"ProvenCare℠"

A Provider-Driven Pay-for-Performance Program for Acute Episodic Cardiac Surgical Care

Alfred S. Casale, MD, Ronald A. Paulus, MD, Mark J. Selna, MD, Michael C. Doll, PA-C, Albert E. Bothe, Jr., MD, Karen E. McKinley, RN, Scott A. Berry, MS, Duane E. Davis, MD, Richard J. Gilfillan, MD, Bruce H. Hamory, MD, and Glenn D. Steele, Jr., MD

Objective: To test whether an integrated delivery system could successfully implement an evidence-based pay-for-performance program for coronary artery bypass graft (CABG) surgery.

Methods: The program consisted of 3 components: (1) establishing implementable best practices; (2) developing risk-based pricing; (3) establishing a mechanism for patient engagement. Surgeons reviewed all class I and IIa "2004 American Heart Association/American College of Cardiology Guidelines for CABG Surgery" and translated them into 40 verifiable behaviors. These were imbedded within a new ProvenCare℠ program and "hardwired" within the electronic health record system, including order sets, templates, and "time outs". Concurrently preoperative, inpatient, and postoperative care within 90 days was packaged into a fixed price. A Patient Compact was developed to highlight the importance of patient activation. All elective CABG patients treated between February 2, 2006 and February 2, 2007 were included (ProvenCare℠ Group) and compared with 137 patients treated in 2005 (Conventional Care Group).

Results: Initially, only 59% of patients received all 40 best practice components. At 3 months, program compliance reached 100%, but fell transiently to 86% over the next 3 months. Reliability subsequently increased to 100% and was sustained for the remainder of the study period. The overall trend in reliability was significant at $P = 0.001$. Thirty-day clinical outcomes showed improved trends (Table 1) but only the likelihood of discharge to home reached statistical significance. Length of stay decreased by 16% and mean hospital charges fell 5.2%.

Conclusion: A provider-driven pay-for-performance process for CABG, enabled by an electronic health record system, can reliably deliver evidence-based care, fundamentally alter reimbursement incentives, and may ultimately improve outcomes and reduce resource use.

(*Ann Surg* 2007;246: 613–623)

From the Geisinger Health System, Danville, Pennsylvania.
Reprints: Alfred S. Casale, MD, Richard and Marion Pearsall Heart Hospital, Geisinger Health System, 1000 East Mountain Boulevard, Wilkes-Barre, PA 18711. E-mail: ascasale@geisinger.edu.

ISSN: 0003-4932/07/24604-0613
DOI: 10.1097/SLA.0b013e318155a996

Healthcare delivery in the United States faces significant quality and cost problems. Medical care is often inappropriate when judged against accepted standards with numerous examples of excess utilization and conversely, appropriately indicated care is frequently not provided.[1] This inconsistency leads to wide, unexplained variation in rates of procedures, expenditures, and outcomes.[2] Landmark publications by the Institute of Medicine and the Rand Corporation[3–5] have focused increased professional and public attention on these issues. Nevertheless, healthcare providers continue to be paid for units of care delivered independent of quality or results achieved. Poor outcomes, such as postoperative complications that require reoperation, often result in more payment.

Care reliability is inconsistent. Best practice guidelines are sometimes based on equivocal evidence, and are often ignored or poorly applied.[6] Translation of even the best guidelines into actual behavior is difficult and slow-paced. The fragmentation of our delivery systems[7] and the influence of diverse and often opposing economic factors can overwhelm the influence of science and well-meaning intentions in determining acceptance and dissemination of best practices.[8]

Strategies to improve this system have included mandates from regulators, federal and state agencies, and payers. Public reports of outcome measures are often derived from administrative databases and have typically had only modest influence on physician and patient behavior.[9] Oversight from medical societies, state licensing agencies, hospital medical staff offices, and specialty societies is not sufficient and is based largely on identifying and reacting to care failures rather than proactively focusing on failure avoidance.

Several innovative provider-initiated programs [eg, Department of Veterans Affairs National Surgical Quality Improvement Program (NSQIP), the American College of Surgeons' NSQIP, the Society of Thoracic Surgeons' (STS) National Database][10–12] have been developed to collect data regarding the outcomes of surgical care and analyze it in a manner allowing comparison of a subunit's performance to overall observed results. These important "translational outcomes research"[13] projects have dramatically improved our capacity to compare results of care across time and platforms,

TABLE 1. Thirty-day Clinical Outcomes

	Conventional Care Group 2005 (n = 137)	%	ProvenCare Group (n = 117)	%	*P*	Statistical Test
Readmission within 30 days	9	6.6	7	6.0	0.99	Fisher exact
Discharged home	111	81	106	91	0.03	Fisher exact
Patients with any complication	53	39	41	35	0.55	Chi-square
Patients receiving blood products	32	23	19	16	0.17	Chi-square
Readmitted to ICU	4	2.8	1	1	0.38	Fisher exact
Pulmonary complications	10	7	3	2.5	0.15	Fisher exact
Operative mortality	2	1.4	0	0	0.50	Fisher's exact
Atrial fibrillation	31	23	30	26	0.58	Chi-square
Deep sternal wound infection	1	0.7	1	0.9	1.00	Fisher's exact

and they are associated with real improvement in outcomes. By providing algorithms that calculate expected rates of adverse outcomes, they minimize the confounding influence of pretreatment severity of illness and comorbidity on outcomes and "level the playing field" among those providing the care. However, it is important to recognize that these programs have not tested actual elements of care, as do conventional randomized clinical trials of specific interventions. Nor have they been designed to encourage use of any specific best practices (although some, such as the Northern New England Cardiovascular Disease Study Group,[14] do include robust mechanisms to share with all participants those practices that high performing units believe are responsible for their success). This moves from the worthy, but limited, goal of reporting outcomes and influencing them by the Hawthorne effect and other indirect means, to actually disseminating practices derived from the insights of participants for voluntary adoption by underperforming units.

Governmental adoption of a standardized system for reporting performance on the delivery of certain process such as the Center for Medicare and Medicaid Services' Surgical Complications Improvement Program[15] compels (and may eventually reward) participation in a reporting program, but even this and other so-called pay-for-participation programs do not fully apply pay-for-performance (P4P) principles.

