Risk Aversion and Public Reporting. Part 2: Mitigation Strategies

David M. Shahian, MD, Jeffrey P. Jacobs, MD, Vinay Badhwar, MD,
Richard S. D'Agostino, MD, Joseph E. Bavaria, MD, and Richard L. Prager, MD

Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; Division of Cardiovascular Surgery, Johns Hopkins All Children's Heart Institute, All Children's Hospital, Johns Hopkins University School of Medicine, St. Petersburg, Florida, and Baltimore, Maryland; Division of Cardiothoracic Surgery, West Virginia University, Morgantown, West Virginia; Department of Cardiothoracic Surgery, Lahey Health, Burlington, Massachusetts; Division of Cardiovascular Surgery, University of Pennsylvania, Philadelphia, Pennsylvania; and Cardiac Surgery Department, University of Michigan, Ann Arbor, Michigan

Part 1 of this review summarizes the consequences of risk aversion and the observational studies and surveys relevant to this phenomenon, almost all of which are derived from cardiac surgery and interventional cardiology. In Part 2, we describe the root cause of risk aversion—the belief by providers that current risk adjustment is inadequate to account for the severity of their highest-risk patients, thereby prejudicing their publicly reported performance scores. Evidence supporting the robustness of current risk adjustment is presented, as well as nine potential strategies to further mitigate risk aversion: optimization of data source, risk models, and performance measures; exclusion of high-risk patients; exclusion of non–procedure-related end points; separate reporting of high-risk patients; reporting by condition or diagnosis rather than by procedures; reporting at the hospital or program level rather than the physician level; collaborative, cross-disciplinary decision making; active surveillance for risk aversion; and improved stakeholder education. Of these, the first is most desirable, widely applicable, and resistant to gaming.

(Ann Thorac Surg 2017;104:2102–10)

As demonstrated in Part 1 of this review [1], risk aversion associated with public reporting undoubtedly exists, although its true extent remains uncertain. Empirical data and examples of this practice are derived mainly from interventional cardiology and cardiac surgery and are more consistent and convincing for the former.

In Part 2 of this review, we examine the root cause of risk aversion—the belief by providers that current risk adjustment is inadequate to account for the severity of their most critically ill patients, whose anticipated worse outcomes might prejudice their report card ratings. We also examine a variety of strategies that have been proposed or implemented to mitigate risk aversion.

Does Risk Adjustment Provide Adequate Protection?

Regardless of the actual extent of risk aversion related to public reporting, many cardiac surgeons and interventional cardiologists believe they are inadequately protected by current risk models when they accept high-risk patients. A detailed discussion of the theory and practice of statistical risk modeling for provider profiling is beyond the scope of this article, and relevant references and examples are available [2–9].

No risk model is perfectly predictive for every patient and outcome, and investigators have described numerous theoretical and practical concerns that any public reporting program should consider [10–15]. For example, not all conceivable risk factors, or combinations thereof, are captured even by the best databases, potentially introducing unmeasured confounding. Some risk factors are included in registries but may have excessive missing data that preclude their use or may be present so rarely that they cannot be modeled. Similarly, random sampling variation makes it difficult to model outcomes that occur infrequently, especially with small sample sizes (eg, individual physician reporting). Intentional upcoding of risk factors (a form of "gaming") by registry participants or inadvertent miscoding caused by poorly specified variables may, over time, dilute the true effect of some risk model predictors by including patients whose actual clinical state does not meet the spirit of the variable. This phenomenon may also give the false impression that risk factor prevalence is increasing in the population, as observed by Green and Wintfeld [16] in the early New York experience.

Finally, in risk model development there is often a tension between models with many predictor variables, which can better accommodate patients with important but uncommon risk factors, and so-called parsimonious models, which are less time consuming and labor intensive to use but may not contain infrequently occurring risk factors. Well-constructed parsimonious models may have overall performance nearly identical to that of

---

Address correspondence to Dr Shahian, Department of Surgery, Massachusetts General Hospital, 55 Fruit St, Bulfinch 284, Boston, MA 02114; email: dshahian@partners.org.

Published by Elsevier Inc.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2017.06.076

full models with more predictors. However, these may underestimate the risk of the few patients who have rare but particularly high-risk characteristics that are not included in the models. Only a few such patients could theoretically affect the risk-adjusted outcomes of a specific provider, especially during short sampling periods or with small volumes.

Given all of these potential issues, what is known about the performance of risk models used in cardiac operations and percutaneous coronary interventions (PCI)? Virtually all published risk models for these procedures have acceptable statistical calibration and discrimination, the most basic tests of risk model performance. A number of studies have also specifically examined the real world protection afforded by risk adjustment in both cardiac surgery and PCI.

Cardiac Surgery Risk Models

In early reports of the New York Cardiac Surgery Reporting System (CSRS), Hannan and colleagues [17] and Chassin and colleagues [18] demonstrated that New York coronary artery bypass grafting (CABG) risk models performed well at the patient level across all strata of expected risk. There was slight overprediction of risk for the most severely ill patients, suggesting more than adequate protection for programs accepting such patients. Importantly, both studies documented a significant negative correlation between expected mortality and risk-adjusted mortality (RAM) in each of the first 4 years of the New York CSRS; that is, programs that cared for the highest-risk patients often had the lowest RAM, and vice versa. This may reflect the excellent protection afforded by these risk models or the superior performance of programs that are willing to accept high-risk patients (a desirable matching of risk and capability), or both.

