Editorial Comments JAMIA

Toward a Measured Approach to Medical Informatics

The paper by Hripcsak et al.[1] is a sparkling example of work that is much needed in our field. The preponderance of empirical studies in informatics are *demonstration studies* concerned with the measured values of a set of variables of interest and what these measured values might say about the efficacy of an information resource. In contrast, Hripcsak et al. have undertaken a *measurement study* to determine the precision, or reliability, with which these variables of interest can be measured.

Why do we need these measurement studies? Science is all about measurement, and a scientist who makes measurements without knowing how well these measurements perform is, after a fashion, flying blind. If medical informatics is going to become a mature science, our measurement methods will have to evolve into a type of reusable technology that researchers can take off the shelf and apply to their study needs. That is to say, we in informatics need measurement methods that can be employed in the same routine way an immunologist does a Western blot assay or a psychologist administers a Minnesota Multiphasic Personality Inventory. This reusable measurement technology will have to be thoroughly pretested and calibrated so that researchers employing the technology know how much error is associated with the measurements they make and, thus, the results they report. Measurement studies, like this one by Hripcsak et al., do exactly that: They determine how much error is associated with a certain measurement process so that other researchers do not have to repeat this painstaking work.

In medical informatics we often bemoan the lack of "gold standards" and often claim that systematic studies of our technology are impossible or impractical for lack of such standards. This is, in fact, a naive view of how science is conducted and reflects an unsophisticated approach to measurement. If we view a gold standard as an attribute that can be measured with perfect reliability and validity (precision and accuracy), we quickly realize that no science is blessed with perfect gold standards. A major objective of all scientific fields is to minimize the amount of "tarnish" with which attributes of interest are measured while recognizing that some tarnish is inevitable. Studies can still be conducted as long as we understand how much tarnish there is and deal with it using appropriate methods that are well developed and can be imported from other fields.

The paper by Hripcsak et al. shows us how to do this. It is a measurement study documenting the precision with which a reference standard diagnosis can be assigned, using expert clinicians as judges, in a specific medical domain. This is a problem of unquestioned importance to our collective work. A key finding of the study is how this precision increases with the number of judges employed. This is a typical result of measurement studies. They can tell us not only the precision of a measurement process as the study was actually conducted but also what the precision would be if the number of observers were changed.

This paper also contains, to the best of my knowledge, the first formal application of generalizability theory (also known as G-theory) to medical informatics. G-theory provides a rigorous framework for conducting measurement studies, using the formalisms of experimental design and analysis of variance that are familiar to many researchers while applying them in a novel way. G-theory allows researchers to estimate the amount of error associated with a measurement process as a whole and also to decompose the error into

Correspondence and reprints: Charles P. Friedman, PhD, University of Pittsburgh Medical Center, Suite 8084 Forbes Tower, 200 Lothrop Street, Pittsburgh, PA 15213; e-mail: ⟨cpf@med.pitt.edu⟩.

Received for publication: 11/22/98; accepted for publication: 11/22/98.

multiple sources. G-theory can be rigorously applied to complex research settings, as is well illustrated in this paper.

The results reported by Hripcsak et al. are a start on a long journey toward sophisticated use of measurement techniques in medical informatics, a journey we must collectively take. Their findings cover a small but important fraction of the measurement problems we must eventually address. I strongly encourage other researchers to conduct measurement studies like this one, to apply to their own work the techniques that are so nicely reported here, and to publish the results so we can all learn from them.—CHARLES P. FRIEDMAN

Reference ■

1. Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. J Am Med Inform Assoc. 1999;6:143–150.

■ **JAMIA.** 1999;6:176–177.

■