Why the hurry? Middleton and his colleagues are trying to build a safer, higher quality, and lower cost health care system right now. And one way to do that, he says, is through well-designed clinical decision support (CDS) systems connected to a nationwide knowledge base of best medical evidence.

Plenty of doctors are dubious about the value of CDS systems. They say they don’t need it; that they are experts in their fields; that they know their patients well, or that their practice is of high quality. “Regrettably, that’s not supported by the evidence,” Middleton says.

Some say it’s also becoming humanly impossible to provide best-evidence medicine without computer support. “There’s too much to know, too much to do; people are overwhelmed,” says John Fox, PhD, a professor in the department of engineering science at the University of Oxford in the United Kingdom.

And the problem is only going to get worse. In addition to staying on top of the 6 million pages of research published in books and journals each year, physicians may soon have to keep track of several hundred thousand genetic variants that could become relevant in medical practice, Middleton says. “It will be impossible for the unaided mind to compute what to do for the patient sitting before you,” he says. Health care providers will need decision support tools.

Today, such tools range from simple alerts, to computerized guidelines that provide recommendations based on electronically stored patient data, to systems that visualize patient data over time or over entire patient populations. Some of these are well-established in various medical institutions around the country; and some are being developed in academia. Others are making the transition to the commercial arena.

So how do we get from where we are now to an efficient national decision support system? It will require incentives for physicians and hospitals to install electronic medical record systems—with both carrots and sticks, Middleton says: carrots to get them to purchase the systems and sticks to make sure care actually improves. It will require a shift in our understanding of what a doctor does—“instead of the final authority, the doctor should be seen as a knowledge manager who helps the patient make the right decision using modern computing tools and decision aids,” Middleton says. It will require a better understanding of what decision support can and can’t do—from the simplest rules to the most complex algorithms—and a way to determine the safety of the system itself. It will require taking the best and most effective academic efforts and bringing them into the commercial arena. And, Middleton says, it will require a national knowledge repository accessible through patients’ electronic records, perhaps as a public web service.

It won’t happen overnight. But it’s enough within reach that Middleton keeps running to work in the morning.
**Building the Plumbing and Standardizing the Data**

Only 15 to 20 percent of physician offices and hospitals in the United States use electronic medical records (EMRs), but that will soon change: the stimulus package passed by Congress in early 2009 is investing nearly $20 billion to incentivize physicians to install EMR systems. “This is (hopefully) a one time frame-shift payment large enough to allow us to wire the country,” Middleton says.

With hundreds of EMRs to choose from, one question is how to ensure that physicians purchase systems that are interoperable. So, as part of wiring the country, the Office of the National Coordinator for Health Information Technology (ONCHIT) is also pushing for data standardization.

“The missing piece is rigorous standardization of the data points that might be used from an electronic medical record for decision support,” says Clem McDonald, MD, a CDS pioneer now at the National Library of Medicine within the National Institutes of Health.

Data can be recorded in ways that confound a computer. For example, to remind a patient to get a mammogram, the computer needs to know when she last had one. But perhaps she had it at a different site, so the system doesn’t know about it. Or perhaps the physician entered the response to a yes/no question (“Have you had a mammogram in the last year?”), in which case the computer still doesn’t know when the reminder should be sent. Or perhaps the EMR contains the patient’s mammograms but can’t easily determine which of many X-ray records actually represents the most recent mammogram.

Or suppose a health care institution wants to contact all of its diabetic patients to inform them of a new treatment option. How does the computer know if a patient is diabetic? Maybe the EMR says so, but perhaps not. So maybe the system looks at whether a patient is on insulin—but there are multiple codes for different types of insulin. Or maybe it looks at lab tests to see if the person is hypoglycemic—but that could be a temporary occurrence.

“Humans will usually make sense of things even if the data is not standardized,” McDonald says. “Whereas the computer has to be able to get at the data.”

“Humans will usually make sense of things even if the data is not standardized,” McDonald says, “whereas the computer has to be able to get at the data.” The problem lies at the interface between the core logic and the facts that feed it—which varies a lot depending on the context and the hospital, McDonald says.

McDonald is hopeful that ONCHIT’s push toward standardization will help. “Everyone invents things in their own way,” McDonald says. “Once that’s solved, all the rest of it gets easier.”

**Smart Alerts—Getting Beyond Simple Rules**

Decision support is not a new idea. Indeed, says Jonathan Teich, MD, PhD, chief medical informatics officer for the health sciences division at the publisher Elsevier, “We’ve already gone through the hype cycle.” Several big studies in the 1990s showed that CDS can prevent medical errors in various circumstances. And, after a 2000 Institute of Medicine report documented the wide extent of medical errors (“To Err is Human”), electronic health record companies created computerized ordering systems and alerts—pop-up windows or alarms that signal the physician should think twice before taking a specified action. “So part of CDS has become mainstream,” Teich says.