Hannan and colleagues [19] further studied the adequacy of risk adjustment using 1990 to 1992 New York CABG data from 31 hospitals and 87 surgeons. This analysis was performed for all 44,918 patients as well as a high-risk ($\geq$7.5% predicted mortality risk) subset of 3,281 patients (7.3% of the total). Observed mortality for the high-risk group was 15.88% vs 1.96% for the remaining 41,637 low-risk patients, but the RAM was actually lower for high-risk patients (2.94% vs 3.02%). No hospital had significantly different RAM for low-risk patients alone vs all patients (ie, with high-risk patients included), and half of the hospitals had lower RAM for all of their CABG patients than for only their lower-risk patients. At the hospital level, RAM for all CABG patients correlated strongly with RAM for low- and high-risk patients separately. Even if a higher threshold were used to classify high-risk patients (ie, three times average, or 9%), the RAM for high-risk patients in 18 of 31 hospitals was less than the state average of 2.99%. All of these findings affirm the adequacy of risk adjustment and do not support the belief that avoiding high-risk patients would improve risk-adjusted outcomes.

Contemporary investigations have verified the protection afforded by modern risk models developed from large clinical registries. Englum and colleagues [20] studied 494,955 patients who underwent isolated CABG between 2008 and 2010 and who were included in The Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database (ACSD). The 1,002 sites that cared for these patients were divided into quintiles by their average expected risk of operative mortality, ranging from 1.46% in quintile 1 to 2.87% in quintile 5. The overall calibration of the STS risk model in this cohort was excellent, although there was slight overprediction of death among the highest-risk 1% of patients (expected risk >20%). Observed-to-expected (O/E) mortality ratios were not significantly different than unity for any but the highest-risk quintile, in which O/E was 0.80 (95% confidence interval [CI], 0.77 to 0.84), indicating better-than-expected performance. Similar findings were observed when all of a hospital's highest-risk patients during a 3-year period were analyzed as if they had occurred in 1 "nightmare year" scenario.

These analyses demonstrated that the STS CABG risk model provides adequate risk adjustment even for hospitals that care for the highest-acuity patients. They challenge the widely held notion that avoiding high-risk patients will improve a provider's risk-adjusted outcomes. Indeed, the STS CABG risk model (like many other risk models) offers some degree of "overprotection" for surgeons and hospitals caring for the highest-risk patients.

Interventional Cardiology Risk Models

Sherwood and colleagues [21] performed a similar investigation for PCI, analyzing 624,286 patients from 1,168 sites that contributed data to the American College of Cardiology National Cardiovascular Data Registry (NCDR) CathPCI data set in 2010. Using the NCDR PCI mortality risk model, they found good calibration over the wide range of predicted and observed mortality rates. Hospitals were grouped into quintiles by overall hospital expected mortality rates, and O/E ratios were estimated for each quintile. Sensitivity analyses ranked hospitals by quintiles by their percentage of shock, cardiac arrest, or other extremely high-risk ($\geq$10% mortality) patients. Overall, O/E ratios for most risk quintiles were close to 1, except for the hospitals in the highest-risk quintile, for whom performance was better than expected (0.91; 95% CI, 0.87 to 0.96); results were similar for the sensitivity analyses. These hospitals also had somewhat lower RAM than the lowest-risk hospitals. When each site's highest-risk patients from 2009 to 2011 were combined into a single, simulated, exceptionally high-risk year, O/E ratios all remained approximately 1, there was no increased identification of outlier hospitals, and there was generally good agreement of O/E ratios between the extreme high-risk year and average-year values. Thus, the authors argued that current PCI risk models are adequate to cover the risk of very high-acuity patients and will not unfairly penalize providers who care for them [22].

Mitigating Risk Aversion

Notwithstanding these reassuring findings, persistent provider anxiety and the resulting potential for risk aversion are a continuing concern, and numerous mitigation strategies have been suggested in response [10, 11, 13, 14, 23–25]. They all share the common goal of enhancing provider trust in the accuracy and fairness of the performance measures and processes used in public reporting, thereby addressing the major driver of risk-averse behavior.

Optimal Data Source, Risk Models, and Performance Measures

Because providers often mistrust the ability of risk models and quality metrics to accurately characterize their performance [26–30], the ideal defense against risk aversion is to optimize all steps of the quality measurement process.

DATA SOURCE. The development of credible performance measures begins with use of the best data, which we believe are prospectively collected, standardized, clinical registry data [23, 31, 32], as exemplified by the STS National Database and the American College of Cardiology NCDR. Although registry data are readily available in cardiology and cardiac surgery, other specialties lag behind, and the establishment of such registries across all of health care should be an urgent priority. These structured clinical data are much more useful and accurate for performance measurement than the largely unstructured information collected in electronic health records; similarly, administrative claims data were designed primarily for billing and not performance measurement [31, 32].

DATA AUDIT AND ADJUDICATION. Even when derived from clinical registries, the accuracy of data used for public reporting must be verified. In their study of the early New York CABG report card experience, Green and Wintfeld [16] reported that between 1989 and 1991, the coded prevalence of numerous risk factors increased dramatically, including renal failure, 0.4% to 2.8%; congestive heart failure, 1.7% to 7.6%; chronic obstructive pulmonary disease, 6.9% to 17.4%; unstable angina, 14.9% to 21.8%; and low ejection fraction, 18.9% to 22.2%. They hypothesize that this might have resulted from upcoding of risk factors by some programs to inflate their expected mortality rates, thereby reducing their mortality O/E ratios and RAM without actually improving outcomes.

