

# **Part 1**

## **Three Grand Challenges for Health Informatics**



# CHAPTER 1

## Introduction

### Diagnosis

Diagnosis seems a good place to start a book about medicine and health care. After all, diagnosis is the first decision that a doctor has to make in the management of a new patient. What exactly do we mean by diagnosis? What is involved in diagnosing an illness? The patient arrives with a story about a problem, a complaint. The doctor first listens to the story, then starts to ask questions. Let us imagine a patient presents at accident and emergency (A&E) with acute abdominal pain and is seen by a junior doctor. As soon as the doctor hears that the patient has acute abdominal pain, he or she will start thinking of the seven or so common (or fairly common) diseases that can cause acute abdominal pain. The doctor might, later on, consider some more unlikely diagnoses as well. He or she will try to establish, through asking a set of questions and performing a simple set of examinations, what the patient's symptoms are.

The trick in diagnosis is to work out, given the symptoms, what the disease is. Or at least what the disease probably is. Or, maybe, what the management should be, given the relative likelihood of a number of possible diagnoses, some more sinister than others. It is, inevitably, a matter of probabilities. As it happens, probability theory gives us a simple equation for dealing with probabilities of this type. It is called Bayes' theorem. In its simplest form, it looks like this:

$$p(D|S) = p(S|D) \times p(D)/p(S)$$

### Bayes' theorem

The notation may look unfamiliar:  $p(D)$  stands for the probability of a disease, which is sometimes called the prevalence, prior probability or pre-test probability of a disease;  $p(S)$  stands for the probability of a symptom. The vertical bar means 'given that'. It expresses the idea that the probability of one thing happening can be altered by the occurrence of another thing. So  $p(S|D)$  is the probability of symptom  $S$  given that the patient has disease  $D$ . It is, therefore, a measure of how good a symptom is as a test for a disease. On the other hand,  $p(D|S)$  is the probability that a patient with symptom  $S$  will turn out to be suffering from disease  $D$ . This, if you think about it, is what the doctor is trying to work out: given these symptoms what is the most likely disease? Bayes' theorem tells him/her how to do it: *the probability that a patient with symptom*

## 4 Chapter 1

*S has disease D is given by the probability of a patient with disease D having symptom S, multiplied by the prior probability of the disease, divided by the prior probability of the symptom.*

Imagine if we actually tried to diagnose using Bayes' theorem. Imagine that a group of people set out to collect data on the thousands of patients who came to their hospital with acute abdominal pain. Imagine that they worked out the prevalence of the various diseases associated with abdominal pain, the prevalence of the relevant symptoms and the probability of each of these symptoms occurring in patients with each disease. Imagine that they programmed a computer to perform the calculations, following Bayes' theorem. Diagnosis would simply be a matter of entering the patient's symptoms into the computer and waiting for the result. Wouldn't that be marvellous? You would get an objective, patient-specific, quantitative, evidence-based statement of the most likely diagnosis. Isn't that the dream that lies behind the subject of this book? Well, it isn't a dream. It was done.

### **AAPHelp**

The first trials of the system now known as AAPHelp (AAP = acute abdominal pain) were published in the 1970s. In 1972, de Dombal *et al.* reported a study in which the system that they created achieved an accuracy of 91.8%<sup>1</sup>. This compared favourably with the accuracy of only 79% achieved by the most senior physician to look at the patients in the study. The junior doctors did much worse. Adams *et al.* reported, in 1986, the results of a multicentre trial involving 16 737 patients<sup>2</sup>. The system raised initial diagnostic accuracy from 45.6% to 65.3%. Observed mortality fell by 22%. In a later European trial the residual diagnostic error rate fell by 40%<sup>3</sup>. The unnecessary operation rate was cut by two-fifths. The perforation rate in appendicitis cases was cut by half. In short, the system proved an astonishing success.