P4P emerged in the late 1990s and has become increasingly prominent in quality improvement programs nationally.[16] In its essence, this involves rewarding improved care. Incentives have included increased financial reimbursement, bonus payments, diminished administrative burden, or inclusion in favorable marketing or promotional campaigns. Ideally, these benefits would be made available to providers (individuals, groups, or systems) who actually achieve improved patient satisfaction and/or better clinical outcomes. However, difficulty in establishing fair, accurate, risk-adjusted measures of meaningful clinical outcomes has often led to the substitution of surrogates for these direct measures of quality. Among these substitutes expected to correlate with or lead to improved outcomes are adherence to adopted guidelines, application of specified care processes and provision of certain minimal infrastructure [eg, intensive care unit staff, electronic health record systems (EHR)]. To date, most P4P programs have been payer-designed and offered (or mandated) to primary care physicians for preventative care in outpatient settings or for management of chronic diseases. In addition, when P4P involves payment to physicians, it has usually represented only a small component of a provider's total compensation. Improvement programs using P4P principles have not targeted complex acute episodic surgical care.[17]

In 2005, Geisinger's executive leadership and Board of Directors challenged clinical leaders to explore innovative, provider-driven quality improvement programs based on P4P principles. We were encouraged to consider major redesign of system-wide processes of care (for both acute episodic and chronic conditions) to promote reliable delivery of consensus-derived or evidence-based best practices. Fundamental change to the piecework reimbursement paradigm, starting with Geisinger's own health plan, was an essential component of the challenge.

MATERIALS AND METHODS

Geisinger Health System ("Geisinger") is a large integrated healthcare delivery system located in central and northeastern Pennsylvania serving a population of 2.6 million in 41 of the state's 67 counties. The system's physician group practice (Geisinger Clinic) has over 650 physicians. Approximately 200 are primary care physicians in 38 community practice locations; the balance consists of specialists. Geisinger has 3 tertiary/quaternary medical centers and a health insurance plan, Geisinger Health Plan (GHP), with approximately 210,000 members. Training of nearly 500 medical students, residents, and fellows, as well as other clinical training programs, reflect the system's commitment to education. Basic science and translational research—including the design and dissemination of new models of healthcare delivery—focus the system's clinical mission. Of particular note, early adoption and system-wide deployment of an EHR (beginning in 1995) has placed Geisinger in the forefront of health informatics—a key to successful large-scale clinical process redesign.[18,19]

Geisinger's Cardiothoracic Surgery Department consists of 8 cardiac surgeons: 4 at Geisinger Medical Center in Danville, PA, and 2 at each of the system's Wilkes-Barre, PA locations, Geisinger Wyoming Valley Medical Center and Geisinger South Wilkes-Barre. Up to 20% of total compen-

sation for physicians in the Geisinger Clinic is predicated on achievement of predefined goals, including measures of clinical care quality.

The ProvenCare Model: CABG

For several reasons, coronary artery bypass graft (CABG) emerged as an appropriate first target for the development of a P4P program focused on acute elective surgical care. First, CABG is performed at all 3 Geisinger hospitals and all the cardiac surgeons are employed by the system. These surgeons function cohesively (despite geography), are actively involved in quality improvement programs, and are supported by a team of experienced physician extenders. The team also has a history of excellent outcomes as measured by the Pennsylvania Health Care Cost Containment Council[20] and the STS database.[21] Nevertheless, there were still idiosyncratic processes of care within this group of surgeons.

Additionally, guidelines for CABG had been recently updated by the American Heart Association (AHA) and the American College of Cardiology (ACC) in the "AHA/ACC 2004 Guideline Update for CABG Surgery."[22] National and regional benchmarks were available and data was already collected in near real-time with the STS database fully deployed at all Geisinger sites.

On the basis of historical performance and case volume, and predicted rates of occurrence of adverse events, we felt the impact of reengineered care processes could be measured over a reasonably short time frame. The size and complexity of the clinical team caring for CABG patients made the prospect of reengineering workflow processes sufficient in scope to be a meaningful test of ProvenCare principles. Finally, the financial impact of CABG to the system is significant, ensuring engagement of financial and administrative leaders in the planning and implementation process.

At the start of our planning, 4 workgroups and a steering committee were convened. The first group, the Best Practice Workgroup, was comprised of all Geisinger cardiac surgeons. They discussed the project's rationale, established individual and group commitment, and identified areas of concern. Initially, there was uniform fear by this physician group that enforcement of best practices might force surgeons to rely on "cookbook medicine". They were also concerned that current best practice guidelines might not evolve in a timely fashion with new scientific findings. To address these issues, the group agreed to critically examine the AHA/ACC Guidelines. All 12 class I and 8 class IIa guidelines (Table 2) were divided among the surgeons (with particular emphasis to assigning a specific guideline to the surgeon most skeptical of its value). Surgeons then went to the primary sources for each guideline, reviewed the referenced papers, and presented their recommendations to their peers. Attesting to the quality of the AHA/ACC committee's work, all surgeons ultimately agreed that each of the reviewed guidelines was appropriate. The generality of the published guidelines prevented their direct application in actual care processes. Therefore, each of the guidelines was translated into 1 or more verifiable, actionable care processes with unequivocal definitions established using STS database or other specialty standards. Each care process change was designed to be consistent with best practices, be practical and measurable, and be accountable to a specific individual. The 20 adopted guidelines yielded 40 separate elements of care (Table 3).

TABLE 2. ACC/AHA Recommendations

ACC/AHA Class I Recommendations
Preoperative antibiotics
Preoperative carotid evaluation
Aspirin use for graft patency
Identify atherosclerotic ascending aorta and adapt operative strategy
Aggressive debridement and flap coverage for deep sternal wound infections
Perioperative beta blockers (or amiodarone) to reduce atrial fibrillation
Statin use
Smoking cessation education and pharmacotherapy
Cardiac rehab
Withholding of clopidogrel for 5 d preoperative if possible
Left internal mammary artery as graft for the LAD artery
ACC/AHA Class II Recommendations
Preoperative use of a CABG operative mortality risk model
Anticoagulation for recurrent/persistent postoperative Afib
Anticoagulation for postoperative anteroapical MI with persistent wall motion abnormality
Carotid endarterectomy for carotid stenosis that is symptomatic or >80%
Intraaortic counterpulsation for low LV ejection fraction
Blood cardioplegia
Delay operation for patients with recent inferior MI with significant RV involvement
Tight perioperative glucose control

Professionals from Geisinger's Department of Clinical Effectiveness documented existing processes of care and patient flows. This revealed considerable care variation, despite having a consistent, experienced group of physician extenders, nurses, and operating room staff. This work group recommended that the project initially include only elective CABG patients because this offered sufficient time to ensure that all 40 elements could be met on each patient (as compared with urgent patients who usually undergo surgery within 24 to 36 hours and, by definition, may be ineligible for certain care components).