Chassin and colleagues [18] asserted that these findings were largely the result of risk factor definition changes during the early years of public reporting, as well as some undercoding in 1989, which was identified and corrected; risk factor prevalence stabilized thereafter. However, periodic data quality issues did persist and were the impetus for ongoing audit. Hannan and colleagues [23] encountered instances not only of inflated risk factor prevalence but also of omitted cases in which the patients had died.

Thus, multiple levels of data audit [18, 23] and adjudication [33] are essential. Data that are out of range, unusually high-risk factor prevalence (eg, upcoding), critical outcomes, and case completeness, should all be reviewed. Submitted case lists should be compared with hospital operating room logs to ensure that patients with poor outcomes were not omitted. STS National Database studies comparing cases submitted from STS sites against procedures billed to Centers for Medicare and Medicaid Services suggest high completeness rates (98% in 2012) [34].

This data quality oversight may be accomplished through a combination of random, nationally supervised audits, such as the annual STS audit of 10% of participants, which consistently demonstrates 96% to 97% data accuracy; and state or regional adjudication conducted by physicians and data managers, such as the Massachusetts Data Analysis Center adjudication process for CABG and PCI report cards [35, 36]. The latter approach is particularly well suited to review all patient records coded for uncommon risk variables with high impact (eg, emergency status or cardiogenic shock), as well as all excluded patients, to ensure that relevant specifications are met [33, 35].

Ideally, clinical registry data used for public reporting should be verified against governmental sources [23], such as the Statewide Planning and Research Cooperative System registry in New York [23], the Social Security Death Master File (currently unavailable) [37, 38], Centers for Medicare and Medicaid Services Medicare Provider Analysis and Review files (primarily for patients aged >65 years) [34, 39], and the National Death Index [40].

RISK MODELS. To reassure providers that their outcomes are being fairly adjusted for inherent patient severity, risk models should be available for all performance measures. These should be developed through collaborations of clinicians and statisticians, using state of the art modeling techniques. All available, clinically plausible risk factors should be evaluated for their contribution to the models, and model performance should be periodically reassessed, especially for high-risk patients.

In some instances, adding new or infrequently occurring but highly influential risk factors to models may reassure providers and diminish risk aversion. In current practice, this has occurred most often in PCI (eg, adding the SYNTAX [Synergy between PCI with Taxus and Cardiac Surgery] score, extreme frailty, or prohibitive surgical risk [10]), especially for acute myocardial infarction (AMI) or for other high-risk presentations. Resnic and colleagues [41, 42] studied the proportion of PCI shock cases before and after institution of public reporting in Massachusetts in 2003. Between 2003 and 2005, the percentage of all PCI procedures with shock decreased from 2.28% to 1.29% (43% reduction), and observed mortality declined from 1.71% to 1.56%. The authors speculate that in a public reporting environment, “case selection creep” or “risk avoidance creep” may occur—a gradual shift in indications from high-risk and high-benefit to low-risk but low-benefit procedures. In response to this undesirable trend, in late 2005 Massachusetts added a new risk variable to its PCI model—

compassionate use—defined by coma on admission (Glasgow Coma Scale <7 and emergency status), active hemodynamic support during PCI (cardiopulmonary bypass, percutaneous ventricular assist device, extracorporeal membrane oxygenation), or ongoing cardiopulmonary resuscitation required before initiation of PCI. Subsequently, from 2005 to 2007, there were 29,784 PCI admissions in Massachusetts, of which 5,588 ST elevation myocardial infarction (STEMI) or shock patients had an overall mortality rate of 5.7%. Within this subgroup, 96 patients (1.7%) had at least one compassionate use feature and their mortality rate was 69.8% vs 4.5% for noncompassionate use patients with shock or STEMI (adjusted mortality odds ratio, 27.3; 95% CI, 14.5 to 47.6); these patients accounted for more than 21% of all PCI deaths in the shock and STEMI group. After the compassionate use variable was included, the declining prevalence of shock cases among the PCI population was reversed, increasing to 1.7% in 2007, which suggests greater provider willingness to accept high-risk cases.

Because of their substantial effect on performance scores, it is critical to prevent gaming by carefully adjudicating all cases coded for these very high-risk factors. In 2009 content experts validated 91 of 128 submissions (71%) of the Massachusetts PCI compassionate use variable, and this had increased to 78% in 2014 [35].

PERFORMANCE MEASURES. Risk aversion is more likely if providers distrust the performance measures used for public reporting. This is not a baseless concern, given the recent proliferation of commercial and "public service" report cards, often based on problematic methods that yield inaccurate and inconsistent results [43, 44]. These are typically not published in the peer-reviewed literature nor are they endorsed by the National Quality Forum, currently the best assurance that a measure is suitable for public reporting (all STS measures are submitted to the National Quality Forum).

Measures used for health care public reporting must address a number of important statistical and practical considerations that affect their validity and reliability [3, 9, 32, 43–47], at a minimum the important features listed in Table 1. Performance metrics should focus on outcomes, because these are, in Donabedian's words, the "ultimate validators of the effectiveness and quality of medical care" [48], yet these need not be the only quality metrics. For coronary revascularization, many have criticized the use of short-term mortality as the sole indicator of performance and have argued for a broader range of measures. These might include complication rates, longer-term outcomes, appropriateness, access to care, disparities, functional status, freedom from angina or other symptoms, resource use, readmissions, and reinterventions [10, 11, 13, 14, 23, 49, 50].