“We have really great technology for simple rules,” says Mark Musen, MD, PhD, professor of medicine at Stanford University. “Making a situation-action rule that says ‘if the patient has a penicillin allergy, don’t give penicillin’ is easy.” The problem is getting beyond these simple rules, he says.

Simplistic alerts are among the most annoying types of CDS to implement, Teich says. They can lead to “alert fatigue,” where doctors start ignoring alerts because they receive too many of them.

But some institutions around the country are taking alerts to the next level. For example, Intermountain Health in Utah developed a system that helps doctors determine the right dose of the right antibiotic. “It’s a complicated space,” Musen says. “You have the patient; the bugs in the environmen-
yet the system summarizes that data and proposes candidate drugs. “That’s a more advanced system that’s out and looks very promising,” Musen says.

The Intermountain Health system also functions in the background, monitoring patients continuously, says R. Scott Evans, PhD, senior medical informaticist in the department of medical informatics at Intermountain Healthcare and professor of biomedical informatics at the University of Utah. It evaluates every new piece of data that comes into a patient’s EMR to determine whether a staff member should receive an email message or page notifying them about it. The system also compiles a report of all “reportable” infections and sends it off to Utah’s public health department. And it monitors for adverse drug events, so that if a nurse records a rash or hives, or a lab result indicates a doubling of creatinine, an indicator of kidney function, an alert goes to the appropriate pharmacist suggesting that perhaps the patient should be checked for a drug allergy.

And alert fatigue has not been a real problem, Evans says. “Intermountain is very careful to only create alerts for high priority problems or situations in which the potential harm is large. In those situations, healthcare providers tolerate false positives more than they would otherwise.” Intermountain also makes sure that alert emails include pertinent data in them, so healthcare providers can determine if something is happening or not. For doctors, Evans says, “the worst thing that can happen is ‘I wish I’d known earlier.’”

Initially, some doctors resist the decision support system, Evans says, “until it provides them with some information that prevents harm. Then they become advocates for it.”

Initially, some doctors resist the decision support system, Evans says, “until it provides them with some information that prevents harm. Then they become advocates for it.”

Vanderbilt’s Process-control dashboard shows real-time feedback for ventilator management in the ICU. For each task the staff must perform for each patient, the dashboard shows a green, yellow or red light indicating whether the task was performed on time according to the guidelines. Source: Stead and Starmer, Beyond Expert-based Practice. Pp. 94–105 in Evidence-based Medicine and the Changing Nature of Health Care. 2007 IOM Annual Meeting Summary, Washington, D.C.: National Academies Press (2008). Reproduced with permission from Bill Stead.
Another example of a CDS system that’s gone beyond simple alerts and has helped save lives is one developed at Vanderbilt University. As nurses, therapists, and doctors follow a standardized guideline for ventilator management in the intensive care unit, the status of each patient against the plan shows up on a “dashboard” screen. For each action that has to be taken, the dashboard displays a red, yellow, or green light indicating whether the plan is on track, is in need of attention, or is off track. No actions are lumped together. This provides everyone with a real-time updated measure of how well the team is performing on the guideline. “The thing that makes the difference, is visualizing any gap while the team still has time to take corrective action,” says Bill Stead, MD, associate vice chancellor for strategy/transformation and chief information officer at Vanderbilt University Medical Center.

Identifying the Right Thing to Do: Making Clinical Guidelines Computable

Handling Complexity

Many medical guidelines are too complex for simple rules to handle. That’s particularly obvious in the area of chronic disease, where a condition evolves over a period of time and might eventually involve multiple diseases co-occurring. Such complex guidelines necessitate a more complex clinical decision support system.

To that end, Musen’s group creates abstract computerized representations of clinical care plans that unfold over time. The idea is to create a decision support system that not only suggests to doctors how to follow a standard plan for chronic disease care but also how to deviate from it in appropriate ways.

The backbone of Musen’s work is a task-specific architecture called EON. “This kind of architecture allows you to get above situation-action rules and talk about problem solving in terms of bigger building blocks,” Musen says.

For example, EON can handle tasks such as decomposing an abstract plan into its constituent parts to make it actionable. Take the guideline: “if there’s been a period of uncontrolled hypertension and the patient has been treated with a given therapy, then consider adding a second line agent.” To translate this abstract rule into a concrete action, such as prescribing a particular drug, EON would determine if that”, are difficult to put into code, says Milton Corn, MD, deputy director for research and education at the National Library of Medicine. “The ambiguity that humans handle is very difficult for computers to handle,” he says.