Representatives of all units involved in CABG patient care then established a new idealized workflow incorporating the 40 best practice elements and developed a mechanism for documenting and tracking each element along with assigning it to an accountable individual. Information technology professionals worked with clinical staff to develop EHR-based tools (checklists, default documentation templates, automated order sets, etc.) required to "hardwire" the reliable delivery of each best practice element. The resulting EHR workflows required that clinicians either comply with best practice or document the rationale for not using a specific best practice element—a step designed to protect physician prerogative and provide ongoing feedback on guideline appropriateness.

A key component to any provider-initiated P4P program is payer acceptance of its terms. Geisinger chose to

TABLE 3.

1. Preadmission documentation:
   a. ACC/AHA indication
   b. Screening for and consultation re: IMI/RV involvement
   c. Treatment options and patient preference
   d. Need for warfarin (anterior MI or wall motion abnormality)
   e. Current user of clopidogrel or warfarin?
   f. Screening for stroke risk
   g. Carotid doppler (if the test is indicated)
   h. Vascular surgery consultation (if indicated)
   i. Ejection fraction
   j. Screening for need to use intra-aortic balloon pump (IABP)
   k. Screening using epiaortic echo (as indicated)
   l. Patient withheld clopidogrel/warfarin for 5 d preoperatively?
2. Operative documentation:
   a. Patient received correct dosing of beta-blocker (preoperative)
   b. Correct use of intra-aortic balloon pump (preoperative → postoperative)
   c. Preoperative antibiotic (within 60 min of incision; Vancomycin within 120 min)
   d. Blood cardioplegia (on-pump patients)
   e. Epiaortic echo of the ascending aorta and the peer consult
   f. Intraoperative hyperglycemia screening
   g. Correct insulin management (as indicated; per protocol)
   h. Use of LIMA for LAD grafting
3. Postoperative patient documentation:
   a. Anteroapical MI within prior 7 d: postoperative echo
   b. Monitoring for atrial fibrillation for >48 h
   c. Anticoagulation therapy (as indicated)
   d. Antibiotic administered (postoperative for 24–48 h)
   e. Aspirin (6 hours postoperative or 24 h postoperative)
   f. Beta-blocker (within 24 h postoperative)
   g. Statin administered (postoperative)
   h. Surgical debridement and revascularization of any sternal wound infection
   i. Plastic surgery consult regarding ongoing management of sternal wound
   j. Tobacco screening and counseling
4. Discharge documentation:
   a. Referral to cardiac rehabilitation
   b. Discharge medications (eg, beta-blocker)
   c. Discharge medication: aspirin
   d. Discharge medication: statin
5. Post-Discharge documentation:
   a. Patient correctly taking beta-blocker?
   b. Patient correctly taking aspirin?
   c. Patient correctly taking statin?
   d. Patient correctly administering anticoagulant?
   e. Did patient resume smoking?
   f. Patient enrolled in cardiac rehabilitation?

pursue a trial of the ProvenCare program initially within its own insurance company. Clinicians, finance staff, administrators, and GHP representatives formed the Financial Workgroup. This group explored the attitudes of commercial purchasers of GHP insurance products toward outcomes-based reimbursement in general and to P4P programs that involved transfer of financial risk from payers to the provider. Initially it was believed that a provider guarantee of adherence to all best practice elements would motivate purchasers. However, the high number of best practice elements (40) and complexity of the clinical content when explained to nonclinicians limited interest in this approach. Further, most purchasers believed that best practices were already being universally applied by all providers or they were skeptical, believing that best practice elements of care would be designed to game program success rather than to achieve a fundamental clinical and cost savings impact.

By contrast, we found that purchasers highly valued financial predictability. They were averse to open-ended risk for the unknown but high costs of postoperative complications and other treatment failures. Ultimately our purchasers were attracted to a package price for elective CABG that included preoperative evaluation and work-up, all hospital and professional fees, all routine postdischarge care (eg, smoking cessation counseling, cardiac rehabilitation), and management of all related complications. Our program's package of postoperative care includes all follow-up care and all rehospitalizations (at a Geisinger facility) for any related postoperative complication occurring within 90 days of surgery. However, not all complications can be eliminated, even with an ideal care process. Therefore, the case rate for ProvenCare patients also included approximately 50% of the mean cost of all postoperative readmissions related to complications as experienced in a 2-year historical financial comparison group. This essentially defined the magnitude of the direct financial reward available to Geisinger for optimal clinical outcomes; the financial risk of managing increased or unchanged rates of complications was transferred wholly to the clinical enterprise. This "warranty," based on our confidence in our ability to deliver reliable care and thereby improve already excellent outcomes, captured the attention of purchasers. High volume purchasers would be better off, while acknowledging clinical reality and providing a financial incentive to the system for ongoing care process improvement.

Successful adherence to ProvenCare processes was included as one component of surgeons' individual compensation, but we avoided tying clinical outcome directly to physician compensation to minimize any hesitance to care for high-risk patients.

Less than optimal communication with patients and among caregivers can lead to lack of coordination and poor continuity of care. Risk is heightened when patients are simply passive care recipients rather than actively engaged as partners in their treatment. As one mechanism of engagement, our Patient Activation Workgroup developed a joint Geisinger/patient agreement, the "Patient Compact" (Table 4) that codified the mutual commitment of the system, the patient, and their family to the redesigned best practice processes.