Process measures based on nationally accepted guidelines offer the advantages of easier calculation and direct actionability, but these are only an indirect measure of what matters most to patients. Patient-reported outcomes and experience of care are also being studied for use as quality metrics because they reflect the judgment of the recipients about the quality of health care they received. Finally, composite measures embracing multiple quality domains (eg, risk-adjusted mortality, morbidity, and process measures) provide more comprehensive quality measurement and a larger effective sample size than single-focus metrics, and these are used preferentially by STS for cardiothoracic surgical procedures [51–56].

Table 1. Essential Considerations in Risk Model and Performance Measure Development

- Data source
- Data quality
- Selection of target population and end point(s)
- High impact
- Adequate denominator population (sample size)
- Sufficient end point frequency (event rates)
- Variation in performance among providers
- Types of measures
- Level and accuracy of attribution (eg, hospital, physician)
- Patient observation period (eg, in-hospital, 30-day, both, longer)
- Provider observation period (eg, 1 year, 3 year)
- Statistical methodology (eg, risk adjustment, management of clustered observations, shrinkage)
- Risk model performance
- Performance measure testing (eg, validity and reliability)
- Outlier determination
- Consumer-friendly presentation formats
- Monitor and mitigate unintended negative consequences

Although there is no widely accepted lower threshold for performance measure reliability (signal-to-noise ratio) [45, 57–59], the STS generally requires a reliability of 0.50 for its quality metrics. The corresponding volumes necessary to achieve this level of reliability may render a substantial number of providers ineligible to receive a quality score [53–56]. Although unfortunate, STS regards this as preferable to assigning scores that have inadequate reliability.

Exclusion of High-Risk Patients

Exclusion of certain high-risk patient subgroups is frequently requested by specialties whose performance is publicly reported, as illustrated by PCI report cards in New York and Massachusetts. Although an intuitive and simple proposal to mitigate risk aversion, the presumption underlying this approach (ie, provider results are adversely affected by treating high-risk patients) is problematic. In an attempt to demonstrate this to interventionalists in New York, Hannan and colleagues [23, 60] analyzed New York CABG and PCI performance with and without shock cases. When these were included, hospitals and physicians often had better O/E ratios for shock cases than for other cases, and outliers were similar with or without shock cases included [60]. In fact, given the excellent protection afforded to very high-risk patients by current risk models, exclusion of these patients might theoretically lead to worse risk-adjusted outcomes [60, 61]. It could also be argued that if particularly high-risk features can be identified as the basis for case exclusions, then they could similarly be used as risk model variables without excluding these cases.

Notwithstanding these findings, New York began excluding AMI-shock patients from PCI reporting in 2006 [23, 60, 61]. In 2005, when these patients were still included, 83 New York shock patients underwent PCI (0.15% of all PCI procedures), and the in-hospital or 30-day mortality rate was 33.7%. During the subsequent 3 years when shock patients were excluded, 133, 146, and 138 PCIs were performed in shock patients (an average 67% increase per year), whereas there was little increase in the number of nonrefractory shock patients; in 2013, 236 refractory shock patients received PCI (0.49% of all PCIs). Mortality rates for shock patients receiving PCI also increased to 45% to 50% [23, 61], suggesting that providers were more willing to treat even higher-risk shock patients. In 2011 patients with documented post-arrest anoxic encephalopathy were also excluded from PCI public reporting in New York [60].

Bangalore and colleagues [62] used the National Inpatient Sample to compare rates of invasive strategies (catheterization/PCI/CABG) in New York with corresponding rates in New Jersey, Michigan, and California for patients with post-MI cardiogenic shock. After the permanent exclusion of cardiogenic shock from public reporting in New York in 2008, a significantly higher proportion of propensity-matched patients received PCI (odds ratio, 1.50; 95% CI, 1.12 to 2.01; $p = 0.005$) compared with the 2002 to 2005 period when shock patients were included, although patients in Michigan had even greater proportional increases in PCI (odds ratio, 1.93; 95% CI, 1.45 to 2.56; $p < 0.001$).

McCabe and colleagues [63] used state all-payer databases in New York, Massachusetts, Michigan, New Jersey, and California to assess the effect of excluding New York cardiogenic shock patients starting in 2006. The comparator states did not publicly report PCI outcomes or did not have shock exclusion policies in their reporting programs. After shock patients were excluded, the PCI rate for New York cardiogenic shock patients increased significantly from 30.5% (2002 to 2005) to 39.7% (2006 to 2012), and the adjusted relative risk (aRR) of receiving a PCI for shock increased (aRR, 1.28; 95% CI, 1.19 to 1.37; $p < 0.001$). In the comparator states, there was a much smaller increase in the rate of PCI for shock in the same time frame (aRR, 1.09; 95% CI, 1.05 to 1.13; $p < 0.001$), although the absolute rate of PCI for shock in New York remained lower throughout the study period. After shock patients were excluded from public reporting in New York, the in-hospital mortality rates for AMI-shock patients in that state decreased from 47.1% to 35.5%, and their adjusted risk of in-hospital death decreased (aRR, 0.76; 95% CI, 0.72 to 0.81) to a greater extent than in comparator states (aRR, 0.91; 95% CI, 0.87 to 0.94), suggesting that more aggressive use of PCI salvaged some shock patients. These findings demonstrate not only the effect of excluding shock patients in New York but also the value of studying all patients with a particular clinical presentation (ie, AMI with shock), not just those receiving interventions, because this more broadly assesses the overall population effect of policy changes.

