Editorial
On Opening "Black Boxes" and Looking Inside

LISA I. IEZZONI

Associate Professor of Medicine, Division of General Medicine and Primary Care, Department of Medicine, Harvard Medical School, Beth Israel Hospital, the Charles A. Dana Research Institute and the Harvard-Thorndike Laboratory, Boston, Massachusetts, U.S.A.

Embedded within many of the best-known initiatives to publicly compare provider performance is an ironic twist. Methodologies used to derive comparative performance figures are increasingly proprietary, with their logic protected as trade secrets. Thus, while demands grow to publicize information on provider performance, details of the specific methods used to produce these figures are withheld; they are a "black box".

One area with particularly strong commercial ties involves tools which quantify severity of illness [1–3]. Although severity adjustment appears an arcane methodological pursuit, findings derived from this process, severity-adjusted patient outcome measures, are shaping the dynamics of some highly-competitive health care marketplaces [4,5]. Comparisons of severity-adjusted patient outcomes for doctors and hospitals are scrutinized by state governments, managed care companies, and even employer coalitions. Along with information on hospital charges, severity-adjusted outcome data are used in contract negotiations between hospitals and health care insurers and by hospitals marketing to attract patients. The presumption is that combining cost and outcome information reveals something about provider "value".

However, determining the true value of the information itself is hampered by the secrecy surrounding the specific methods used to produce it. The general rationale for performing severity adjustment seems irrefutable. Some hospitals and doctors obviously treat sicker patients than others. Therefore, when comparing patient outcomes across providers, the major challenge involves teasing apart mortality differences which are due to variations in quality and effectiveness of care from those caused by other factors, such as patient characteristics and random chance. Despite this theoretical underpinning, how best to measure severity is unclear. The various commercial severity products measure severity in different ways. They can also produce different results, for example, by flagging different hospitals as mortality outliers [6–8] or producing very different severity ratings for individual patients [9,10].

The developers of severity measures have often invested years and considerable resources into deriving their products, and their hesitancy about disclosure thus seems natural. Few questions arise when they are sold primarily to individual hospitals which voluntarily seek the methods for internal use and willingly accept these terms. However, public scrutiny of methods seems merited when providers are compelled, either legally or by marketplace pressures, to use a particular product. For example, in 1986, Act 89 of the State Legislature required all Pennsylvania hospitals to adopt MedisGroups® [11,12], a commercial severity measure, as part of a statewide effort to publish information to guide purchasers of health services. The original agreement did not require MedisGroups® to reveal how it worked. Some Pennsylvania hospitals, forced to respond to public reports about poor performance, were hampered by uncertainty about how MedisGroups® rated their patients. Recently, the vendors of MedisGroups® have opened part, but not all, of their system for external scrutiny. Other vendors have been less willing to share...
details of their methods in ways that can be understood.

Queries about how “black box” severity measures work lead inevitably to the real question, “do they work?” Do they provide any insight about quality of care? Reason suggests that, before using a method to evaluate provider quality, the measure should itself prove that it does indeed reflect quality. In the current health policy environment, however, the rules of evidence and proof appear to be reversed. Because it is often the only measure available, severity-adjusted information, such as mortality rates, will continue to be used as an indicator of provider quality until someone proves, definitively, that it is not. Such a definitive study is unlikely in the near future: it will be very expensive and poses the daunting challenge of defining an irrefutable “gold standard” quality measure. Nonetheless, severity-adjusted outcome information could significantly affect health care marketplaces. For instance, in Pennsylvania, insurers are already selectively contracting with certain hospitals based largely upon MedisGroups-derived information [13]. While most developers are, at least rhetorically, vigorously committed to improving health care quality, this verbal commitment is not enough given the tremendous power vested in these measures.

Given this potential for affecting individual patients and institutions, a formal process to evaluate severity measures, as well as other information packages that purport to measure quality, is justified. Opening the “black box” is a minimal first step. We also need to know that the information generated is valid for its intended goal — quantifying quality.

REFERENCES