CHAPTER 1

Introduction

Good patient care, at all levels, depends upon competently managing and analyzing patient data. Good patient care also depends on understanding the determinants of desirable outcomes and on applying dependable criteria for evaluating care. To identify these determinants and criteria we must explicitly connect the data about our patients and what we do to them with the data about how things turn out. That is what this book is about.

We commonly believe that because we work hard and care deeply about what we do, we must be doing a good job.[1] Rarely do outcomes data support this relationship. Although one may have years of experience in these areas, experience does not reliably enhance competence.[2-5] For experience to inform future action, we must reflect on all that has happened, not just what was most memorable; we must give proper weight to each observation; and we must discern how some observations relate to others.

We may manage and analyze patient data at multiple levels. At the individual level, we work with the data to formulate a patient's problems and plan of care. At the group level, we work with the data to determine how well our system of care serves our patients. Together, the data from these two levels can largely describe our work. However, in actuality, we seldom achieve such comprehensive description of our work and we have few dependable criteria by which to evaluate it.

School-age children ask the riddle: "Why do you always find things in the last place you'd think to look for them?" The answer, of course, is "Because that's when you stop looking." Our daily work is too complicated and demanding to accept having to rummage through a patient chart, with little idea of exactly where to find the information we seek. A central aim of patient information management is to ensure that you get the information you need in the first place you look. But this may require profound advances in how an organization's members collect, manage, and interpret data; how they learn from their experience; and how they share the learning among all who can benefit.

Many of us work in organizations that continue in some degree to rely on handwritten records – a technology essentially unchanged for about 100 years. What has changed is the amount of information and processing activity such records reflect. It is now vast. Of course, we have tools to help, but commonly these are mismatched to the complexity of the problems we tackle.

This book aims at this mismatch of problems with tools. Part I introduces you to getting your arms around data: designing, implementing, and administering an electronic database of patient information. Part II introduces you to asking

questions of data, obtaining well-founded answers, and making sense of those answers.

The subject matter is complex. I have strived to present it simply and clearly, mindful of Albert Einstein's observation: "You do not really understand something unless you can explain it to your grandmother." Even so, I encourage you to read this book unhurriedly and be sure every idea is clear before you move on. Ahead lays new competence to manage the ever-growing amount of information you must handle and new ways to interpret what you see.

When the only tool you have is a hammer, everything looks like a nail
                                              ABRAHAM MASLOW

I think we clinicians often succumb to the vast amount of complex information in our daily work by changing our perception of reality or by not letting it all in. We formulate simplified versions of complicated situations, confident that we have included the essence of the problem (see Ch. 30 in Ref. 6; and Ch. 4, especially p. 87, in Ref. 7). In making a complex medical decision we may need to process in our heads more information than we can handle.[8] Perhaps for this reason, we apply to this vast amount of information decision-making methods that are often tacit (see p. 28 in Ref. 9). Of course, the resulting knowledge is tacit too.[9] Applying such imprecise methods as we do, is it any wonder that our care and documentation vary widely?[10,11] Some claim this variation is integral to the "art of medicine," a notion traditionally thought to embody the wisdom of experience. I think Eddy counters best: "When different physicians are recommending different things for essentially the same patients, it is impossible to claim that they are all doing the right thing."[8]

Here's the point. By our traditional information processing methods, we can't account for all important determinants of the outcomes we observe, nor do we have a sound basis to determine exactly which outcomes we ought to observe. We poorly understand the quantitative dimension of our work. We cannot accurately account for all the relevant evidence. The ever-growing demands of processing information in our heads are overwhelming. We are in trouble.

This information overload was recognized to some extent long ago when paper-based information systems were set up. But a paper-based system is inadequate.[12,13] To recall a lab value or remember to check on an X-ray, we may jot the information down on an index card or on the back of an envelope. Even better, so we don't lose it, we might write it on the leg of our scrub pants. A computer-based information system is potentially far better.

