EDITORIAL COMMENT

Measuring the Effectiveness of Medical Care Delivery*

James G. Jollis, MD, FACC
Durham, North Carolina

There are four stages to improving medical therapeutics: 1) the discovery of new therapies, 2) the application of such treatments to humans, 3) the demonstration of their therapeutic efficacy in clinical trials, and finally, 4) the application of efficacious therapies during routine medical practice. Relative to the other stages, this last important step is supported by relatively few resources, possibly owing to the implicit assumption that physicians and medical systems routinely adopt efficacious therapies without the need for measurement or support. However, in almost every instance in which medical practice is examined, treatment and outcomes vary widely, and therapies and clinical trials identified as beneficial are often surprisingly underutilized.

Lacking adequate resources to measure the effectiveness of medical practice, hospitals, health-care purchasers, government and the media often turn to administrative data. In this issue of the Journal, Tu et al. (1) suggest a measure of medical effectiveness based on Ontario administrative data. Analyzing 52,616 patients hospitalized for acute myocardial infarction between 1994 and 1997, they identify 11 claims descriptors most associated with 30-day and 1-year mortality. The work goes on to externally validate the findings, demonstrating that these same descriptors and their corresponding regression model coefficients are similarly associated with mortality in administrative data from Manitoba and California.

In choosing from among 43 candidate variables for the final Ontario prediction rule, these researchers use reasonable selection criteria, including previously published risk factors, strength of association with mortality and clinical intuition. Two other considerations that may add to the suitability of claims diagnoses in mortality comparisons include the likelihood that a diagnosis represents a coexisting illness rather than a complication, and the potential for “gaming” of the variable. A significant concern regarding the use of claims data in mortality comparisons involves the difficulty in discerning diagnoses present on admission from those that accrue during hospitalization. In the latter case, the inclusion of complications in regression models gives hospitals credit for complications. For example, a patient with a large myocardial infarction who is not treated with reperfusion therapy may go on to develop shock or congestive heart failure, and these diagnoses would increase expected mortality in regression models, effectively “lowering the bar” for the involved hospital and masking less effective care. Claims data from Ontario may potentially avoid this dilemma, because hospitals in this province are required to add an additional level of detail to hospital discharge diagnoses that identifies them as either preexisting or occurring after hospital admission, and only diagnoses identified as present on admission are used for risk adjustment. For systems lacking the level of detail available in Ontario, diagnoses in the suggested prediction rules such as acute renal failure and cardiac dysrhythmias may represent complications for which hospitals “earn credit” as noted earlier.

Regarding gaming, in settings where hospital claims are used to rank health-care providers, public scrutiny may encourage providers to code high risk diagnoses whenever possible. Such a strategy would lead to an apparent improvement in outcomes due to manipulation of hospital discharge abstracts, rather than any real improvement in care. The reporting of diagnoses, such as shock, present among 2.5% of patients in the Ontario data, and approximately 7% of patients in clinical data, should be monitored at the hospital level to differentiate actual improvements in mortality from “coding creep,” or increased specification of mortality model components (2).

These authors consider the ability of their models to account for mortality differences through a variety of measures, including calibration, or a comparison of actual to expected mortality, for 10 subgroups sorted by increasing risk. Most prominently featured is the area under the receiver operating characteristic (ROC) curve or c-statistic. This measure divides patients into two groups, alive or dead, and examines how often the model assigns higher mortality risk to the patient that died for all potential pairs of patients. A value of 1 represents discrimination 100% of the time or perfect discrimination, while a value of 0.5 implies discrimination 50% of the time or similar discrimination to that of a coin toss. The values herein of 0.77 to 0.79 are reasonably good for mortality models. In considering these measures, one should be aware that the inclusion of complications in mortality models improves model performance characteristics such as ROC areas and calibration despite obscuring the identification of most effective care. Thus, deciding if data are suitable for outcomes comparisons should also consider factors beyond these performance measures.

Most remarkable is the finding that the Charlson index (a group of 12 illnesses found to be associated with 1-year mortality among 559 New York University general medi-
Evidence and routine medical practice, what are the next steps? The refinement of claims data to identify effective medical therapy using the Ontario approach remains a practical and reasonable first step. This approach, however, should be considered an interim solution. In order to resolve fully the gulf between evidence from clinical trials and medical practice, information needs are much greater than current claims data can provide. Physicians and other health-care providers need to know if and when effective medical therapies are being employed, whether treatments found to be efficacious in clinical trials are similarly effective when applied to the general population and where resources may be best directed to improve care.

Recent experience during a European Society of Cardiology/American College of Cardiology exchange program suggested that we are at the threshold of a fundamental change in medical care. Despite differences in language and customs across countries, one feature that remained consistent across U.S. and European hospitals was identical computer technology. This new international electronic standard has placed information systems capable of measuring effectiveness within the reach of health systems in most of the developed world. With such innovation, the vast array of data generated during routine practice may actually be used to systematically improve practice. Rather than further refining claims data, our greatest efforts should involve redirecting the traditional medical practice of meticulously recording information in free text and paper-based systems toward the specification of such information in common electronic formats suitable for comparison across hospitals and health systems. This reconfiguring of the medical information process should be supported by the integration of information from other medical care activities, including scheduling, pharmaceutical administration and test results reporting. These efforts must also be supported by the development of sensible standards to protect confidentiality without overly obscuring information such that medical care cannot be evaluated. The featured Ontario work demonstrates that the tools exist to convert regional electronic medical data into information about medical effectiveness. Ultimately, relatively coarse tools such as administrative data comparisons will be replaced by systems that can directly monitor and guide medical effectiveness. Such an enhancement to the final critical step of therapeutic development will ensure that efficacious therapies proliferate beyond clinical trials so that all of society may benefit from medical discovery.

Reprint requests and correspondence: Dr. James G. Jollis, PO Box 17969, Durham, North Carolina 27715. E-mail: james.jollis@duke.edu.

REFERENCES

2. Hochman JS, Boland J, Sleeper LA, and the SHOCK Registry