From Patient Data to Medical Knowledge
The Principles and Practice of Health Informatics
To Ailsa and Ewan

In real life a mathematical proposition is never what we want. We make use of mathematical propositions only in making inferences from propositions that do not belong to mathematics to other propositions that likewise do not belong to mathematics.

Wittgenstein *Tractatus Logico-philosophicus*
From Patient Data to Medical Knowledge
The Principles and Practice of Health Informatics

Paul Taylor
Centre for Health Informatics and Multiprofessional Education (CHIME)
University College London, Archway Campus, Highgate Hill
London, N19 5LW
Contents

Acknowledgements, vi
About this book, vii

Part 1: Three Grand Challenges for Health Informatics
Chapter 1  Introduction, 3
Chapter 2  Reading and writing patient records, 15
Chapter 3  Creation of medical knowledge, 32
Chapter 4  Access to medical knowledge, 50

Part 2: The Principles of Health Informatics
Chapter 5  Representation, 69
Chapter 6  Logic, 82
Chapter 7  Clinical terms, 98
Chapter 8  Knowledge representation, 122
Chapter 9  Standards in health informatics, 143
Chapter 10 Probability and decision-making, 158
Chapter 11 Probability and learning from data, 182

Part 3: Achieving Change
Chapter 12 Information technology and organisational transformation, 207
Chapter 13 Achieving change through information, 217
Chapter 14 Achieving change through information technology, 230
Chapter 15 Conclusions, 244

Index, 259
Acknowledgements

I would like to thank my friends and colleagues at CHIME, especially David Ingram and Jeannette Murphy. I also owe a significant debt to John Fox who introduced me to the field. The other alumni of the Advanced Computation Laboratory have been an enormous influence. I have learned a great deal from the students of the UCL Graduate Programme in Health Informatics, among whom Chris Martin stands out as a friend and collaborator. Finally, above all I must thank Jean McNicol. I could not have written this without her support, encouragement and understanding.
About this book

The best way to learn about a subject, I now realise, is to write a book about it. Another good way is to teach it. In 1999, University College London (UCL) started a postgraduate programme in Health Informatics. As the programme director it was largely my responsibility to define the curriculum, a somewhat daunting task in a new and ill-defined subject. I decided, early on, that students should take an introductory module that would give them a grounding in the necessary theory and would also provide a survey of the different problems and applications that make up the field of Health Informatics. The module was called ‘Principles of Health Informatics’. But what are the principles of Health Informatics?

The course, and the introductory module, has now run five times. Our students are all part-time and mostly work in information or clinical roles in the National Health Service (NHS) or other health care organisations (we recruit a small number of international students). They have brought with them a wealth of experience and practical intelligence. Each year I have presented the introductory module in a different way and each year the students have responded to some aspects and not to others. As a result, over the years, my feeling for what the essence of Health Informatics is has changed. Eventually it developed to the point where I felt my understanding of what mattered could be set out in a short book that could serve as a text for our course and for other similar courses.

Writing the book has been complicated by the fact that the UK government is in the process of pushing through an unprecedented programme of investment in information technology, which has raised the profile of the field and also introduced some new and quite specific challenges. I have tried to deal with these, while recognising that specific agenda may well have moved on again by the time this book comes to press. The field is inevitably a rapidly changing one.

The book has three parts. Part 1 consists of an introductory chapter and three further chapters, each of which deals with one of the ‘grand challenges’ I identify for Health Informatics. This part provides a broad introduction to the field of Health Informatics. Part 2 deals with various techniques used in Health Informatics and the theory behind some of them. A key element of this is the question of how we can represent clinical concepts in computer programs such as electronic health care records or decision support systems. I argue that many applications of Health Informatics can be seen as drawing on techniques from computer science that, in turn, are based on logic. I therefore provide a brief introduction to logic and then to subjects that, in some sense, involve the application of logic: controlled clinical terminology,
knowledge representation, ontologies and clinical standards. By way of a contrast I also discuss probability, in two chapters, one of which deals with decision making and the other with statistics, an element in research but also in machine learning and data mining. Part 3 explores attempts to apply Health Informatics in practice. This includes a chapter on theories of organisational change and two further chapters: one dealing with attempts to change clinical practice by improving the dissemination of information and the other on the change management issues raised by attempts to introduce new technology into health care organisations. I also offer some closing thoughts in a final concluding chapter.

I hope that the book will be of interest to anyone who has cause to think about how we use information in health care, and I have tried not to make assumptions of any form of prior knowledge about information, IT, computer science or health care. I live and work in the UK and the overwhelming majority of my students have been employees of the NHS. Many of the examples I discuss are drawn from this experience. I hope, however, that the subject and the themes are nevertheless relevant to a wider audience.
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
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%1. 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 patients2. 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%3. 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’4. 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. 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. 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