A Steering Committee was established comprising clinical and executive leaders from cardiac surgery, the health system, and the system's health insurance plan (GHP). It supported and complemented the workgroups, ensuring that the ProvenCare program was on track, targeting specific

TABLE 4. ProvenCare: CABG Patient Compact

My Role in Proven Heart Care

The Geisinger heart surgery team has your health and safety as its chief concern. That is why we established the *ProvenCare℠* Heart Program. The *ProvenCare* Heart Program includes all of the care steps necessary to ensure the highest quality care before, during and after your heart operation. Your *active* participation is one of the most important parts of the Geisinger *ProvenCare* Heart Program. Medical research has shown that the more involved you are in your own care - and the stronger the partnership between you and your caregivers - the better your results will be. Even though the Geisinger heart surgery team *always* strives to provide all of the elements of the ProvenCare Heart Program, you will get the best result when you, your family and your Geisinger heart surgery team are all *active* partners in your care.

COMMITMENT TO COMMUNICATE AS A TEAM

- I will alert my heart surgery team when I don't understand something, when anything worries me, or if anything unexpected occurs, knowing that my heart surgery team will work with me until I am satisfied.
- I will discuss all of my current medications, non-prescription products, vitamins or herbs as well as all of my current and past medical problems, recognizing how important this information is in guiding my care and making me safer.

COMMITMENT TO INVOLVE MY FAMILY AND LOVED ONES

- I will have a trusted family member or loved-one present with me during my hospitalization and clinic visits - to help support me during my care.
- I will work with my heart surgery team to develop a sensible plan for my transition from the hospital back to my home.

COMMITMENT TO COMPLETE IMPORTANT CARE STEPS

- I will alert my heart surgery team before I stop or start any of my medications so that we can discuss how any change might impact my care.
- I will work with my heart surgery team to develop a sensible schedule for my after-surgery care, follow-up visits and rehabilitation.

COMMITMENT TO IMPROVED HEARTH AND PREVENTION

- I will complete a cardiac rehabilitation program, understanding that it will give me a better, quicker and more lasting recovery.
- I will work with my heart surgery team to stop my use of any tobacco products - forever.
- I will discuss with my heart surgery team the important role that life-long nutrition, weight management, exercise and medications play in keeping my heart healthy.

I realize that my decisions and behavior have a significant positive impact on my long-term health. Because I want to become and stay healthy, I fully accept my role as a partner in the *ProvenCare* Heart Program.

Sincerely,

Insert Name _____ Date _____

goals, and that ProvenCare methods and results were communicated to the broader Geisinger community. All-or-none assessment[23] was adopted as the measure of performance in the ProvenCare program.

Testing of the ProvenCare processes, tracking mechanisms, and financial adjudication began in January 2006. The full program was implemented a month later—on February 2, 2006.

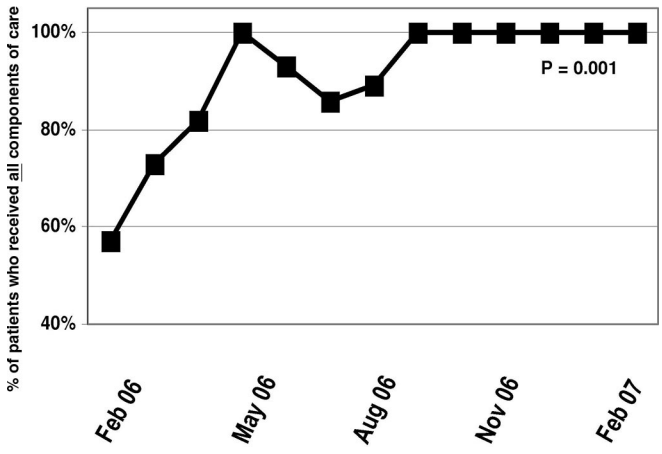

FIGURE 1. Coronary artery bypass: Reliability.

Geisinger's Institutional Review Board approved the analysis of patient de-identified, aggregated data for the patients undergoing elective CABG during the ProvenCare program (the ProvenCare Group), and those patients who underwent elective CABG in 2005 (the Conventional Care Group). The Institutional Review Board waived the requirement for informed consent since only data already routinely collected for clinical use would be accessed.

RESULTS

Before initiation of the ProvenCare program, 59% of all elective CABG patients were found to have received all 40 elements of the ProvenCare process (Fig. 1). Within 3 months, performance rose to 100%. During the second 3 months, reliability transiently dropped to 86%; however, by the end of the first 6 months, reliability rose and thereafter remained at 100% throughout the study period. The overall trend for ProvenCare process adherence was significant at $P = 0.001$, using the Cochran-Armitage Trend test.

Clinical Outcomes

All 117 elective CABG patients treated within the Geisinger system in a 1-year period (from February 2, 2006 through February 2, 2007) were managed under this program. There were an additional 290 nonelective CABG patients

who did not undergo systematic application of the ProvenCare processes of care and were not included in the outcome analysis. Nevertheless, these patients usually had most elements included by surgeon choice.

The cohort of ProvenCare patients was compared with the 137 elective CABG patients (operated upon in 2005) whose care was provided before the program's initiation.

The preoperative and operative characteristics used in the STS outcome predicting algorithms were similar in both cohorts except that left main coronary stenosis greater than 50% occurred more frequently in the ProvenCare Group (23% vs. 12%), in keeping with the trend toward expanded use of percutaneous catheter intervention in patients with lesser degrees of coronary artery disease (Table 5).

Importantly, the observed rates of adverse events were already lower than predicted by the STS algorithms in the 2005 (Conventional Care) group. These excellent clinical outcomes before care process reengineering established a high bar for further clinical improvement.

TABLE 5. Comparison of Conventional Care Group to ProvenCare Group: Preoperative and Operative Characteristics