Or did it? If I began to suffer from abdominal pain and staggered out of my office into the A&E department of the hospital where I work, would I benefit from this system? No. Why not? Well, because it is not in routine use in this hospital or, as far as I know, in any hospital. Why not? Well, that is a longer story than the one I have just told and one with important lessons about health care, about diagnosis, about computer systems and about all kinds of things. This book is, in part, an attempt to explain that story.

The impressive results I have quoted above were not the only findings to be published. While de Dombal *et al.* were broadcasting good news in the *British Medical Journal (BMJ)*, another group was printing bad news in the *Lancet*: 'Computer systems based on Bayes's formula have no useful role in the diagnosis of acute abdominal pain'<sup>4</sup>. Others came to the same conclusion. Inevitably there was argument about the methodology of the trials, the interpretation of the results and so on. Many people felt that the system was not given a fair evaluation because clinicians saw it as a threat. Other arguments centred on the usability of the system: remember that this was a

long time ago in terms of user interfaces and processing power and, indeed, in terms of the number of computers readily available in hospitals.

The team behind AAPHelp regarded themselves as pioneers. Inevitably they made a number of pragmatic decisions about which diseases to include, which data items to collect, how to perform the calculations and how to present the results. They were prepared to do the best they could and then to expose the results to empirical tests, to use the system in practice and see if it worked. The clinical evidence about the system's success is, perhaps, mixed. The verdict of history is, however, unequivocal: the system pioneered by de Dombal has not led to the development of a tool used in the management of large numbers of patients.

It is worth thinking about the reasons for the failure of such a promising project. There are many possible objections to the use of AAPHelp. Some of them are quite specific, and have to do with details of the machine's operation and the practicality of its use in a particular setting. Some are more general and would apply to all systems of this type, that is, all systems that attempt to make predictions based on statistical calculations. Other even broader criticisms would apply to almost all attempts to introduce technology into clinical practice. I want to look at some of these criticisms in the rest of this chapter and in so doing to introduce some of the challenges faced by health informatics today.

## **Criticisms of AAPHelp**

### **Technology in medicine**

The most general criticisms reflect concerns about the way technology is used in medicine. Many clinicians are ambivalent about new technology. A doctor who has devoted years of education and training to acquiring and refining a particular skill will inevitably be reluctant to accept a new development that seems to make all that effort redundant. This was true in 1819 when Laennec introduced the stethoscope, and it remains true today<sup>5</sup>. Any hostility towards, or scepticism about, new technology is not necessarily Luddite or reactionary. New technology will generally be accepted if it makes it easier for doctors or nurses to perform the services that they regard as valuable. The difficulty comes when the technology seems either to get in the way of traditional ideas of good practice or to infringe on territory that clinicians regard as requiring expert judgement. Hence, radiologists welcome new and better imaging techniques, because they realise that such developments allow them to become better radiologists. Computer software that could help them interpret X-rays, however, poses a greater challenge to their belief in the value of their own expert knowledge and their existing ways of working.

For over 160 years after the development of reliable thermometers, they were not routinely used to monitor the progress of fevers<sup>6</sup>. The root cause of this long delay was not a reluctance to adopt new technology but rather that the notion of fever was ill defined in the medical thinking of the time. The

## 6 Chapter 1

few studies that were attempted using thermometers failed to show a correlation between temperature and the severity of other symptoms because the researchers had a unitary notion of fever. It was only when researchers developed a classification of distinct fevers that the thermometer became indispensable.

AAPHelp was a particularly problematic system for clinicians. It did not provide the physician with additional information about the patient as a thermometer or a positron emission tomography (PET) scanner does. Most medical technology aims to help the physician by revealing otherwise inaccessible information about the patient's state. The physician's expert judgement is helped by such technology and his or her decisions are better informed. AAPHelp is different. It takes the same information that the physician has, but does something different with it and then confronts him or her with the result. One of the lessons that system designers have had to learn, given the reception of AAPHelp and many similar projects, is that computer systems are most likely to be accepted if they are designed to complement clinical expertise. Decision support systems are now commonplace but the most successful ones are very different from AAPHelp. Computer aids have proved most effective in other decisions; e.g. in prescribing or in generating reminders or alerts<sup>7</sup>. There have been relatively few, if any, successful attempts to apply decision support to diagnostic decisions.