In addition, there is a great deal of inconsistency in the way recommendations are written. A recent study led by Richard Shiffman, PhD, of Yale University examined about 1,275 guideline recommendations derived from the National Guideline Clearinghouse and found that 32 percent of them did not include a reliably identifiable recommended course of action, and more than half did not indicate the strength of the recommendation. As part of the GLIDES project (Guidelines Into

“If we could apply statistical methods, many would regard that as the ideal way to do [decision support],” Fox says. “But we can’t. We don’t know what the numbers are.”
bers are.” So Fox developed a language called PROforma that provides a way of reasoning about what may be the case or what ought to be done. “It’s reminiscent of the way people think,” Fox says. It involves evaluating the arguments for or against a course of action. “And it lets you model any kind of medical decision or clinical process,” Fox says. “It’s very powerful and versatile for such a simple language.”

PROforma lies behind Arezzo, a decision support tool sold by the British company Infermed. Arezzo provides computer interpretable protocols for managing a patient over time. Rather than develop its own knowledge base, Infermed uses guidelines provided by third-party sources such as medical publishers; professional bodies such as the academies of the various specialty services; or the National Institute of Clinical Excellence (Britain’s equivalent of the AHRQ). Typically, a government payer or other significant healthcare organization sponsors the development of an Arezzo guideline. As ambiguities arise, Infermed works closely with experts at the sponsoring organization. “Any questions around these ambiguous or vague recommendations are really exposed quickly when you’re trying to execute them,” says Robert Dunlop, MD, the clinical director at Infermed. “This engages the people who will be using the content so they are more likely to use the system subsequently.”

New Zealand is one of the largest users of the Arezzo system. Twenty-five percent of the country’s family physicians use Arezzo, and last year they accessed the guidelines more than one million times. The guidelines are hosted on a central server and linked through Web services to all five of the EMR systems used by New Zealand’s family physicians. “When the family physician opens a patient record, the system will contact Arezzo through the patient data and recommend the next steps in the patient’s treatment,” Dunlop says. The system is focused mostly on chronic diseases, particularly those that tend to co-occur such as diabetes, hypertension, kidney disease, and ischemic heart disease. The New Zealand system also includes referral triage and management of accidents and injuries.

In the commercial arena, Arezzo is somewhat unusual in providing a system that goes beyond alerts and situation-action rules. “When we talk to customers, we have to explain that the value proposition we bring is very different from these other systems,” Dunlop says. “Once we get past that knowledge barrier, people realize there is nothing else like it in commercial use.”

Dunlop also distinguishes Arezzo from Musen’s work at Stanford. “It’s not like an algorithm where you have a specific pathway you follow,” Dunlop says. Because illness doesn’t follow a specific pathway (from A to B and B to C), the engine has to be able to navigate to whatever part of the guideline content.
is relevant to the patient at a particular time. “What if the patient doesn’t start at A or is suddenly at Q?” he says. “Arezzo fits the guideline to the patient rather than the other way around.”

Like Intermountain Health’s “background” decision support system, Arezzo can also trigger guidelines without the doctor having to realize that he needs them. “We call that the guardian angel approach,” Dunlop says, “where you ensure that if the patient record is updated, with a path to Arezzo behind the scenes, then Arezzo will send an alert that the patient might need to be reviewed according to these guidelines.”

According to Dunlop, New Zealand doctors are finding that Arezzo delivers what they need. “It’s our experience that the physicians are the hardest to convince but they become the greatest advocates once it’s in production,” he says.

**Guessing What The Doctor’s Doing: Using Computerized Guidelines Unobtrusively**

Another option is to let the computer figure out what a doctor is trying to do and then help him or her with that task. Yuval Shahar, MD, PhD, professor of medical informatics at Ben Gurion University in Israel is developing tools that can determine which—if any—guideline a physician is trying to follow. “We can compare the temporal pattern of care over time to see if the physician is actually using any guideline.” Thus the computer might observe that the doctor is trying to apply a particular anti-hypertensive guideline (JNC7) because the physician’s actions fit that guideline the best. The computer might then intervene to say (in a constructive fashion), “if that’s what you’re trying to do, then let me point out that you should now really consider switching medications.” It’s a way of non-obtrusively using artificial intelligence to watch the physician’s actions and try to help them. Shahar has created such a guideline library. “This is still under development,” he says. “It will lead to, I hope, a new kind of medicine in the 21st century.”