Despite the apparent effect of shock exclusion on provider willingness to perform high-risk PCI, only 0.25% of all PCIs were actually excluded from public reporting in the first 3 years after this policy was implemented in New York [63]. This suggests that the psychological reassurance and sense of fairness provided by these changes was disproportionate to the number of patients actually affected. As noted by Hannan [60], exclusion of high-risk patients does not substantially decrease the number of patients in the denominator population but does eliminate a disproportionate percentage of the deaths; the mortality rate for the remaining patients may then be so low that meaningful differentiation among providers is problematic (eg, 0.65% for New York PCI patients in 2011 without shock, hemodynamic instability, or an MI within 24 hours). Hannan recommends exclusions only for rare presentations with very high expected mortality, especially if the risk for specific patients is difficult to quantify with the available data (eg, the wide range of cardiogenic shock severity).

To further explore the implications of the New York shock exclusion policy, Hannan and colleagues [61] recently analyzed the effect on risk-adjusted mortality rates and outliers of excluding or including all shock cases from New York PCI public reporting compared with the current practice of excluding only refractory shock cases. When the investigators compared risk-adjusted physician and hospital mortality rates and outliers for the current reporting criteria and the two alternative approaches, the correlation coefficients were 0.99 and 0.92 or greater, respectively. Outliers were very similar; however, at the physician level, 10% to 15% of interventionalists who were outliers by one exclusion method were not by the other. Despite the substantial overall concordance of results regardless of the specific shock exclusion rule, Resnic and Majithia [49] noted that these results demonstrated a 13.5% false-positive and 20% false-negative outlier designation at the physician level when the modeling strategy included all shock patients.

Overall, PCI providers enthusiastically support exclusions, and risk aversion appears to decrease when these rules are implemented. However, the care of critically ill AMI patients with intractable shock or uncertain neurologic status, in which the decision for or against intervention must be made quickly and without certainty about the patient's survivability, is somewhat unique; this may not be a paradigm that is readily transferable to other areas. Existing public reporting programs in cardiac surgery focus on commonly performed, largely nonemergency procedures with adequate sample size and end points (eg, isolated and combined CABG and valve procedures). Excellent risk models cover most exceptional risk presentations, such as salvage status or shock, which are uncommon. Patients with infrequent high-risk scenarios (eg, emergency operations for patients with aggressive endocarditis) [25] might be potential exclusion candidates. Whether such cases are excluded or included with adjustment, to make truly informed decisions patients and their families must be aware of not only the high risks but also the potential benefits of intervention [25].

When exclusions are used to mitigate risk aversion, audit of all excluded patients is essential to ensure they meet the explicit, prespecified criteria. This is much more difficult to achieve on a national scale compared with the state level, where local committees of clinical experts can be periodically convened and provided with access to detailed patient record information [33, 35]. For example, in 2009 Massachusetts excluded exceptionally high-risk patients (extreme high-risk characteristics not included in the risk models, and PCI deemed the best or only option for improving survival) from reporting. That year, the state adjudication committee of clinical experts validated only 7 of 26 patients (27%) coded by programs as exceptional risk [35]. In the fiscal year 2014 report 5 years later, only 18.2% of patients submitted as exceptional risk were validated [35]. These high rejection rates by expert adjudication panels demonstrate the potential for intentional gaming or unintentional misapplication of exclusion criteria and illustrate the necessity for careful review of every excluded patient.

Separate Reporting of High-Risk Patients

Rather than completely excluding high-risk patients from reporting, another option is separate reporting of these patients, ideally with separate risk models, which acknowledges the inherent heterogeneity of patient populations for many procedures. For example, two separate PCI reports are issued annually in Massachusetts, one for patients who have sustained a STEMI or who are in shock and another for the remaining patients. The overall mortality in their fiscal year 2014 report was 5.04% for the first group and 0.40% for the remaining patients [35]. New York State separately reports all cases and nonemergency cases [64].

The only major disadvantage of this approach is smaller sample sizes for the separate groups, which may reduce measure reliability and ability to discriminate performance. Perhaps the optimal strategy is to provide reports documenting low-risk and high-risk patients separately, as well as a combined report including all patients, which would have larger sample size and reliability.

Report at Program or Hospital Level Rather Than Individual Physician Level

Although the STS has historically reported outcomes at the STS participant level, typically a hospital or practice group, there is increasing interest by the media and consumers in individual provider outcomes. Health care outcomes do vary at the physician level [59, 65], but reporting of individual practitioner outcomes raises numerous issues [10–12, 14, 26, 29, 41, 59, 66, 67]. First, most complex interventions are a team effort, and focusing only on the surgeon or proceduralist may not be appropriate. Also, from a statistical perspective, sample sizes for particular diagnoses or procedures are typically small for individual doctors. To achieve satisfactory measure reliability [57, 59], it is often necessary to aggregate multiple years of data or to broaden the outcomes and types of patients included. For example, the STS individual surgeon composite for adult cardiac surgeons [65], which has a high reliability of 0.81, includes data on five of the most commonly performed adult cardiac procedures, both morbidity and mortality outcomes, and rolling 3-year data windows.

Reporting individual practitioner performance has also been criticized because this is not the level at which most quality improvement activities occur (ie, the hospital). Reporting at the individual provider level may thus diminish the actionability of public reporting programs by uncoupling the measured entity and the usual effector of quality improvement activities.