There are other objections to the use of technology in medicine. People are suspicious of it because they feel that it makes medicine cold and impersonal. Clinicians and their patients generally believe that medicine needs a human touch, that patients have to be treated as individuals and that an understanding of the social context and background to a case is often important. The writers of television dramas and hospital-based soap operas clearly believe that their viewers prefer doctors who connect with their patients at an emotional level. A number of health informatics interventions, notably certain attempts to provide telemedicine via videoconferencing, have foundered on the failure to recognise that a medical consultation is not just an occasion for the transfer of patient data and medical advice but is also a social encounter in which the participants have established roles and expectations. Technology that is suspected of dehumanising the consultation is often rejected. But this is not always the case. Patients sometimes express a preference for more technical interventions, perhaps believing that they result in better outcomes (see, e.g. Wallace *et al.*<sup>8</sup>). Such is the penetration of computers elsewhere that many people would be a little surprised if their doctor did not have a computer on his or her desk.

### **Statistical approaches to decision support**

The second class of criticisms concerns the use of what we might call statistical, probabilistic or Bayesian techniques. The controversy about AAPHelp can be seen as part of a wider debate that has its roots in an anxiety about the extent to which medical practice is truly scientific. In the early post-war years,

the accepted view of the role of science in medicine held that the physician was an artisan with a scientific education; a skilled practitioner who understood and applied scientific knowledge but did so using the intuition and experience and skill required to treat unique patients. By the 1970s, however, the editorials of influential clinical journals had begun to argue that there were fundamental problems with this, and to use the term 'scientific' to describe how medicine should be practised. It was argued that medical practice was not the application of a science that is located elsewhere but was, or should be, itself a scientific activity.

Of course, the assertion that medical practice should be more scientific in character can be used to support more contentious proposals. Berg identifies two distinct views of what scientific medicine might be<sup>9</sup>. On one side writers argued for the standardisation of terminology, more rigorous and better structured history taking and the use of flow charts and decision tables to guide diagnostic reasoning. Medicine, on this view, is not an art informed by scientific knowledge but is itself a scientific process in which questions are defined, data collected, recorded, analysed and used to test hypotheses. On the other side were those, like de Dombal, who argued that humans were simply unable to carry out the task of diagnosis with the precision that could be achieved by mathematical tools. The limitations of short-term memory mean that we cannot retrieve and hold in our minds all the necessary facts. We are unable to see all the information that is present in the data, and intuition is hopelessly flawed when it comes to performing probabilistic computations.

Both sides argued for the introduction of new tools and new ways of thinking, but took very different approaches. The kinds of tools that de Dombal and others developed were sharply criticised by opponents who argued that the apparent rationality of statistical methods was deceptive. The messy reality of actual clinical practice meant that countless compromises, pragmatic judgements and unwarranted assumptions had to be made in the design and application of Bayesian systems. Furthermore, the output of such systems – a set of statistical scores – was alien to clinical thinking because the conclusions could not readily be interpreted as an explanation of the salient details in the patients' history.

In the three decades that have followed the development of AAPHelp, two distinct strands of research in decision support can be traced: one is the development of increasingly sophisticated approaches to the use of probabilities in clinical decision making; the other is the attempt to model the logical rules used in making decisions. Many researchers have argued that we should not attempt to build Bayesian systems, in part because in all but a few cases we do not have the required statistical data<sup>10</sup>. Many successful decision support systems have been built using sets (sometimes very small sets) of relatively simple logical rules that can be incorporated into electronic patient record systems or prescribing systems to perform tasks such as checking for allergies or drug interactions<sup>7</sup>. A great deal of the work described in this book

## 8 Chapter 1

aims to provide enhanced patient record systems that will be able to give exactly this kind of support. Much of it draws on work in computer science on the representation of knowledge, and much of that work is, in turn, ultimately based on logic.