**Data Analysis Support: Helping Docs Understand the Patient**

CDS systems are not just about computerized guidelines, though; there is a large amount of data available in patient records that could be harnessed to improve medical decisions. This is particularly true for chronic disease. In the U.S., 80 percent of healthcare system expenses are due to chronic illness—even though chronic illness affects only 25 percent of the patients. Because patients with hypertension, cancer, AIDS, or kidney or heart disease are being treated and monitored for a long time, they generate a lot of data. “We need to help doctors grasp the significance of these data, says Shahar. “Without that, we risk treating

Shahar and colleagues at the VA Hospital in Palo Alto, California, found that doctors could more quickly and accurately answer key questions about cancer patients’ status when data was visualized over time. Here, the KNAVE II knowledge browser shows the bone-marrow transplantation ontology on the left with panels containing raw clinical data and their abstractions on the right. Panels are computed on the fly and displayed when a raw or abstract concept is selected within the left hand browser. Here we see visualizations (top to bottom) of a transplant patient’s bone-marrow toxicity (myelotoxicity) states, platelet-count states and white blood cell (WBC) states over a four-month period. Users can zoom in on specific time periods or select icons to the right. The “KB” icon, for example, defines the concept in that panel. Reprinted from Martins, S.B., et al., Evaluation of an architecture for intelligent query and exploration of time-oriented clinical data, Artificial Intelligence in Medicine 43, 17-34 (2008) with permission from Elsevier.
these patients in a non-optimal fashion or spending too many funds (unnecessarily), or both.”

**INTERPRETING AND VISUALIZING CLINICAL DATA OVER TIME**

Shahar is therefore working to visualize patient data over time in ways that doctors will find helpful. “The key is to apply medical knowledge to these data, thus displaying meaningful concepts emerging in the patient’s data over time,” he says.

In a study published in 2008 in Artificial Intelligence in Medicine, Shahar collaborated with Mary Goldstein, MD, and Suzanne Martins, MD, at the Veteran’s Administration Hospital in Palo Alto, California, to test a tool aimed at helping physicians determine whether a cancer patient’s chemotherapy treatment needs adjusting. They first asked oncologists to identify the questions they’d need answered in order to make such a determination. The questions ranged from simple to complex: Is there anemia? Is there low white blood cell count? Is there a continuous period of liver dysfunction? Is there any pattern of organ toxicity (defined as having 2 out of 3 organs involved)? “It’s not simple at all, at least for humans doing it on their own,” Shahar says. “It requires some really difficult conceptual and cognitive work in putting these values together and drawing a conclusion.” Shahar developed a tool that could visualize the answers to these questions by applying clinical knowledge to raw data, such as hemoglobin values, or raw liver enzyme measurements.

After training for only 10 to 20 minutes, physicians were timed as they answered the questions using three different sources of information (in randomized order): a traditional paper patient record; an Excel spreadsheet containing the patient data; or the data as visualized by Shahar’s tool (KNAVE II), which showed not just the data but also patterns and abstractions of the data. “They were looking at an interface that displayed the patient’s problems through a filter of knowledge,” Shahar says.

Using paper or Excel records, the physicians often took 15 minutes or more to answer all of the more difficult questions, and they did so accurately only 57 percent of the time. By contrast, the KNA VE group answered each of the difficult questions in about 10 seconds—the same time as for the easy questions—and answered with 92 percent accuracy. Since physicians typically see a patient for at most seven to eight minutes, this difference really matters, Shahar says.

“Essentially, humans are probably not very good at noticing temporal trends from clinical or other types of time-step data in a spreadsheet,” Shahar says.

“Essentially, humans are probably not very good at noticing temporal trends from clinical or other types of time-step data in a spreadsheet,” Shahar says.

**ANALYZING PATIENT POPULATION DATA**

In addition to providing a picture of individual patient data, CDS can be used to analyze data on patient populations to provide better care. This often evolves from an institution’s quality assurance and compensation program. For example, Partners Healthcare looks at outcomes for groups of patients as part of the physician compensation scheme, Middleton says. Doctors get a bonus if they are up to snuff for quality measures for certain groups—say diabetes patients or heart disease patients. “The docs love it because they get to see how they are doing compared to peers or national benchmarks,” he says. Once such a quality system is in place, decision support can be applied to patient populations. Doctors can also drill down from the population to the individual level. For example, they could identify patients in the population who are outliers in terms of how they are responding to treatment. “It helps target attention on those people who need it most,” Middleton says.

As McDonald points out: “It’s valuable to see everything at once and act on that collective rather than one off. There might be more leverage in doing that rather than focusing on the person in the office or the ones who happen to call a lot.”