Finally, compared with hospital or program-level reporting, physician or surgeon reporting may be more likely to stimulate risk aversion, given the potential reputational effect on the individual. The potentially worse outcomes of a few very high-risk patients are absorbed into the overall results of a Department and their public reporting effect is thus diluted. However, surgeons fear that one or two additional deaths in high-risk patients could have a proportionately greater effect on their individual results, as the denominator is much smaller [28]. Because individual doctors, not Departments, generally make the final treatment decisions about high-risk patients, the potential for risk aversion may therefore be higher with surgeon-level reporting.

Report Outcomes at the Condition or Diagnosis Level Rather Than by Procedure

Some commentators [10, 14, 49] have argued for reporting at the condition or diagnosis level (eg, all patients with AMI, including those with shock) rather than, or in addition to, the procedural level (ie, only patients who received PCI). This strategy would enable better monitoring of the effect of risk aversion on patients denied an intervention and would make the relevant clinicians responsible for an entire population of patients, not just those receiving an intervention. For example, as described in Part 1 of this review [1], results of PCI for AMI patients may appear excellent, but what is not apparent from this procedure-based metric is the worse outcomes of high-risk patients who were denied an intervention.

Although a feasible approach for AMI, this strategy might be difficult to extrapolate to other conditions or diagnoses. AMI is a discrete diagnosis with only a few subcategories and a limited number of potential interventions (PCI, CABG, or medical therapy), which is not true for many other conditions. Also, although credible clinical data and risk models are available for many procedures, they may not be readily accessible for all patients with a given diagnosis, including those who do not receive an intervention, thus making risk-adjusted outcomes comparisons difficult. Nevertheless, this approach may be more feasible in the future with the greater availability of high-quality clinical data and evolving reimbursement models that reward excellent care of an entire population, not just those receiving interventions [14].

Exclusion of Nonprocedure Related Outcomes

Some have argued that only outcomes clearly related to the procedure should be included in public reporting [10]. However, this approach would require highly subjective assessments that may not be applied consistently, it would be difficult to scale, and it may increase the potential for gaming.

Collaborative Decision Making

For high-risk procedures, it is optimal care for teams of experts from the same and related specialties (eg, the heart team) to collaborate in the final recommendation regarding interventions in high-risk patients and to present these to patients and their families. This proactive approach reduces the decision making burden on individual providers and leverages peer pressure to promote management that is in the best interest of the patient, which may include high-risk procedures. However, human nature suggests that those members of the expert team not likely to be personally involved in caring for the patient may be more inclined to recommend these interventions because the potentially poor outcomes will not be attributed to them.

Active Surveillance for Risk Aversion

The presence and extent of risk aversion must be periodically studied in public reporting environments, as has been done for PCI in AMI patients, and results should be monitored over time to assess the effectiveness of mitigation strategies. One potential strategy is peer auditing of all initially referred patients who are rejected for treatment.

Stakeholder Education

Many stakeholders do not understand how to read or interpret public report cards [68]. Consequently, providers are concerned that the public will misinterpret report card results, and this fosters risk-averse behavior [26–30]. Education must be provided to providers, patients, regulators, and payers so they can better comprehend report cards and their limitations, including the correct interpretation of “risk-adjusted” results [69].

How report card information is presented also facilitates stakeholder understanding. Professor Judith Hibbard and her colleagues [70] have been leaders in efforts to make public report cards more comprehensible for typical readers, including a reduction in the amount of material presented and the use of visual aids, such as star ratings.

Conclusions

Numerous strategies to mitigate risk aversion have been proffered, of which robust audit and optimization of data sources, risk models, and performance measures are the most appealing, widely applicable, and resistant to gaming. Providers often request exclusion of high-risk cases, but this is probably appropriate for only a tiny fraction of cases. When exclusions are allowed, each excluded case must be carefully adjudicated to prevent gaming. Ultimately, elimination of risk aversion depends on the professionalism of providers and their commitment to do their best for every patient, including those at high risk, because these patients may potentially benefit the most from appropriate but challenging interventions.

References

1. Shahian DM, Jacobs JP, Badhwar V, D’Agostino RS, Bavaria JE, Prager RL. Risk aversion and public reporting. Part 1: observations from cardiac surgery and interventional cardiology. Ann Thorac Surg 2017;104:2093–101.
2. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. Chicago: Health Administration Press; 2012.
3. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. Circulation 2006;113:456–62.
4. Ash A, Fienberg S, Louis T, et al; the COPSS-CMS White Paper Committee. Statistical issues in assessing hospital performance. Commissioned by the Committee of Presidents of Statistical Societies. Available at https://www.Cms.Gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/statistical-issues-in-assessing-hospital-performance.Pdf. 2011. Accessed May 30, 2017.
5. Normand SL, Shahian DM. Statistical and clinical aspects of hospital outcomes profiling. Stat Sci 2007;22:206–26.
6. Shahian DM, O’Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. Ann Thorac Surg 2009;88(1 Suppl):S2–22.
7. O’Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009;88(1 Suppl):S23–42.
8. Shahian DM, O’Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3—valve plus coronary artery bypass grafting surgery. Ann Thorac Surg 2009;88(1 Suppl):S43–62.
9. Shahian DM, He X, Jacobs JP, et al. Issues in quality measurement: target population, risk adjustment, and ratings. Ann Thorac Surg 2013;96:718–26.
10. Gupta A, Yeh RW, Tamis-Holland JE, et al. Implications of public reporting of risk-adjusted mortality following percutaneous coronary intervention: misperceptions and potential consequences for high-risk patients including nonsurgical patients. JACC Cardiovasc Intervent 2016;9:2077–85.
11. Young MN, Yeh RW. Public reporting and coronary revascularization: risk and benefit. Coron Artery Dis 2014;25:619–26.
12. Rosenbaum L. Scoring no goal—further adventures in transparency. N Engl J Med 2015;373:1385–8.
13. Kirtane AJ, Nallamothu BK, Moses JW. The complicated calculus of publicly reporting mortality after percutaneous coronary intervention. JAMA Cardiol 2016;1:637–8.
14. Wasfy JH, Borden WB, Secemsky EA, McCabe JM, Yeh RW. Public reporting in cardiovascular medicine: accountability, unintended consequences, and promise for improvement. Circulation 2015;131:1518–27.
15. Cutlip DE, Ho KK, Kuntz RE, Baim DS. Risk assessment for percutaneous coronary intervention: our version of the weather report? J Am Coll Cardiol 2003;42:1896–9.