Not all work in health informatics is underpinned by logic or probability: e.g. work in telemedicine or on the design of user-friendly websites for the general public. But most of the systems discussed in this book attempt to represent information, either about patients or about medicine. Some of these representations use sets of symbols to represent facts and the relationships between facts. Others depend on numbers, on probabilistic calculation rather than logical inference.

The use of statistical methods to support clinical decision making remains controversial. Clinicians are trained to deal with patients as individuals, whereas probabilistic calculations deal with populations. Most doctors, like most other people, find the mathematics of probability difficult. Practising clinicians have been shown to come to dramatically incorrect conclusions when asked to assess clinical information expressed in terms of mathematical probabilities<sup>11</sup>. But as medical knowledge advances in the post-genomic era we will learn more and more about the genetic basis for disease, and much of what we learn will be about susceptibility and risk. Already we know enough about the risk factors for certain cancers and for cardiovascular disease to mean that the effective communication of information about risk is a key component of preventative medicine. It is not easy to convey an accurate idea of risk: one study has reported that educated American women massively overestimated the incidence of breast cancer, believing that they had a 1:10 chance of dying of it within 10 years when the true likelihood was about 1:200. The development of effective tools for communicating information about risk is a fertile area of research in health informatics.

### **Collecting and analysing patient data**

The final class of criticisms of AAPHelp deals with specific features of the system's operation. There is only one we need to look at here: the use made of patient data. Consider the processes involved in creating and using a system such as AAPHelp. The first step is to collect the data from which the statistics will be calculated. You might think this is easy enough, simply a matter of trawling through the notes and counting up how many times a patient with symptom X turned out to be suffering from disease Y. Well, not quite. Say symptom X is not mentioned in the notes. Does that necessarily mean the patient did not have the symptom? You cannot be sure. The only way to ensure that the statistics accurately reflect the symptoms and diseases of the patients is to collect all the data prospectively. Worse, it is also necessary to set out in advance exactly what questions are to be asked and how the answers are to be recorded. The process of data collection requires the standardisation not just of the set of data items to be recorded for each patient but also the terms used to record patient history. This will inevitably change the way



patients are interviewed and managed. de Dombal described his method thus:

First we created a long list with the items mentioned in the literature. Then we got rid of those items the majority of our clinical colleagues wouldn't do or where they could not agree on the method of elicitation. The reproducibility of the item is important: we have thrown out typifications of the pain as 'boring', 'burning', 'gnawing', 'stabbing'. They haven't gone because people don't use them, they've gone because people can't say what they are . . . . Another example which fell off was back pain with straight leg raising: an often mentioned sign. But nobody agrees on what they are talking about. What should the result of the test be? A figure? The angle the leg makes with the table? . . . We could not get a group of rheumatologists, orthopedic surgeons and general practitioners to agree about what they should call 'straight leg raising' so we abandoned that.<sup>9</sup>

The need for a robust and well-defined set of data items to use in the Bayesian calculations clearly biases the process of history taking. If you cannot agree on how a term should be defined, it cannot go on the form. And if the term is not on the form, it is not in the history, it is not on the record and it is not available to help make a diagnosis. This is one of the most commonly remarked observations on failings of Bayesian systems; critics argue that the 'soft' data items that tend to be dropped are often the most important. Stripping out subjective impressions or observations that have to be understood in terms of a social context deprives the patient history of much of its human character and that obviously worries physicians. Human beings are able to use language to communicate pretty well – most of the time. With computers, things are very different. Although we get by, using words that have no clear, crisp definition, as soon as a computer is introduced into the process things begin to break down.

