Sisyphus and 30-Day Heart Failure Readmissions

Futility in Predicting a Flawed Outcome Metric*

Marvin A. Konstam, MD, Jenica Upshaw, MD, MS

And I saw Sisyphus in agonizing torment trying to roll a huge stone to the top of a hill. He would brace himself, and push it towards the summit with both hands, but just as he was about to heave it over the crest its weight overcame him, and then down again to the plain came bounding that pitiless boulder. He would wrestle again, and lever it back, while the sweat poured from his limbs, and the dust swirled round his head.

—Homer, The Odyssey, Book XI:593 (1)

In 2012, the Centers for Medicare and Medicaid Services (CMMS) began penalizing hospitals on the basis of the number of all-cause hospital readmissions within 30 days following an index hospitalization for heart failure, acute myocardial infarction, and pneumonia (2). If a hospital’s number of these events exceed the number expected, the hospital is penalized by a percentage of total Medicare revenue; that is, it is denied full payment for the services that it has provided to Medicare beneficiaries during that year. The maximum penalty has risen from an initial 1% to 3% of revenue. In understanding the construct of this metric, it is worth remembering its purpose: to save money for Medicare, a worthy goal, that is, to reduce the estimated $26 billion annual price tag associated with the 18% 30-day readmission rate across the country (3).

It is important, to both CMMS and providers, to predict the probability of readmission from available data related to a hospital’s patient population or from individual patients. For CMMS, the prediction drives estimation of the expected readmission rate. Both CMMS and providers seek the best possible estimation in order for the metric to operate fairly. Prediction also allows providers to cost effectively allocate resources needed to prevent readmission. However, as pointed out by Krumholz et al. (4) in this issue of JACC: Heart Failure, existing models constructed both inside (5) and outside (6,7) CMMS have relatively poor discrimination and predictive ability (C-statistic generally ranging from 0.5 to 0.7). Krumholz et al. (4) postulate that model discrimination could be improved by the addition of patient-derived socioeconomic, health status, adherence, and psychosocial characteristics. They interviewed 1,004 patients enrolled in the TeleHF (Telemonitoring to Improve Heart Failure Outcomes) trial (8) between 3 and 30 (median of 12) days following discharge from a hospitalization for heart failure. They constructed their model by combining variables derived from the interview with demographic and clinical variables available from the clinical record. The results were disappointing, with a C-statistic of 0.65, using a risk-score derived from all variables, compared with 0.62, using a risk-score derived solely by examining the medical record. In addition, with the large number of candidate variables considered for inclusion and the use of automated selection techniques, the modest discrimination seen may in fact be an overestimate due to model overfitting.

The 30-day readmission metric has generated something of a cottage industry, with hospitals and commercial vendors developing interventions expressly to reduce the penalties associated with failure to achieve the benchmark generated by CMMS. These
efforts, including specialized clinics and home-based disease management programs, target those patients falling within the denominator of this measure and often focus exclusively on the 30-day post-discharge period. Hospitals often deploy risk tools to funnel those patients perceived to be at highest risk into these programs in order to maximize return on investment. Therefore, improvement in the performance of predictive models could help reduce re-hospitalization rates and provide substantial value to hospitals.

The efforts of Krumholz et al. (4) are to be applauded for seeking information directly from patients and reaching beyond traditional medical factors. Because the revised model failed to substantially improve model discrimination, the authors discuss additional unmeasured factors that may be in play, including health system quality and a hospital’s propensity for admission. One potentially important limitation to the investigation is the varying time separation between discharge and interview. Interview timing, with the 30-day test period commencing at that point, may be relevant, because rehospitalization rates may decline over time following discharge (9), and answers to some patient-reported questions might vary depending on when assessed. Additionally, patients were excluded if they died or were hospitalized before the interview, posing a challenge to the discriminatory power of clinical variables ascertained during the hospitalization.