A computer is a wonderful tool to keep track of more things than we can in our heads. But most computers and the programs they run are not designed to handle tacit knowledge and implicit processes. If we are to benefit from the things computers can do enormously better than humans can, we must think *explicitly* about clinical data elements. This entails articulating precise methods for collecting, managing, and analyzing this data – methods quite

different from traditional, paper-based, tacit approaches. Few clinicians are solidly trained in these skills.[14] It's understandable, therefore, that when many clinicians imagine a computer managing their patient information, they imagine the computer doing it the same way the clinicians have always done it, but just faster and with less direct clinician involvement. But to "computerize" patient information management, that is, to represent and conduct our daily work with the aid of a computer, *we must change the way we typically think and work*. To represent our work in a computer we must precisely specify as many essential details about our work as possible. And to conduct our work we must understand the structural relationships among the data elements.

Changing the way we think and work can be enormously disruptive, affecting workflow, and framing what we perceive. For the same reason, it can be positively transformative. Incorporating computer-aided information management processes in our daily work can enable us with capabilities we never before dreamed of. Computer-aided information management forces us to think about the fine structure of our work and our results, can help us make fewer errors (see Framework for Strategic Action, p. 2–3, in Ref. 15), and can help us understand quantitative dimensions of our work (see Ch. 30 in Ref. 6).

To understand quantitative dimensions of our work is to make sense of the enormous amounts of data generated by our work, both at the individual and the group level. It is to know explicitly why we collect each data element that we do, to know what might be different by knowing the value of each data element we collect. It is to progress from relying on what appears intuitively obvious to suspending judgment until results emerge from a carefully analyzed and interpreted body of data. And it is to connect that body of data to other established knowledge. The key tools for making sense of our data are database software and statistical software.

Database software facilitates collecting data, but it only provides data to admire. Statistical software facilitates interpreting aggregate data, but it does not do the interpreting, it only entails computing. Reasoning from the data is something we still must do in our heads. This book explains how to collect data worth admiring and explains powerful methods to analyze that data. The book also explains interpretive reasoning from the analytic results. Because I am a neonatologist, I shall illustrate concepts using data sets from neonatal situations. I assure you, the exposition is nonetheless ecumenical.

Part I of this book is about data management. Section I frames the problems of managing health care data. Section II describes how to plan solutions. Sections III and IV describe how to implement the plan, largely illustrated by the actual patient information management solution that I developed and use daily: eNICU. The eNICU software is on the CD accompanying this book.

What you learn here about database theory, software implementation, and administration should enable you to

- appreciate how these things contribute to managing patient information.
- recognize, understand, and interact productively with the database interfaces you increasingly encounter in your daily work.

4   Chapter 1 Introduction

- use and modify the accompanying eNICU software to support your daily clinical work and evaluation activities.
- recognize the boundaries of your new knowledge and select appropriate further reading material.

If you worry that the software platforms or the hardware that anchor the exposition will become outdated, then don't. They will – as will any other tool you use to manage data. My focus is on illustrating essential concepts that I expect will remain relevant to future software and hardware.

You may question why you need to know about database theory, software implementation, and administration at all, particularly if you plan to invest in a ready-to-use electronic medical record (EMR) product. Well, if that is your plan, then that is precisely why you should learn about these things. You must evaluate the EMR model and implementation – not just its bells and whistles. You must understand why a well-designed database product will nonetheless disappoint you if it is not properly administered. And because others use some of our data, you must understand how to share it.

Part II of this book is about why, and how, you must review collected data to learn accurately from clinical and investigational experience. The patterns in data and the connections among factors contributing to an outcome may be obscure and may surface only after carefully planned analysis. Intuitively "obvious" associations may fade in the light of careful analysis. Part II is also about how to formulate questions that pilot these activities. Some questions, better than others, point the way toward informative analysis and interpretation. Moreover, it is largely by explicitly formulating the questions you seek to answer that you determine the data you really need to collect. And that is the reason for combining in a single book the topics typically covered by two.

What you learn here about data analysis and interpretation should enable you to

- develop genuine analytical competence.
- clearly identify the boundaries of your new skills.
- add coherence and functionality to what you already know – though previous biostatistics coursework is not a prerequisite to master the material presented here.

To develop competence in managing and analyzing vast amounts of data, you must work with database and statistical software. Following are the software applications you'll need in conjunction with your reading:

- Microsoft® Access 2002 or 2003 (Part of Microsoft® Office XP or 2003 Professional).
- Pendragon Forms 4.0 or 5.0: by installing eNICU, provided on the CD at the back of this book, you activate a complementary 2-week trial version of Pendragon Forms. Version 5.0 operates on both Palm OS® and Microsoft® Windows Mobile 2003™ devices.
- Stata[16] (preferably Version 8 or 9; many graphics commands have changed from earlier versions and earlier versions cannot retrieve data directly from a database).