16. Green J, Wintfeld N. Report cards on cardiac surgeons—assessing New York state's approach. N Engl J Med 1995;332: 1229–33.
17. Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York state. JAMA 1994;271:761–6.
18. Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. N Engl J Med 1996;334:394–8.
19. Hannan EL, Siu AL, Kumar D, Racz M, Pryor DB, Chassin MR. Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients? Med Care 1997;35:49–56.
20. Englum BR, Saha-Chaudhuri P, Shahian DM, et al. The impact of high-risk cases on hospitals' risk-adjusted coronary artery bypass grafting mortality rankings. Ann Thorac Surg 2015;99:856–62.
21. Sherwood MW, Brennan JM, Ho KK, et al. The impact of extreme-risk cases on hospitals' risk-adjusted percutaneous coronary intervention mortality ratings. JACC Cardiovasc Interv 2015;8(1 Pt A):10–6.
22. Sherwood MW, Peterson ED. Risk adjusted mortality ratings and public reporting for high-risk PCI. JACC Cardiovasc Interv 2015;8:1134–5.
23. Hannan EL, Cozzens K, King SB, 3rd, Walford G, Shah NR. The New York state cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. J Am Coll Cardiol 2012;59:2309–16.
24. Peberdy MA, Donnino MW, Callaway CW, et al. Impact of percutaneous coronary intervention performance reporting on cardiac resuscitation centers: a scientific statement from the American Heart Association. Circulation 2013;128:762–73.
25. Bavaria J. Presidential address: quality and innovation in cardiothoracic surgery: colliding imperatives? Presented at The Society of Thoracic Surgeons 53rd Annual Meeting, Houston, Texas, January 21–25, 2017.
26. Sherman KL, Gordon EJ, Mahvi DM, et al. Surgeons' perceptions of public reporting of hospital and individual surgeon quality. Med Care 2013;51:1069–75.
27. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care—a survey of cardiovascular specialists. N Engl J Med 1996;335:251–6.
28. Jarral OA, Baig K, Pettengell C, et al. National survey of UK consultant surgeons' opinions on surgeon-specific mortality data in cardiothoracic surgery. Circ Cardiovasc Qual Outcomes 2016;9:414–23.
29. Narins CR, Dozier AM, Ling FS, Zareba W. The influence of public reporting of outcome data on medical decision making by physicians. Arch Intern Med 2005;165:83–7.
30. Burack JH, Impellizzeri P, Homel P, Cunningham JN Jr. Public reporting of surgical mortality: a survey of New York state cardiothoracic surgeons. Ann Thorac Surg 1999;68: 1195–200.
31. Shahian D, Silverstein T, Lovett A, Wolf R, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. Circulation 2007;115:1518–27.
32. Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? J Thorac Cardiovasc Surg 2005;129: 1309–17.
33. Barringhaus KG, Zelevinsky K, Lovett A, Normand SL, Ho KK. Impact of independent data adjudication on hospital-specific estimates of risk-adjusted mortality following percutaneous coronary interventions in Massachusetts. Circ Cardiovasc Qual Outcomes 2011;4:92–8.
34. Jacobs JP, Shahian DM, He X, et al. Penetration, completeness, and representativeness of The Society of Thoracic Surgeons Adult Cardiac Surgery Database. Ann Thorac Surg 2016;101: 33–41.
35. Mass Data Analysis Center, Department of Health Care Policy, Harvard Medical School. Adult percutaneous coronary intervention reports. Available at http://www.massdac.org/reports/cardiac-study-annual/. Accessed May 12, 2017.
36. Mass Data Analysis Center, Department of Health Care Policy, Harvard Medical School. Adult coronary artery bypass graft surgery annual reports. Available at http://www.massdac.org/reports/cardiac-study-annual/. Accessed May 12, 2017.
37. Jacobs JP, O'Brien SM, Shahian DM, et al. Successful linking of The Society of Thoracic Surgeons database to Social Security data to examine the accuracy of Society of Thoracic Surgeons mortality data. J Thorac Cardiovasc Surg 2013;145: 976–83.
38. Jacobs JP, Edwards FH, Shahian DM, et al. Successful linking of The Society of Thoracic Surgeons database to Social Security data to examine survival after cardiac operations. Ann Thorac Surg 2011;92:32–7.
39. Jacobs JP, Edwards FH, Shahian DM, et al. Successful linking of The Society of Thoracic Surgeons Adult Cardiac Surgery Database to Centers for Medicare and Medicaid Services Medicare data. Ann Thorac Surg 2010;90:1150–6.
40. Morales DL, McClellan AJ, Jacobs JP. Empowering a database with national long-term data about mortality: the use of national death registries. Cardiol Young 2008;18(Suppl 2): 188–95.
41. Resnic FS, Welt FG. The public health hazards of risk avoidance associated with public reporting of risk-adjusted outcomes in coronary intervention. J Am Coll Cardiol 2009;53:825–30.
42. Resnic FS, Normand SL, Piemonte TC, et al. Improvement in mortality risk prediction after percutaneous coronary intervention through the addition of a "compassionate use" variable to the National Cardiovascular Data Registry CathPCI dataset: a study from the Massachusetts Angioplasty Registry. J Am Coll Cardiol 2011;57:904–11.
43. Shahian DM, Normand SL, Friedberg MW, Hutter MM, Pronovost PJ. Rating the raters: the inconsistent quality of health care performance measurement. Ann Surg 2016;264: 36–8.
44. Shahian DM, Blackstone EH, Edwards FH, et al. Cardiac surgery risk models: a position article. Ann Thorac Surg 2004;78:1868–77.
45. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. J Am Coll Surg 2013;217:336–46.
46. Krell RW, Hozain A, Kao LS, Dimick JB. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. JAMA Surg 2014;149:467–74.
47. Osborne NH, Ko CY, Upchurch GR Jr, Dimick JB. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. J Vasc Surg 2011;53:1–5.
48. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44:166–206.
49. Resnic FS, Majithia A. Rationalizing our report cards. JACC Cardiovasc Interv 2017;10:232–4.
50. Pepper JR, Coonar AS. High-risk surgery: the courage to fail. J R Soc Med 2015;108:44–6.
51. Shahian DM, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 1—conceptual framework and measure selection. Ann Thorac Surg 2007;83(4 Suppl):S3–12.
52. O'Brien SM, Shahian DM, Delong ER, et al. Quality measurement in adult cardiac surgery: part 2—statistical considerations in composite measure scoring and provider rating. Ann Thorac Surg 2007;83(4 Suppl):S13–26.
53. Rankin JS, Badhwar V, He X, et al. The Society of Thoracic Surgeons mitral valve repair/replacement plus coronary artery bypass grafting composite score: a report of The Society of Thoracic Surgeons quality measurement task force. Ann Thorac Surg 2017;103:1475–81.

