Using Administrative Codes to Measure Health Care Quality

The US spends more annually per capita on health care than any other country, with inpatient care accounting for approximately 20% to 30% of total expenditures.\textsuperscript{1,2} In an effort to eliminate waste, the US health care system is evolving away from fee-for-service payment models toward value-based reimbursement models. High-stakes applications of health care quality measures, such as attaching financial incentives to performance and public reporting of data regarding quality or cost, are central to the effort to optimize health care spending. Because administrative data (ie, diagnosis and procedural codes) are standardized and can be accessed at a low cost in virtually every practice setting, these codes are commonly used to construct process and outcome quality measures used in value-based payment and quality assessment programs.

Although the use of administrative data for quality measurement is convenient, this approach creates opportunities to “game the system” because hospitals can optimize their coding practices to maximize reimbursement or perceived performance.\textsuperscript{3-6} For outcome quality measures, risk adjustment has a critical role by accounting for underlying variation in the level of risk across patient populations, thereby ensuring that measurement, comparisons, and reimbursement are fair and balanced and have face validity. Because hospitals are reimbursed more for treating more severely ill, higher-risk patients, administrative data can either reflect the true level of risk in a hospital’s patient population or can overestimate—or underestimate—this risk.

Introducing financial incentives into quality assessment is a concern because it encourages upcoding (ie, maximizing a patient’s apparent number of comorbid conditions or the severity of these conditions through administrative coding).\textsuperscript{3-7} Although existing Centers for Medicare and Medicaid Services (CMS) fraud and abuse programs imply there is no evidence to suggest upcoding is significant or meaningful, mounting evidence indicates current coding practices cloud the interpretability of quality metrics composed of claims-based data in a variety of care settings.\textsuperscript{4,5,8,9} The related question is whether linking quality measures composed of hospital-assigned diagnosis and procedure codes to reimbursement accurately and reliably characterizes care quality or simply captures the comprehensiveness of hospitals’ administrative coding.

With increasing regulatory and financial pressures, hospitals will remain incentivized to maintain quality and maximize reimbursement, and balanced and have face validity. Because hospitals are reimbursed more for treating more severely ill, higher-risk patients, administrative data can either reflect the true level of risk in a hospital’s patient population or can overestimate—or underestimate—this risk.

Introducing financial incentives into quality assessment is a concern because it encourages upcoding (ie, maximizing a patient’s apparent number of comorbid conditions or the severity of these conditions through administrative coding).\textsuperscript{3-7} Although existing Centers for Medicare and Medicaid Services (CMS) fraud and abuse programs imply there is no evidence to suggest upcoding is significant or meaningful, mounting evidence indicates current coding practices cloud the interpretability of quality metrics composed of claims-based data in a variety of care settings.\textsuperscript{4,5,8,9} The related question is whether linking quality measures composed of hospital-assigned diagnosis and procedure codes to reimbursement accurately and reliably characterizes care quality or simply captures the comprehensiveness of hospitals’ administrative coding.

Pitfalls of Using Administrative Data to Measure Quality

Two characteristics of quality measurement programs likely drive hospital coding practices: financial incentives (eg, pay-for-performance programs, value-based purchasing programs) and reputational incentives (eg, hospital ranking, public reporting). Because risk-adjusted quality measures based on administrative data are central to most quality and reimbursement programs, hospitals are incentivized to optimize their coding practices to ensure that the complexity (or at least the apparent complexity) of their patient population’s risk is captured. From a financial standpoint, within existing hospital prospective payment models, a higher-severity diagnosis-related group (DRG) can increase reimbursement more than 2-fold.\textsuperscript{8} As such, comprehensive coding is in the best interest of hospitals to ensure true underlying differences in the case-mix and the inherently higher costs associated with treating patients with more severe illness (eg, addition of “sepsis” to an admission for pneumonia) within a given DRG are captured.