There are likely to be more fundamental difficulties in predicting the occurrence of the CMMS metric. We have previously expressed concern about the appropriateness of this metric (10), which we feel does not ideally serve the patient. Our concerns relate to each principal characteristic of the metric: the focus on hospitalization to the exclusion of mortality; the inclusion of all readmissions, rather than limiting the numerator to those with the same cause as the index admission; and the 30-day window. Just as each of these characteristics diminishes the metric’s relevance to improving patient care, they each also likely render the metric challenging to predict.

Mortality, estimated in Medicare fee-for-service beneficiaries at 10.7% during the 30 days following heart failure discharge (11), is a competing risk for hospitalization; that is, patients who die cannot be hospitalized, so perversely, an increased rate of death reduces the probability of rehospitalization. Heart failure clinical trials usually account for this issue by including both death and hospitalization in a composite endpoint. Being blind to death, the readmission metric is a utilization metric but should not be considered a quality metric. Treatments shown to reduce mortality have also reduced heart failure hospitalization (12-16). To the degree that appropriate medical care reduces mortality, this effect will partially offset the benefit on readmission. In fact, several analyses have demonstrated an inverse relationship between post-discharge mortality and hospitalization rates (17,18). Prediction models for hospitalization that do not account for the competing risk of death are predisposed to biased estimates due to informative censoring (19). Our own recently reported multistate prediction model demonstrates the potential application of multistate models to generate estimates of semicompeting risks such as hospitalization and death (20).

We know a lot about prognosis and treatment in heart failure, and use of that knowledge allows us to both predict and prevent post-discharge heart failure-related morbidity and mortality. However, rehospitalizations for all other causes combined are far more difficult to predict and to prevent, given the diversity of possible hospitalization diagnoses. It is far more difficult to devise a disease management strategy to prevent any admission than it is for admission from a specific known etiology. In fact, heart failure disease management studies have sometimes shown a modest increase in non-cardiovascular hospitalization (21,22), likely due to increased surveillance and detection of other illness. Such findings are an advantage to the patient, allowing for more expedient treatment. Furthermore, as Krumholz et al. (4) point out, there are multiple sources of “hospitalization stress,” resulting in a “transient syndrome of generalized risk” (23). We currently lack methodologies for measuring that stress. Cause-specific hospitalization is more valid in yielding undiluted rewards for implementing care processes known to improve outcomes. Likewise, the weakness in CMMS’ models for predicting all-cause hospitalization call into question the validity of distribution of penalties.

Finally, 30 days is a fleetingly short period for both improving a patient’s life and for predicting adverse outcomes, particularly when care directed toward a 30-day result may be at odds with that for long-term outcome benefit. The decision to hospitalize is subjective. A clinician incentivized around the 30-day metric might be biased to avoid hospitalizing a sick patient 28 days post-discharge, delaying a needed hospitalization until day 31. Moreover, a metric-conscious clinician might choose to avoid therapies known to improve long-term outcomes if those therapies might also threaten short-term stability. Beta-blockers and mineralocorticoid receptor antagonists (MRAs) are examples, even though their delayed initiation reduces the probability of downstream utilization (24,25). Although MRAs reduce long-term
mortality and hospitalization rates in patients with heart failure and reduced left ventricular ejection fraction (15), in-hospital initiation of MRAs in Medicare beneficiaries was found to be associated with increased 30-day rehospitalization (26). Paradoxical (and potentially harmful) clinician behavior in the face of an artificially short-term goal may have contributed to stymying the efforts of Krumholz et al. (4). Importantly, systems should be incentivized to build interventions designed for long-term patient health, not for 30 days.

Metrics that are flawed in driving patient benefit may also be difficult to predict, rendering their assessment even more flawed. They seem to create a perfect storm for suboptimal care, both by sidestepping the best interest of the patient and by thwarting assessment of risk for both clinicians, in their care, and for CMMS in its attempt at “fair” adjudication and penalty assignment. In the larger picture, we need to move entirely away from artificial metrics and penalties and toward greater direct responsibility of health care systems for quality and efficiency, with rewards linked to long-term patient benefit, through innovative approaches to care. Otherwise, those who have wished to devote their lives to healing will instead forever be endlessly rolling a boulder toward a remote and unreachable precipice.

**REFERENCES**