54. Shahian DM, He X, Jacobs JP, et al. The STS AVR + CABG composite score: a report of the STS quality measurement task force. Ann Thorac Surg 2014;97:1604–9.
55. Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons isolated aortic valve replacement (AVR) composite score: a report of the STS quality measurement task force. Ann Thorac Surg 2012;94:2166–71.
56. Badhwar V, Rankin JS, He X, et al. The Society of Thoracic Surgeons mitral repair/replacement composite score: a report of the Society of Thoracic Surgeons quality measurement task force. Ann Thorac Surg 2016;101:2265–71.
57. Adams J. The reliability of provider profiling: a tutorial. Rand Health; 2009. Available at https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR653.pdf. Accessed May 12, 2017.
58. Huffman KM, Cohen ME, Ko CY, Hall BL. A comprehensive evaluation of statistical reliability in ACS NSQIP profiling models. Ann Surg 2015;261:1108–13.
59. Hall BL, Huffman KM, Hamilton BH, et al. Profiling individual surgeon performance using information from a high-quality clinical registry: opportunities and limitations. J Am Coll Surg 2015;221:901–13.
60. Hannan EL. The public reporting risk of performing high-risk procedures: perception or reality? JACC Cardiovasc Interv 2015;8:17–9.
61. Hannan EL, Zhong Y, Cozzens K, et al. The impact of excluding shock patients on hospital and physician risk-adjusted mortality rates for percutaneous coronary interventions: the implications for public reporting. JACC Cardiovasc Interv 2017;10:224–31.
62. Bangalore S, Guo Y, Xu J, et al. Rates of invasive management of cardiogenic shock in New York before and after exclusion from public reporting. JAMA Cardiol 2016;1:640–7.
63. McCabe JM, Waldo SW, Kennedy KF, Yeh RW. Treatment and outcomes of acute myocardial infarction complicated by shock after public reporting policy changes in New York. JAMA Cardiol 2016;1:648–54.
64. New York State Department of Health. Percutaneous coronary interventions (PCI) in New York state 2011–2013. Available at https://www.health.ny.gov/statistics/diseases/cardiovascular/docs/pci_2011-2013.pdf. Accessed May 12, 2017.
65. Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons composite measure of individual surgeon performance for adult cardiac surgery: a report of The Society of Thoracic Surgeons quality measurement task force. Ann Thorac Surg 2015;100:1315–25.
66. Westaby S. Publishing individual surgeons' death rates prompts risk averse behaviour. BMJ 2014;349:g5026.
67. Drozda JP Jr, Hagan EP, Mirro MJ, Peterson ED, Wright JS. ACCF 2008 health policy statement on principles for public reporting of physician performance data. J Am Coll Cardiol 2008;51:1993–2001.
68. Donelan K, Rogers RS, Eisenhauer A, Mort E, Agnihotri AK. Consumer comprehension of surgeon performance data for coronary bypass procedures. Ann Thorac Surg 2011;91:1400–5.
69. Shahian DM, Normand SL. Comparison of "risk-adjusted" hospital outcomes. Circulation 2008;117:1955–63.
70. Hibbard JH, Peters E, Slovic P, Finucane ML, Tusler M. Making health care quality reports easier to use. Jt Comm J Qual Improv 2001;27:591–604.