Since the implementation of value-based payment models, such as Medicare’s Hospital Readmission Reduction Program (HRRP), hospitals have significantly increased coding rates of secondary diagnoses, especially sepsis.\textsuperscript{4} Although the Hospital Readmission Reduction Program appears to have reduced readmission rates, these modest improvements disappear when accounting for upcoding of disease severity.\textsuperscript{3} Importantly, specific definitions and criteria for instances in which upcoding a clinical episode may be appropriate are not well-defined and are thus subjective. This creates an opportunity for hospitals to modify coding practices (in the absence of a specific clinical indication to do so) to enhance their risk adjustment for the purposes of quality evaluation, to capture higher reimbursement, or both. A 2021 report from the US Office of Inspector General suggests that in the period preceding the COVID-19 pandemic, inpatient stays coded at the highest severity increased by nearly 20%, from 2014 to 2019, with a simultaneous, paradoxical decrease in length of stay.\textsuperscript{8}

Strategies to Enhance the Reliability of Administrative Coding

The increase in coded disease severity noted for inpatient stays has 3 possible, not mutually exclusive, explanations: (1) the US population is aging and patients simply have more (or more severe) comorbid conditions, suggesting hospitalized patients may have developed greater illness severity over time; (2) incentives from quality programs, reimbursement programs, or both have pushed hospitals to change their coding practices
to more accurately capture their patients’ disease severity; and (3) in the absence of clear guidance of when upcoding is appropriate (or effective safeguards against this practice), US hospitals have increased upcoding to improve reimbursement and their apparent performance in quality assessment programs.\textsuperscript{3,4,6–8} All 3 elements may be contributing factors, although it is currently unclear what the respective contributions are. An inability to disentangle actual quality signals from those distorted by inaccurate coding practices is a potential threat to the goals of efficiently improving quality and ensuring the validity and reliability of the data provided by existing quality assessment programs.

Two potential solutions could help mitigate the risk of upcoding and help ensure that quality improvement and reimbursement programs are accurately and reliably characterizing hospital performance. The first involves ensuring the rigor and robustness of the administrative data used in such programs. CMS-approved qualified clinical data registries (QCDRs) have been developed across a spectrum of clinical specialties and diseases. The data and outcomes reported by QCDRs require transparency in data collection practices and the availability of details regarding the specific data elements they contain and are intended to have a greater level of rigor than administrative data used for billing purposes. In addition to their role in quality improvement activities, CMS could use the existing QCDR framework and leverage these data to audit hospital coding practices. This could allow QCDRs to serve as a separate, externally controlled source of data to validate hospital coding practices.

A second potential solution involves either modifying the type of data used when quality measures are tied to reimbursement or the manner in which such data are collected. Specifically, incorporating clinical information from the electronic health record could help ensure greater objectivity in the data used for risk adjustment. For instance, when comparing outcomes for heart failure and pneumonia across hospitals, improvements in patient outcomes are overestimated when using claims-based covariates relative to clinical data abstracted from the electronic health record, with clinical data better predicting patient illness severity.\textsuperscript{9} A variety of alternative approaches to collecting clinical data could be operationalized. Electronic health record systems are now ubiquitous in the US health care system and have increasing sophistication. As natural language processing and machine learning techniques evolve and improve, these applications could create opportunities to incorporate objective clinical data. For example, some machine-learning algorithms perform as well as radiologists in grading knee osteoarthritis with greater reproducibility.\textsuperscript{10} Scores derived from such technology could be incorporated to standardize a key covariate inherent to knee replacement quality measurement.

With increasing regulatory and financial pressures, hospitals will remain incentivized to maintain quality and maximize reimbursement. Although improving quality will largely be accomplished through local and regional efforts, the current approaches to quality measurement and risk adjustment remain susceptible to potential gaming through upcoding. If the goal is to evolve the US health care system toward higher reliability and to provide patients with better value care, this will require existing quality and reimbursement programs to modify their data collection and quality measurement practices to ensure the data reported and reimbursement provided truly reflect the patient population being treated, not just the robustness of the data that are coded.

\section*{REFERENCES}