Of course there is a counter-critique: one could argue that the fact that people cannot agree on the meaning of a particular term raises questions about its value in clinical reasoning. One of the interesting conclusions reached in the work of de Dombal and others was that much of the improvement in performance that followed the introduction of AAPHelp was actually due not to the information that the statistical calculation provided but to the use of a standard data entry form that the computer system required clinicians to use in collecting the history<sup>4</sup>. In order for AAPHelp to generate a prediction, someone had to enter the patient's symptoms into the computer. They had to be collected in a standard format, to match the data stored in the computer. In order to manage the process efficiently, a form was designed that took the doctor through a standard set of questions. Doctors had to sit down with patients and spend between 5 and 20 min going through a checklist of the questions that all doctors know must be asked of such patients but that some of them sometimes forget. Many people believed that at least

## 10 Chapter 1

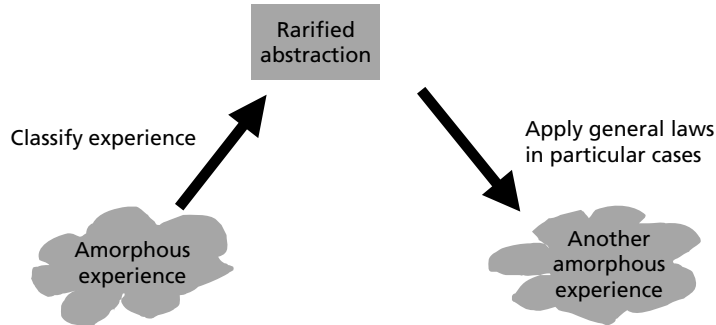
some of the improvement attributed to the software was due to the use of the form rather than the computer-generated predictions. Certainly the team accepted that the standardisation of both terminology and the process of history taking was valuable.

One conclusion that the project team drew from the experience was that ‘databases do not travel’. Part of the reason doctors in different sites had different perceptions of the value of the system was that it performed better in some places than in others. There are, perhaps surprisingly, real differences in the ways clinicians define even the most obvious symptoms and even the best understood diseases. These differences again reflect underlying differences in geography, economics and organisational norms. A system that depends on the capacity of a clinical user to record a history in a standard way will run into difficulties as soon as it is moved into a setting where the users are poorly trained, trained in a different way or simply unfamiliar with the assumptions built into the design of the system. The prior probability that a patient with acute abdominal pain has appendicitis is not the same for a patient who turns up at A&E and another who is referred to the chest ward. Equally, if you install the system in a rural hospital in the north of England, you will get a different mix of patients to those seen in an urban hospital in East London. If the senior clinician in the unit is supportive of the system, it will be used in the management of different kinds of patient than will be the case if the senior clinician is reluctant to get involved.

The predictions generated by AAPHelp would be sensitive to changes, because the data the system uses to calculate the probabilities are specific to the place in which the data were collected. We should be careful about the meanings we attribute to clinical data. They carry information not just about patients but also about the time and place in which they were recorded. They are moulded by all sorts of things, from the internal politics of the institution to the social geography of the surrounding population. Crucially, they are products of the organisational processes through which they were collected.

### **Scientific medicine and the description of experience**

At the heart of the controversy about statistical systems is a question about what use we can make of patient data, other than as an element in the patient’s story. How can we capture what we need to record about a patient’s signs and symptoms in terms that allow us to use them as the raw material of calculations that will inform the care of future patients? The interesting point, if we relate this back to the controversy between the Bayesians and their opponents who advocated a scientific but not a statistical approach to diagnosis, is that the standardisation of terminology and the structured recording of patient histories were first put forward by members of the second camp. And, actually, the difficulties involved in attempting to impose rigid definitions on the terms used to describe clinical conditions crop up all the time in ‘scientific’ medicine. The point is illustrated diagrammatically in Figure 1.1.