	Conventional Care (2005) n = 137		ProvenCare^{SM} (2006) n = 117			
	n or Mean	% or SD	n or Mean	% or SD	*P*	Test Type
Demographics						
Age (mean in yr)	66	10	66	11	0.89	Equal T test
Male gender	100	73.0%	92	76.8%	0.30	χ^2
White race	136	99.3%	115	98.3%	0.60	Fisher exact
Weight (mean in kg)	89	22	89	20	0.99	Equal T test
Height (mean in cm)	171	11	172	10	0.36	Equal T test
BSA (mean in m^2)	2.05	0.29	2.06	0.26	0.80	Equal T test
BMI	30.4	6.2	30.1	5.7	0.60	Equal T test
Risk factors						
Diabetes	48	35.0%	44	37.6%	0.63	χ^2
Hypertension	106	77.4%	87	74.4%	0.58	χ^2
Renal failure	8	5.8%	7	6.0%	1.00	Fisher exact
Dialysis	4	2.9%	4	3.4%	1.00	Fisher exact
Hypercholesterolemia	111	81.0%	90	76.9%	0.42	χ^2
Immunosuppressive therapy	0	0.0%	0	0.0%	NA	
Cerebrovascular disease	15	10.9%	11	9.4%	0.84	Fisher exact
Peripheral vascular disease	19	13.9%	17	14.5%	1.00	Fisher exact
Smoker	81	59.1%	64	54.7%	0.48	χ^2
Current smoker	23	16.8%	23	19.7%	0.55	χ^2
Preoperative creatinine (median in mg %)	1.00		1.00		0.54	Wilcoxon
Stroke	10	7.3%	11	9.4%	0.65	Fisher exact
Previous CV Interventions						
Prior CABG	4	2.9%	5	4.3%	0.74	Fisher exact
Preoperative cardiac status						
MI	37	27.0%	29	24.8%	0.69	χ^2
CHF	8	5.8%	4	3.4%	0.55	Fisher exact
Class NYH:						
I	10	7.4%	8	6.8%	0.89	χ^2
II	60	43.8%	49	41.9%		
III	66	48.2%	58	49.6%		
IV	1	0.7%	2	1.7%		
Arrhythmia	12	8.8%	6	5.1%	0.33	Fisher exact
Hemodynamics and cath						
Left Main Disease ≥50%	17	12.4%	27	23.1%	0.031	Fisher exact
Ejection fraction (mean)	52.3	13.1	54.6	10.6	0.12	Equal T test
Mitral insufficiency	24	17.5%	13	11.1%	0.16	Fisher exact
Number of diseased vessels = 3	106	77.4%	88	75.2%	0.77	Fisher exact
Aortic stenosis	2	1.5%	4	3.4%	0.42	Fisher exact
Operative						
IABP	4	2.9%	2	1.7%	0.69	Fisher exact

Comparison of clinical outcomes between the 2 groups shows that most adverse events occurred less often in the ProvenCare Group than in the Conventional Care Group, though only the likelihood of being discharged to home was statistically significantly different (Table 6). There was no deterioration in outcomes as some have feared might occur with the rigid application of guidelines.[24] The concern that limitation of physician autonomy would be dangerous was not supported.

Financial Outcomes

Although median postoperative length of stay was the same at 4 days for both groups, average total length of stay fell 16% from 6.3 days in the Conventional Care Group to 5.3 days in the ProvenCare Group and was reflected in a 5% reduction in hospital charges. The 30-day readmission rate fell 15.5%, from 7.1% in the period used to establish a financial baseline to 6% in the ProvenCare Group.

On the basis of the level of reliability with which care processes were delivered, individual surgeons earned 100% of the incentive compensation available for quality improvement work in 2006.

DISCUSSION

This project demonstrates that a large integrated healthcare delivery system, enabled by an EHR, can successfully reengineer complicated care processes to reliably deliver consensus-derived and evidence-based best practices system-wide, while fundamentally altering the reimbursement paradigm.

We found the cultural milieu established by the Geisinger Board of Directors and executive leadership team essential to our successful redesign of clinical systems to support fully reliable care. In addition, the active engagement of all clinical, administrative, and financial stakeholders led by both clinical champions and clinical effectiveness professionals proved important. Precise and thorough definition of workflows and their reengineering to eliminate idiosyncratic variation and incorporate carefully chosen best practices created the mechanism that motivated clinicians to improve reliability. Importantly, in this model a surgeon's prerogative to individualize patient care was maintained while assuring that care delivery is not idiosyncratic or oblivious of best practice.

The advisability of direct financial incentive to individual providers for improved quality has been questioned. Damaging effects on professionalism and job satisfaction have been predicted. In contrast, incentives to organizations, systems and groups to support, justify and validate quality improvement work may be a superior model for P4P with complex clinical situations involving different types of providers.[25] Including ProvenCare process adherence (rather than specific outcome targets as one component in a broader preexisting incentive plan) helped motivate the involved physicians while avoiding any inducement to inappropriately limit care.

Quintessential application of P4P principles would directly reward individual providers financially for improved

TABLE 6. Comparison of Conventional Care Group to ProvenCare: Outcomes

	Conventional Care (2005) n = 137		ProvenCare℠ (2006) n = 117			
	n	%	n	%	*P*-Value	Test Type
Blood Products Used	32	23.4	19	16.2	0.17	χ^2
Reintubated during hospital stay	4	2.9	1	0.9	0.38	Fisher exact
Total ventilation hours (median)	8		7.5		0.64	Wilcoxon
Operative complication	8	5.8	5	4.3	0.78	Fisher exact
Re-op for bleeding	5	3.6	3	2.6	0.73	Fisher exact
Re-op for other cardiac problem	1	0.7	1	0.9	1.00	Fisher exact
Re-op for other noncardiac problem	1	0.7	1	0.9	1.00	Fisher exact
Perioperative MI	1	0.7	1	0.9	1.00	Fisher exact
Infection: Sternum - Deep	1	0.7	1	0.9	1.00	Fisher exact
Neurologic complication	2	1.5	1	0.9	1.00	Fisher exact
Pulmonary complication	10	7.3	3	2.6	0.15	Fisher exact
Prolonged ventilation	8	5.8	3	2.6	0.23	Fisher exact
Pneumonia	1	0.7	0	0.0	1.00	Fisher exact
Pulmonary embolism	1	0.7	0	0.0	1.00	Fisher exact
Renal failure	0	0.0	1	0.9	0.46	Fisher exact
Atrial fibrillation	31	22.6	30	25.6	0.58	χ^2
Any complication (by STS database definition)	53	39.0	41	35.0	0.55	χ^2
Postoperative length of stay (median)	4		4		0.25	Wilcoxon
Readmission to ICU	4	2.9	1	0.9	0.38	Fisher exact
Operative mortality	2	1.5	0	0.0	0.50	Fisher exact
Discharge location = Home	111	81.0	106	90.6	0.033	Fisher exact
Readmit <30 days from procedure	9	6.6	7	6.0	0.99	Fisher exact

outcomes. In acute episodic surgical care, when dealing with any individual patient, this pure form of P4P is unworkable. It is axiomatic that precise outcomes for an individual patient are unpredictable. Even if surrogates for outcome, such as adherence to promised processes are adopted, the timely and accurate adjudication of a claim for payment in a given patient is difficult. The enormity of the data tracking needed, the imperfection in that data, and the possibility of "gaming" limit the utility of such an approach when substantial payments are at risk.