**Nationwide Knowledge Representation**

Many of the cutting edge CDS systems described here are available at only a handful of key institutions such as Intermountain Health, Partners Healthcare in Boston, and the Veterans Administration system. But with the stimulus package funding EMRs all around the nation, the potential for widespread CDS will soon exist. The question is: How to make it happen in the most efficient and effective way?

Middleton says it would cost $25 billion per year for physicians to do the knowledge engineering themselves to put the knowledge needed for CDS into their EMRs. He and his colleagues at Partners Healthcare therefore propose a more cost-effective option: Creation of a national knowledge
repository—a federal facility that delivers knowledge artifacts in a form that every EMR can access and use for decision support.

Using a two-year grant from the AHRQ, Middleton launched the Clinical Decision Support Consortium to build a prototype system. “We’ll aim to stuff it full of knowledge from Partners, from the Riegenstett Institute, from the VA, from Kaiser hopefully, and other members of the CDS consortium,” he says. “And we’ll build web services off of that knowledge repository, so that a remote EMR in Iowa could subscribe to a publicly available Web service and benefit from the knowledge repository.”

The tricky part is in the knowledge representation, which is still an active area of research, Middleton says. The most widely used method is the so-called Arden Syntax, which describes a way to procedurally represent knowledge so it can be used in rule-based systems in EMRs. “It has a host of problems and issues, but is still the best-known example of how to describe knowledge in a way that many people can use and uptake,” he says. “I think as the world moves more toward a service-oriented architecture, it will be easier to represent knowledge in Web services that can be subscribed to in a service catalog.”

Interestingly, Partners’ pilot system is still largely built on simple rules. Middleton says he’s dealing with a different problem than Musen and Fox. “It’s more about accessing knowledge and using it remotely than it is about expressing knowledge in a knowledge formalism,” he says. “I hope in the end that we converge on a general theory of knowledge representation that is both practical and addresses the theoretical limitations of approaches tried to date.”

Teich, at Elsevier, shares Middleton’s urge to make decision support practical to a large segment of the population. In academia, Teich says, you are developing a system in a controlled environment where you can be there every day to make adjustments and tweaks. “When you do it on a larger scale, you have to create things that are more flexible—that can be adjusted and operated by people who may not have full-scale informatics training,” he says.

So although Elsevier is trying to do relatively advanced decision support,
Teich says, “we’re also looking at what drives care needs and quality needs at thousands of hospitals. It’s a different focus than at an academic institution.”

To Teich, the goal is to have a repository of CDS tools that hospitals can download to the local EMR where they’ll start running. You could even have knowledge bases scattered all over the world and tools that integrate those knowledge bases with patient records at any medical center.

“That’s the way life should be,” he says. “We need that kind of drop-in CDS service if we’re going to make this work at 6000 hospitals and many thousands of ambulatory practices.”

The Grand Challenge of Scaling Up

But even if the country is wired with EMRs and has downloadable CDS knowledge representation, there remain some grand challenges for CDS research, according to a January 2009 report by the National Research Council of the National Academies. One of those is to get beyond dealing with patients primarily as a series of transactions, and to provide an overarching view of the individual patient, says Stead, who co-authored the report. “Think what it would be like to examine the world with 1000 ground robot views but no satellite view,” Stead says. “That’s in essence the problem with today’s health care IT.”

Fox agrees: “These systems have to be lovingly—and in some cases, painfully—crafted by hand,” he says. “And while these systems can be very useful, we’re a long way from automation of any systems that have the versatility of a human clinician.” On the other hand, he says, once you have a powerful tool like PROforma or Protégé (a tool developed by Musen), you can apply experience and knowledge to build new applications faster and faster.

“There are research problems in this space,” Stead says, “that are as important as the Human Genome Project a decade ago.”

Funding for the Future of Decision Support

Meeting the grand challenges will require a significant investment. “Unfortunately,” Stead says, “we haven’t had a focused national research agenda in this space.” Some of the problem is institutional. The NIH doesn’t view health care delivery and efficiency as a major part of its mission, Corn says. The NLM has invested in the field because of its interest in how computer science can help in health care delivery and efficiency. “But once we’ve demonstrated that something can be done, is efficient, and might be safer, we’re in no position to make it happen,” he says. “That has to be taken care of outside the Institutes.”

According to Middleton, what’s missing is a single place within the NIH where clinical informatics and its related specialties—including computational science, cognitive science, and information science—are being studied in a coherent and aggressive way. The NLM and AHRQ fund some important work, and the new Office of the National Coordinator for Healthcare IT is a great thing, he says, but there’s no clear focal point for coordinated research activities in inform-