**Figure 1.1** Learning from experience involves abstraction.

The goal of most quantitative clinical research is to cast observations about a patient's experience in terms that allow a connection to the experience of other patients. This involves abstraction. It involves extracting something from a messy, complicated, amorphous, individual story that is sufficiently clear and well defined to serve as the raw material of scientific study. It will involve a task not unlike that which confronted the doctors using the AAPHelp system who had to characterise their patients' pain as chronic, acute or cholicky. It will be a matter of putting pegs that are never entirely round or exactly square into holes that are either one thing or the other.

### What have we learnt?

How would we do things differently now, 30 years later? What kind of system might we envisage to support a junior doctor in A&E at the start of the twenty-first century? Perhaps the most obvious difference between a new tool and the one developed by de Dombal *et al.* would be the hardware we would use. A&E departments are complex, flexible and busy environments. We would therefore perhaps want to deliver a system on a hand-held computer connected via a wireless network, something that was certainly not possible for de Dombal. What information might we expect the doctor to obtain from the system? We would be interested in three distinct types of information:

- 1 About the patient – we would want to provide the doctor with the fullest possible access to the patient's record, not just access to notes about previous visits to A&E or previous investigations carried out in the hospital but also his or her general practitioner's (GP's) record, and summarised information about current prescriptions, known allergies and other relevant episodes.
- 2 About the hospital's facilities and procedures – the doctor should be able to consult relevant guidelines, protocols and care pathways to find out about the availability of beds, theatre slots and also be able to order investigations and issue prescriptions electronically.

## 12 Chapter 1

- 3 On clinical evidence and published research – the doctor might consult estimates of the extent to which genetic and environmental factors predisposed patients towards certain illnesses.

### **Evidence-based medicine**

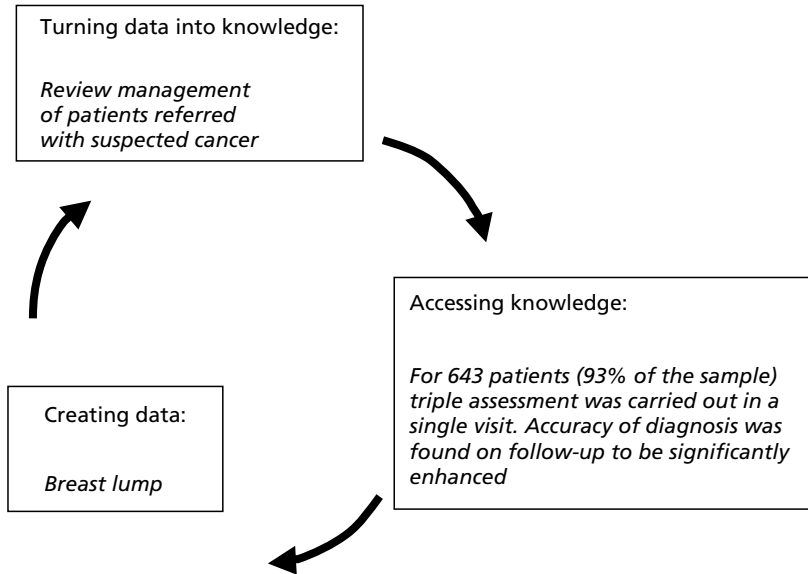
In recent years a movement has grown within medicine, arguing that the pace of change in medical research demands that clinicians should consult the scientific evidence before deciding about the treatment of individual patients. This is simply the most recent expression of the anxiety that sparked off the debate about Bayesian statistics – the belief that too much clinical decision making is arbitrary and idiosyncratic. Its proponents do not think it is enough that the latest advances are taught in medical schools or as part of clinicians' continuing education. If patients are to reap the benefits of new research, they believe clinicians must get into the habit of actively looking for clinical evidence when making decisions about diagnosis and management. This movement is known as 'evidence-based' medicine.

The challenge of evidence-based medicine is to treat each patient as an individual while interpreting his or her unique experience in the light of what has been learned from the experience of others. The project of health informatics – and the subject of this book – is to build tools that maximise the benefits of abstracting from the particular while minimising the costs. Evidence-based medicine is about moving from the abstract to the particular, applying clinical evidence to the amorphous experience of individual patients. Health informatics attempts to support both steps in the process: the creation of evidence out of data, and the application of evidence in the management of patients.