ProvenCare avoids these pitfalls by having the clinical enterprise at Geisinger accept the risk for complications related to the index operation for 90 days in exchange for the increment in payment for each CABG patient (equivalent to one-half of the average cost of caring for complications in past experience). It is then incumbent on Geisinger's clinicians to strive to deliver optimal care, both because it is the right thing to do and because the system and they will benefit by decreasing complications and improving patient outcomes. The ProvenCare processes of care operationalize this effort by enabling reliable delivery of best practices.

The goal of this project was not to validate any individual best practice guideline that was adopted. The AHA/ACC Guidelines were selected as a basis of the elements of ProvenCare for CABG because they were the best reflection of specialty consensus available at the time. Their adoption allowed us to focus on the program's main goal: establishing a mechanism to enable a complicated clinical enterprise to deliver chosen elements of care with total reliability, rather than to quibble over which practices were in fact "best."

The observed overall outcome trend resulting from the reliable delivery of all of the AHA/ACC Class I and IIa guidelines was positive. The ProvenCare methodology established a mechanism that allows for the reliable delivery of a set of best practices that can be added to, subtracted from, or otherwise modified as knowledge evolves. Potentially, the ProvenCare process could itself serve as a platform to test the importance of changes in practice by randomizing patients to groups differing only in the reliable provision of 1 or more putative best practice elements.

Limitations and Caveats

Although the power of changing workflows to ensure use of best practices is significant, the ability to scale up (or down) requires careful consideration. Geisinger's resources used to develop the ProvenCare process have been substantial. Although most of the processes that were developed are now being applied to redesigning other episodic care services, a critical mass of engagement by many parties was necessary to implement and sustain the project. Further, expansion to units larger than ours, to systems with more hospitals or those without a provider-owned insurance company, would add significant logistical complications that have not yet been assessed.

Whether nonintegrated delivery system provider teams can develop similar process change is unclear, particularly if there is no structural alignment of hospital and physician financial incentives. Despite the clinical transformation success, to date ProvenCare has had little effect on market demand from employers or purchasers of healthcare. Feedback suggests that this is at least in part because CABG impacts only a small fraction of an employer's total healthcare spending. Demand is expected to increase as additional clinical areas deploy ProvenCare and these are offered as part of a larger package within the Geisinger insurance company's products.

Next Steps

Throughout the Geisinger system, all elective CABG patients are actively managed within the current ProvenCare program. Tracking of clinical and financial outcomes is ongoing. The original workgroups have now reformed to examine the ProvenCare elements and further refine them. New guidelines from the STS,[26] the Institute for Healthcare Improvement,[27] the Agency for Healthcare Research and Quality,[28] and ACS NSQIP will be examined. Expansion to nonelective CABG and to valve and combined valve-CABG patients is underway. ProvenCare programs for both percutaneous catheter intervention and for management of acute myocardial infarction are being designed, as are projects focused on total joint replacement, cataract extraction and the use of biologic specialty treatments, such as erythropoietin.

ACKNOWLEDGMENTS

The authors thank the surgeons and physician assistants of the Geisinger Heart Institute's cardiac surgery service. Their diligent work and commitment to their patients' well being has been instrumental in establishing ProvenCare. The cooperation and input from the nursing and anesthesia staff of Geisinger's acute care hospitals are gratefully acknowledged. The efforts of the Geisinger Health Plan staff, the Department of Clinical Effectiveness, and the Center for Health Care Research are appreciated. Special thanks are extended to Craig Wood for providing statistical support, to Melinda Reed for managing the STS database and especially to Sandra Buckley for her expert assistance in preparing this manuscript.

REFERENCES

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348:2635–2645.
2. Coronary Artery Bypass Grafting Surgery in Redding, California. Regional variations in rates of coronary artery bypass grafting [Internet]. Dartmouth atlas of health care: studies of surgical variation. Available at: www.dartmouthatlas.org/atlases/Cardiac_report_2005.pdf.
3. Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err is Human: Building a Safer Health System*. Washington: National Academy Press; 2000.
4. Committee on Quality Health Care in America, Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington: National Academy Press; 2001.
5. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q*. 1998;76:517–563.
6. Steinbrook R. Guidance for guidelines. *N Engl J Med*. 2007;356:331–333.
7. Gauthier A, Schoenbaum SC, Weinbaum I. *Toward a High Performance Health System for the United States* [Internet]. New York: The Commonwealth Fund; 2006. Available at: www.cmwf.org/usr_doc/Gauthier_highperformance_909. pdf.
8. Balas E, Boren S. Managing clinical knowledge for health care improvement. In: *Yearbook of Medical Informatics.* Bethesda, MD: National Library of Medicine; 2000:65:–70.

