LETTERS

Enhancing Friedman's "Fundamental Theorem of Biomedical Informatics"

In a recent issue of *JAMIA*[1] Charles P Friedman proposed a "Fundamental Theorem of Biomedical Informatics" to aid others in understanding the mission of the profession. The theorem states:

> "A person working in partnership with an information source is 'better' than the same person unassisted."

The theorem was accompanied by three corollaries, briefly:

(1) Informatics is more about people than technology.
(2) The information resource must contain something the person does not already know.
(3) Whether the theorem holds depends on an interaction between person and resource.

The discussion of the theorem was accompanied by a figure showing that a person's brain plus a computer are together greater than the person's brain alone (figure 1).

In this author's view, the figure's focus on the computer provides too narrow a view of bioinformatics and informatics. Missed is the opportunity to emphasize the crucial role of the scientific method in enhancing the learning process.

The computer is an instrument designed to gather, store, and manipulate data: numeric, graphic, handwritten, vocal, visual, or streaming. To handle the vast size and variety of biomedical data a computer is a necessity. However, it is easy to misuse a computer and reverse the "greater than or equal" symbol given in the figure. Further, a computer standing alone is not, as the theorem suggests, an "information source." It is a source of *data*.

What is often overlooked in discussions of Biomedical Informatics is the important difference between information and data. Information rests within data much like a valued ore within rock. Information may indeed be present within a data set but it must be extracted. Like rock without ore, data can contained errors, biases, or be pure noise. Missing in the figure with its display of a computer, and hidden in the theorem's statement "working in partnership," is an appreciation of the scientific method and its role in identifying information in data.

The scientific method begins with an idea, conjecture, or hypothesis. Data are acquired, either through experimentation or from historical sources. Essential statistics and graphics are obtained to illuminate the hypothesis. Models are postulated, fitted, and checked to obtained quantita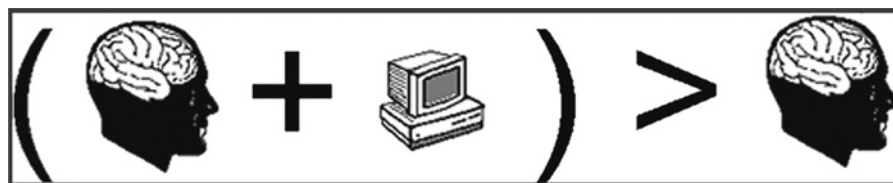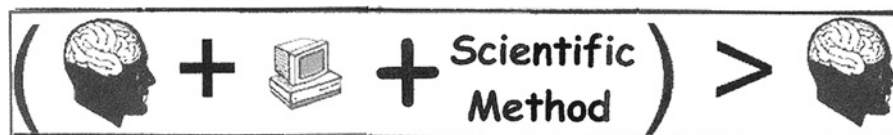tive measures (forecasts) which, when coupled with their standard errors (measures of uncertainty) are compared against prior estimates. Commonly new hypotheses are inferred from these analyses and new data requested (experiments planned), and the iterative process of learning from data, creating new knowledge, begins. On other occasions decisions are made based on costs and the new state of knowledge. The cycle of hypothesis, experiment, data, and analysis leading to a new hypothesis, or to a decision, elucidates the scientific method.

An example of a simple application of the scientific method begins when a physician serves a patient. At the initial greeting preliminary hypotheses are formed. The patient is then examined, and data are requested and analyzed. A more informed hypothesis leads to a decision for treatment of the patient. Will the hypothesis be confirmed? New information now waits on the physician's "experiment," the consequences of the prescribed treatment. These new data lead to newer hypotheses and the iterative process of learning continues. The rational physician evokes the scientific method. This learning process may or may not require a computer.

Finding structure in massive data sets would seem to demonstrate the use of the computer without a need for a formal scientific method approach. Not so. Here the initiating hypothesis is that statistical measures of association (signals) exist within the numerous (noisy) data. Analysis now consists of demonstrating that any discovered data structure is both probabilistically significant and useful as a forecast or prediction. The remaining crucial step is to design experiments to obtain new data to test the discovered structures (the new hypotheses). Newer hypotheses may follow and the learning process continues. Computer scientists at Aberystwyth University[2] recently reported the successful development of a computer that not only generates its own functional genomics hypotheses, but also plans, runs, and analyzes its own experiments and generates new hypotheses. The computer learns and accumulates knowledge, all by itself. Admittedly, humans later confirm the computer's work. But note the scientific method: the cycle of conjecture, experiment, data, analysis, and thence to new conjecture persists.

Learning from data is the objective. Doing it well emphasizes the need for a greater awareness of the role of the scientific method and its associated use of statistical tools. In support of Dr Friedman's "Fundamental Theorem of Biomedical Informatics" and to show the combined roles of the computer and the scientific method, an alternative figure is offered (figure 2).

This new visual hopefully adds to the libretto and enhances the timbre of the informatics message.

Figure 1

Figure 2

J Stuart Hunter

Princeton University, Princeton, New Jersey, USA

Correspondence to Dr J Stuart Hunter, Professor Emeritus, 15 Meadow Lakes 7U, Hightstown, NJ 08520, USA; stu@princeton.edu

REFERENCES

1. **Friedman CP.** A "Fundamental theorem" of biomedical informatics. *J Am Med Inform Assoc* 2009;**16**:169–70.
2. **King RD,** Rowland J, Oliver SG, et al. The automation of science. *Science* 2009;**324**:85–9.

The author's response

I thank Dr Hunter for his insightful comments on my paper. I would like to add some observations in response to his.

I agree that the "fundamental theorem" as expressed in the paper is silent on method and does not explicate how the person-plus-technology actually becomes better than the person unassisted. The scientific method

certainly plays a significant role in this process when the pertinent domain of activity is scientific research.

At the same time, the fundamental theorem applies to many domains of work. These domains include research, of course. In that case, the "person" portrayed in the theorem is a scientist. Other pertinent domains are healthcare (in which case the person is a practitioner or a consumer), education (where the person is a student or a teacher), and administration (where the person is a manager).

While I agree that the scientific method plays a profound role in making the person better in the domain of research, it is less clear that the scientific method applies directly to the domains of healthcare, education, and administration. For example, in healthcare, a significant body of literature suggests that clinicians do not routinely use hypothetico-deductive reasoning, as described by Dr Hunter. As powerful as this approach can be for the discovery of new knowledge, it is generally inefficient for the application of existing knowledge. Evidence suggests that experienced clinicians use a highly efficient inductive pattern-matching process to arrive at most diagnoses, and employ the scientific method only when the patient's problem does not fit a known pattern.

So, I would agree that the scientific method plays a prominent role in making the theorem work but does this primarily in only one of the domains to which the theorem applies. Its role in the other domains is less clear and almost certainly less profound. For that reason, I am much less confident that the scientific method should be included in a general reformulation of the theorem.

Charles P Friedman

Department of Health and Human Services, Washington, DC, USA

Correspondence to C P Friedman, Department of Health and Human Services, 200 Independence Ave SW, Suite 729-D, Washington, DC 20201, USA; charles.friedman@hhs.gov