Before we move on to Part I, let's make sure we share a common understanding of key terms.

- *Database*: A collection of data. Surprisingly simple? The term does not necessarily imply computers or software program applications. The card catalog that is now disappearing from libraries is an example of a useful, low-tech database. Computerized, high-tech databases aren't just faster versions of low-tech databases; they usually reflect more sophisticated design principles than the traditional, low-tech databases. Computerized databases can also be connected to each other to achieve search results undreamed of with older technology.
- *Data management*: Meaning depends on context.
    - **(i)** As a clinical process, it refers to the overarching process that begins with planning which data to collect, collecting them, manipulating the data, interpreting them, and reporting both data and interpretation in a specified way.
    - **(ii)** As a database software process, it refers to the software features that control and manipulate the data that reside in the database.
- *Database management system (DBMS)*: A software application that manages data in a database. The DBMS is the means by which you can add, delete, and update data in a database. And it's the means by which you can configure the data in a myriad of ways for viewing or printing.
- *Table*: A container for holding data that share common attributes. Tables have rows (horizontal divisions) and columns (vertical divisions). Each column describes one attribute of whatever the table is intended to describe. Each row contains one instance of the table's attribute set, one observation of the thing the table describes. Each row is also called a record. Each column is also called a field. If you had a table for storing several attributes of your patients, each row would contain the information for one patient (one record), with each column recording the information for each attribute (field). In database jargon, a table is also called a relation.

    Even though you think of a table as a rectangular structure that's neatly subdivided and has data in each of those subdivisions, your computer doesn't. The table appears that way on the screen only because your computer is trying to relate to people. Your computer actually stores the data as magnetic charges distributed not so uniformly on the computer disk. This gives the computer and DBMS software lots of flexibility to manipulate and represent the data in tables.
- *Form*: A way to interface with the data in the database tables without direct risk to the base tables. If you have provided information to a page displayed by an Internet browser, then you have probably worked with a database form. The data you entered in the form went to one or more tables in the website's database.
- *Query*: A question you ask of a database. You ask the question operationally, providing a set of instructions for finding a subset of the data in the database.
- *Report*: Specifically designed printed output of a requested subset of data.

6   Chapter 1 Introduction

- *Relationship*: The logical connection between information in one table (relation) with the information in another table (relation). When one record in table A can relate to only one record in table B, that's a one-to-one relationship. Thus, each patient can have only one set of admission vital signs because a second set would no longer describe the condition at admission. When one record in table A can relate to many records in table B, that's a one-to-many relationship. One mother, for example, can have more than one infant.
- *Primary key*: One or more fields uniquely identifying each record in a table. That is, for each record, the value entered in the primary key field(s) is unique among all records in the table. Without a primary key, records in a database may become confused and the database content degraded.
- *Foreign key*: A good way to link a record in one table with a record in another table is for each record to share some common attribute value. Thus, if we wish to connect a particular record in a table of infant data with a particular record in another table of maternal data, we would ensure the infant data table includes a field containing the mother's unique identifier – primary key value. Such a linking field is called a foreign key. The linked records together describe one instance of a higher entity, the mother/infant dyad in our example, constituted by the various tables in aggregate.
- *Normalization*: A set of design rules specifying what each of the multiple tables in a database is about, and the attributes that belong with each table. These rules generally optimize data storage and retrieval by anticipating the things you'll want to do with the data, and ensuring you'll be able to carry them out. Normalization thus enables reliable queries.

If you want to work with the software straight away

Readers desiring to connect from the outset what they see on the pages of Part 1 with what they see on the PC or PDA screen may look ahead to Chapters 9, 11, 13, 17, and 14 – in that order – and then install eNICU from the accompanying CD. Be mindful, however: simply *owning* a specialized tool does not necessarily make you an expert *user*. To gain a coherent understanding of eNICU and database technology, return here to study the chapters in numerical (and logical) sequence – you will discover anew those to which you looked ahead.