9. Epstein A. Do cardiac surgery report cards reduce mortality? Assessing the evidence. *Med Care Res Rev*. 2006;63:403–426.
10. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. *Ann Surg*. 1998;228: 491–507.
11. Jones RS, Brown C, Opelka F. Surgeon compensation: "pay for performance," the American College of Surgeons National Surgical Quality Improvement Program, the Surgical Care Improvement Program, and other considerations. *Surgery*. 2005;138:829–836.
12. Grover FL, Shroyer AL, Hammermeister K, et al. A decade's experience with quality improvement in cardiac surgery using the Veterans Affairs and Society of Thoracic Surgeons national databases. *Ann Surg*. 2001; 234:464–472.
13. Warshaw AL, Flum DR, Sarr MG, et al. Introduction: the rise of translational outcomes research. *Surgery*. 2005;138:813–814.
14. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery: the Northern New England Cardiovascular Disease Study Group. *JAMA*. 1996;275:841–846.
15. Surgical Care Improvement Project Information. [Internet]. Washington: Medicare Quality Improvement Community; 2007. Available at: www.medqic.org/dcs/ContentServer?cid=1122904930422&pagename=Medqic%2FContent%2FParentShellTemplate&parentName=Topic&c=MQParentsSCIPCMS.
16. Board on Health Care Services, Institute of Medicine. *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington: National Academy Press; 2007.
17. Rosenthal MB, Frank RG, Li Z, et al. Early experience with pay-for-performance: from concept to practice. *JAMA*. 2005;294:1788–1793.
18. Hospitals and Health Networks [Internet Webpage]. "2006 Most Wired Hospitals and Health Systems"; 2006. Available at: www.hhnmag.com/hhnmag_app/jsp/articledisplay.jsp?dcrpath=HHNMAG/PubsNewsArticle/data/2006July/0607HHN_FEA_CoverStory_WinnersList&domain=HHNMAG.
19. 2007 Verispan IHN 100® [Internet]. Yardley, PA: Verispan; 2007. Available at: www.verispan.com/images/content/2007IHN100.pdf.
20. Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery [Internet]. Harrisburg, PA: Pennsylvania Health Care Cost Containment Council. Available at: www.phc4. org/reports/cardiaccare.htm.
21. Duke Clinical Research Institute. *Data Analyses of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database*. Chicago, IL: Duke Clinical Research Institute; 2000–2006.
22. Eagle KA, Guyton RA, Davidoff R, et al. American College of Cardiology. American Heart Association. ACC/AHA 2004 guideline update for coronary artery bypass graft surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J Am Coll Cardiol*. 2004;44:e213–e310.
23. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA*. 2006;295:1168–1170.
24. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ*. 1999;318: 527–530.
25. Galvin R. "A deficiency of will and ambition": a conversation with Donald Berwick. *Health Aff (Millwood)*. 2005;W5-1–W5-9.
26. Practice Guidelines [Internet]. Chicago, IL: The Society of Thoracic Surgeons; 2007. Available at: www.sts.org/sections/resources/practiceguidelines/.
27. IHI Launches National Campaign to Reduce Medical Harm in US Hospitals [Internet]. Cambridge, United Kingdom: The Institute for Healthcare Improvement; 2006. Available at: www.ihi.org/NR/rdonlyres/A96E0639-A979-44F6-A0FF-20329E492F44/4230/5MillionLivespressreleaseupdatedDec26FINALFINAL.pdf.
28. Agency for Healthcare Quality and Research [Internet]. Washington: US Department of Health and Human Services; 2007. Available at: www.ahrq.gov/.

Discussions

DR. R. SCOTT JONES (CHARLOTTESVILLE, VIRGINIA): If one looks at the processes that Dr. Casale and his colleagues employed to develop this work, it is, I believe, a generalizable model that can certainly help other institutions improve the quality of their care and their organization. There are several points that I think are particularly interesting about this work.

First of all, it was done in an integrated system of care under one management system, and it also included the insurance company along with the providers. There are relatively few opportunities in health care today where we can work with this particular model. So, in a sense, it is really a great test.

I was particularly impressed with the development of the quality improvement. The way they translated the guideline information into clinical pathways I think shows that this is the way it should be done. It was done with consensus and high reliance on the best evidence from all of these practices and I would urge any folks doing this kind of work also to do that.

An important point about the processes of care is that they have now documented that with such leadership and organization they can achieve 100% compliance with these evidence-based practices. And that is commendable.

Another point is that they have a complete electronic health record, and they use it optimally to decrease a lot of manual work in terms of collecting and analyzing their data. And I think that is also something we should strive for.

I would like to make a comment about the outcome statistics. When I read this paper and looked at it on the first pass, what I saw in the statistical analysis was that nothing is significant, and I thought about this a little bit more. I first focused on the raw data, which tells me, in this particular circumstance, more than your statistical method for the following reasons: First, you were already using the STS database for a period of 10 years, so you were actually looking at a system of care that was always under scrutiny with outcome data. Second, the numbers of patients in this study with this kind of statistical analysis are far too small to even have much meaning with your statistical methods. So I hope you continue to do this work, and with more patients. With each increment in patients you can get, you can have more confidence in your statistical methods.

The one thing that I would have liked to see done a little bit better in this particular study is how you address pay for performance. Now, that is a difficult challenge, and many folks have addressed this. But I would have expected a little bit more information about how you intend to manage the income, and perhaps even the expenses, from this particular project.

In other words, your system will receive income attributable to this project. I assume that income will be collected and managed by the insurance component of your organiza-

tion. I think it would be important for you to tell us, or inform us in some way, what methodology or process you can employ to determine how much of that income remains in the insurance component, how much of it goes to the hospital, and what methodology you plan to use to provide the financial incentive back to the people who actually did the work of looking after the patient and organizing the system.

So I close my comments by saying I think it is a great paper and I think it will help all of us do a much better job with our quality improvement. And I hope all the people will use this.

Dr. Alfred S. Casale (Danville, Pennsylvania): You are certainly right that one of the important things to getting buy-in from the clinicians for this was the careful translation of general guidelines into verifiable, measurable behaviors. The AHA ACC guidelines for coronary grafting are about as good as any guidelines we have focusing on surgical procedure. But even they are very general, almost like "eat your vegetables." It is hard to measure that. We then translated those generalizations into specifics like "eat 2 cups of broccoli every 24 hours," because that could be measured, that could be quantitated and followed.

The fact that the number of "statistically significant differences" was very small originally put us off as well, until we acknowledged that that was really only a secondary goal of the project. The project was designed to re-engineer processes of care, and the actual guidelines that were used could be changed as science progresses.

We started from a very high bar. Because of years of using the STS database as a benchmark, and many ongoing quality improvement efforts, our observed to expected mortality ratio and complication rates were already very good. So that even a statistically insignificant change, for example, return to the ICU from 3% to 1%, was, we thought, meaningful in a clinical way.

You asked about what happens to the extra funds that are available by virtue of this project. So far, this has been an internal experiment. We are transferring funds from our health insurer to the clinical enterprise. By virtue of the small increment in case rate that is equivalent to half the cost of complications as experienced in a pre-program baseline period, some potential for a net increase in reimbursement exists if we reduce severity or incidence of post-op problems beyond that level. But the real innovation here is that we, as a system, have essentially transferred the risk for complications above that level from the insurer to us. Basically, we have provided a "warranty" that our process improvement efforts will reduce complications and cost or we will provide the care extra care needed gratis. But the system is really not losing or gaining any money at this point until we get further into selling to actual purchasers of care. But the bet we are making is that if we can reduce the cost of complications and readmissions to 50% of what the baseline period was, then the system and the clinical enterprise, will in fact be doing better as patients do better.