### **Health informatics and evidence-based medicine**

Figure 1.2 is an attempt to illustrate the process by which patient data are transformed into clinical evidence. Three stages are identified. In the first, the data are created. It is worth clarifying the claim that is being made here. Data are not just waiting to be gathered, collected or recorded. Data are created. Recording patient history is not a simple matter of writing down observed facts. The observations emerge from the conversation between the clinician and the patient; they are a product of that conversation and take their meaning from it. Similarly when data are transmitted from one professional to another as the patient moves from primary care to an acute hospital, they alter. Patient histories are continually resummarised, recontextualised and recreated. Even the simplest statements will be reinterpreted in the light of new information, new possibilities and changing priorities.

The process of care comes to a conclusion, if treatment is successful, when the patient stops being a patient and returns to being an active healthy individual. But that is not necessarily the end of the story for the data. The details that have been recorded in the management of this patient are coded



**Figure 1.2** Three stages in a 'virtuous circle' of health knowledge management.

and classified to compile statistics about the management of patients with this disease, at this institution, in this region, and used to answer a range of questions. Clinical audit, clinical research and management scrutiny all depend on data. This is the second stage in the process, the transformation of clinical data into various forms of medical knowledge.

In the third stage, the loop is closed and the knowledge obtained from the data is used to inform the management of future patients. Again, the ideal of evidence-based medicine is that the essence of the aggregated data about past patients provides the empirical basis for decisions about current and future ones.

### This book

The AAPHelp system attempted to do exactly that: to use data about past patients to inform the treatment of current and future patients. It attempted to complete all three arcs of the circle shown in Figure 1.2. This book describes other, more recent systems, techniques and ideas that also aim to realise the potential of IT to improve the flow of information around that circle.

The argument of this book is that the creation of systems to support clinical work has proved harder than de Dombal and other pioneers envisaged. Most medical researchers, in other fields, devote their professional lives to work that promises at best an incremental improvement in how one disease is

## 14 Chapter 1

managed or treated. Researchers in health informatics believed that they could achieve a step-change in the accuracy of diagnosis and efficacy of treatment across a swathe of common conditions. It is the scale of that potential gain rather than the track record of success that continues to motivate work in the field.

The three stages in the graphic correspond to the three 'grand challenges' for health informatics, the three generic tasks involving health information. Chapters 2–4 address each of these in turn.

## References

1. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. *BMJ* 1972;2:9–13.
2. Adams ID, Chan M, Clifford PC. Computer-aided diagnosis of acute abdominal pain: a multicentre study. *BMJ* 1986;293:800–804.
3. de Dombal FT, de Baere H, van Elk PJ, *et al.* Objective medical decision making: acute abdominal pain. In: Beneken EW, Thévenin V, eds. *Advances in Biomedical Engineering*. Amsterdam, The Netherlands: IOS Press, 1993:65–87.
4. Sutton GC. How accurate is computer-aided diagnosis? *Lancet* 1989;2(8668):905–908.
5. Reiser SJ. *Medicine and the Reign of Technology*. Cambridge: Cambridge University Press, 1978.
6. Worth-Eskes J. Quantitative observations of fever and its treatment before the advent of short clinical thermometers. *Med Hist* 1991;35:189–216.
7. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998;280(15):1339–1346.
8. Wallace P, Haines A, Harrison R, *et al.* Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. *Lancet* 2002;359(9322):1961–1968.
9. Berg M. *Rationalizing Medical Work*. Cambridge, MA: MIT Press, 1997.
10. Fox J, Das S. *Safe and Sound*. Cambridge, MA: MIT Press, 2000.
11. Gigerenzer G. *Reckoning with Risk: Learning to Live with Uncertainty*. London: Penguin Books, 2003.