Dr. Bartley P. Griffith (Baltimore, Maryland): Performance-based outcome and pricing are on the horizon and you are leading the way.

I would say that health service research is about effectiveness, and you have shown that you can effectively decomplicate complicated health care by instituting step processes that you can now adhere to.

Because you picked your low risk patient population in which you are a premiere provider already, it is hard to show improvement. The modest improvement that you showed in discharge days may be based on your patient contract. It would be interesting to look at a validating instrument called a patient activation measure. There is a Hibbard test in which you can actually sort out how well that contract is being followed by the patients.

Finally, is this whole exercise actually improving care? I do not think you will really learn that until you get to the more complicated patient populations. I would love to know whether you think it is going to be generalizable beyond the rather homogeneous group in your neck of the woods to my very heterogeneous group in West Baltimore.

Dr. Alfred S. Casale (Danville, Pennsylvania): The issue of generalizability is one we are very interested in as well. We have noticed that even though only elective cases were included in this original design, it is very hard, once you do things right, to not do them right for everybody else.

So we went back and looked at the reliability of delivering these care elements to patients who were not "included formally in the ProvenCare process," and it was very high. We have also happily noticed things like the CRNAs getting used to giving antibiotics appropriately; it is hard for them not do this in orthopedic cases and urology patients, et cetera. So, one of our happy consequences of a cardiac surgery focused project was increasing the level of performance across a broader part of our entire surgical program.

And we share your enthusiasm for the patient activation part of this. Clearly, that was one of the important parts from the very beginning. We do think that some measure of compliance with the "compact" is important and will certainly look into the patient activation measure you mention. We considered but rejected as overly draconian at this time doing things like measuring nicotine levels during follow-up or canceling the "deal," essentially "voiding the warranty" if patients missed appointments, did not participate in rehab etc. Patients need to join us in recognizing that we are not doing an operation to them, but that an operation is becoming part of a life-changing event that they need to carry on with.

DR. RICHARD J. SHEMIN (LOS ANGELES, CALIFORNIA): I have a few brief questions.

Obviously, many of the pay-for-performance processes are so elementary and basic; it is ridiculous that doctors should be paid bonus dollars to do what should be part of our medical responsibilities.

You developed a composite of 40 quality processes. Do you have any sense as to which of the 40 are most important?

Your institution is part of a very unique health care system. How much do you think the success of your program is due to the integration and the size of the system?

I see you have trademarked ProvenCare℠. What are the implications of that? Is your platform for sale?

DR. ALFRED S. CASALE (DANVILLE, PENNSYLVANIA): No, we have not picked apart which of the 40 elements of care we focused on as the most important. We adopted the AHA/ACC guidelines in toto and translated them, as we discussed, into verifiable behaviors. We did not design the project to validate any one, but to develop a mechanism to reliably deliver them all. We are now revising the best practices to add newly identified elements and to modify the existing ones to take new science into account.

The system is integrated, but the variation among the processes that all 8 surgeons had at baseline was really enormous. So the part of this program that involved developing consensus within a group of high functioning individuals accustomed to using prerogative frequently, I think, is very generalizable.

The electronic health record and a consistent group of mid-levels helping out are important, and, I think, we will have to wait to see how other people generalize things like this. But using principles to apply evidence-based care has to become the wave of our future regardless of where we practice.

We service marked the name ProvenCare℠ only to allow us to maintain it as a shorthand for Geisinger's best practice focused care redesign that tries to offer an alternative to the piecemeal payment system generally applied in acute care situations.

DR. FREDERICK L. GROVER (DENVER, COLORADO): I became very involved with this issue in my position at the STS last year when it was obvious that Congress and CMS would roll this out very quickly and were unsure as to how to do it. We wanted to be present and at least advise them. Jeff Rich testified twice before the Health Subcommittee of the House Ways and Means Committee on the Virginia demonstration project indeed showing some of the things that you showed, that if you decrease complications you can save money. And actually, the programs in Virginia have the best ratios at the lowest cost—even going into the question what does one complication cost in terms of hospital costs? Renal failure, for example, adds $50,000 onto the bill.

We then proposed to Congress that if we were able, through a national educational program, like you are using at your own locale, to decrease the complication rate by 10%, that would totally fund the pay-for-performance for cardiothoracic surgery. I asked Congressman Thomas, who chaired the House Ways and Means Committee one time, if he would allow us to keep that, and quite frankly, he was pretty vague on that subject, as you can well imagine.

I want to call your attention to a supplement in the April issue of *The Annals of Thoracic Surgery*, because I think how to measure performance applies across the board to all specialties---particularly all surgical specialties, a paper by Dave Shahian, first author from Boston, but using the Duke Clinical Research Institute statisticians and the Harvard biostatisticians, on developing a model for performance based on processes of care and risk adjusted outcomes. The model actually weights the relative importance of each of these things statistically. It weights the operative mortality of adjusting mortality more than the process factors, but it does not have the bias that it would if we did it ourselves.

A question I would like to ask you is one that we have discussed in NQF multiple times. And that is, if you do pay for performance, how do you decide who gets rewarded? Is it the top 10%? Is it the top half? We are dealing with the STS now with Blue Cross-Blue Shield and United Health Care developing relationships using the STS database in the performance criteria that I just mentioned a minute ago. But who get rewarded? What percentage gets rewarded? Are there different amounts of reward?

Also, I think a really important point is, so that you are not hurting the ones in the lower group that struggle with the results by taking money from them and giving it to those that are doing better, what do you have in the way of upward mobility? Do you have a reward for pay for performance, if you were to design this yourself, that rewards improvement, not just who is in the top 10% or 25%.

DR. ALFRED S. CASALE (DANVILLE, PENNSYLVANIA): Dr. Grover's leadership in the STS in getting the database to its preeminent position as the best cardiac surgery dataset was really very important in providing a robust data collection, analysis and benchmarking system for doing any of this quality improvement work.

When pay-for-performance was first advocated years ago, it was used very specifically to describe a scheme in which new money would be made available as a reward for increase in quality however defined and measured. The evolution of this from a rewarding to a punitive process is indeed disconcerting and we as a profession should carefully consider our response.