Use and Effect of Quantitative Clinical Feedback for Learning and Quality Improvement in Anesthesiology

by

Andrea Brovender

A thesis submitted in conformity with the requirements for the degree of Master of Science
Institute of Health Policy, Management and Evaluation
University of Toronto

© Copyright by Andrea Brovender, 2015
Use and Effect of Quantitative Clinical Feedback for Learning and Quality Improvement in Anesthesiology

Andrea Brovender

Master of Science

Institute of Health Policy, Management and Evaluation
University of Toronto

2015

Abstract

Providing quantitative performance feedback to individual clinicians and clinical groups has been suggested as a useful strategy to reduce preventable morbidity and mortality within medicine. Anesthesiology departments wishing to design a program of quantitative performance feedback will encounter few published accounts of how such systems are influenced by their contexts and the features that promote effectiveness and usability. Two case studies are presented here which highlight how program design is related to context and perceived effectiveness. Design features which supported usability included prioritizing actionability when choosing metrics, involving the group or end-users during metric creation and goal setting and providing frequent communication of results. Contextual features which were found to support usability included resource availability, a mature safety and learning culture within the organization, and the provision of active leadership to create a sense of purpose, harness motivators and overcome barriers.
Acknowledgments

I would like to express my appreciation for my thesis supervisor, Dr. G. Ross Baker, for his guidance during this journey. His generosity, insightfulness and curiosity were crucial to my graduate school experience and the preparation of this thesis. And also to my committee member Dr. Conor McDonnell, for his help, and for the many conversations we’ve had about clinical problems in anesthesiology. Thanks also go out to Drs. Gofton, Applegarth, Grant, Umedaly and Mayson, my colleagues in arms, who supported the premise of this work and encouraged me to pursue it and Drs. P O’Connor and R Dutton for their help and advice.

Most of all I would like to thank my husband who allowed me to temporarily drag him away from family, friends and skiing to pursue this line of inquiry, and my sons, whose smiles can always manage to brighten the darkest day.
Table of Contents

Acknowledgments .................................................................................................................. iii

Table of Contents ................................................................................................................ iv

List of Tables ......................................................................................................................... viii

List of Figures ......................................................................................................................... ix

List of Appendices ............................................................................................................... x

1 Chapter 1 Introduction .................................................................................................... 1
   1.1 Background and rationale ......................................................................................... 1
   1.2 Overview of thesis chapters ..................................................................................... 2

2 Chapter 2 – Literature Review ....................................................................................... 4
   2.1 Outline ....................................................................................................................... 4
   2.2 Perioperative complications are common and cause considerable suffering. Quality anesthesia care has the potential to decrease this burden. ........................................ 4
   2.3 How performance measurement has been used as part of a larger quality improvement strategy .............................................................................................................. 7
   2.4 The use of information technology for performance measurement and quality improvement in perioperative care ................................................................................. 14
   2.5 Providing performance feedback as a pedagogical strategy for clinical anesthesia providers ......................................................................................................................... 17
   2.6 Would performance measurement influence organizational culture to promote learning and safety? .................................................................................................................. 23
   2.7 Performance Measurement and Feedback can Result in Improved Care but Feedback Structure Matters .............................................................................................................. 31
   2.8 Unintended Effects of Performance Measurement and Feedback ............................. 34
   2.9 Summary ..................................................................................................................... 38
   2.10 Purpose and Research Questions ............................................................................. 41
Chapter 3 - Research Procedures

3.1 The role of case study research in understanding quality improvement efforts

3.2 Data Collection

3.2.1 Sampling

3.2.2 Site Visits

3.2.3 Collection of data from documents and administration

3.2.4 Collection of Semi-Structured Interview Data

3.3 Data Transcription, Management and Analysis

3.3.1 Interview Data Transcription and Management

3.3.2 Case Database

3.3.3 Analysis

3.4 Researcher Reflexivity

3.5 Ethical Considerations and Consent

3.6 Judging the Quality of Case Study Research Design

Chapter 4 – Descriptions of the sites and their quality and data systems

4.1 Descriptions of Study Sites

4.2 Resources for Quality Improvement and Provider Learning/ CME

4.3 Description of data and feedback systems

4.3.1 Program Inception and Development

4.3.2 Measure Choice, Collection and Communication

4.3.3 Use of the performance feedback systems

4.4 Summary

Chapter 5 – Findings and Discussion

5.1 Learning and Practice Improvement generated by the Measurement of Quality Metrics

5.1.1 Personal or Individual Learning
5.1.2 Group Learning

5.1.3 Motivation

5.1.4 Strategies for Self-Assessment and Sources of Feedback

5.2 Performance measure choice

5.2.1 Useful metrics are representative

5.2.2 Useful metrics are salient

5.2.3 Useful metrics are designed to be actionable

5.2.4 The number of metrics should not be overwhelming

5.3 Communication and Collaboration

5.3.1 Indicator and Report Interpretation

5.3.2 Communication around program results and implications

5.4 The Role of Leadership

5.4.1 Site I – Limited leadership

5.4.2 Site II – Supportive leadership from multiple institutional levels

5.5 Group norms and culture

5.5.1 Site I - Individualism and personal responsibility

5.5.2 Site II – Team effort and group responsibility

5.5.3 Performance measurement can emphasize differences between types of providers

5.5.4 Anesthesiologist Identity

5.6 Summary

6 Conclusions

6.1 Different performance measurement and performance improvement models were selected by Site I and Site II

6.1.1 Design features of feedback systems that support learning and QI activities

6.1.2 Use of electronic health information systems to support data collection and analysis
6.1.3 Supporting structures and resources ................................................................. 134
6.2 Integration of quantitative feedback into self-assessment and learning strategies .... 135
6.3 Perceptions of quantitative feedback systems ....................................................... 137
6.4 Contextual elements that support or impede learning and change efforts in response to quantitative feedback ......................................................................................... 138
6.5 Guidance to other centers and areas for future research .................................... 143
7 Appendix A - Descriptive and contextual data collection form ............................. 150
8 Appendix B – Interview guide .................................................................................... 153
9 Appendix C – List of abbreviations ........................................................................... 156
10 References ............................................................................................................... 159
List of Tables

Box 1: IOM dimensions of quality

Table 2.1: Factors at the organizational, team and individual levels that may have an impact on the development, use and efficacy of performance measurement in anesthetic care

Table 4.1: Major characteristics of each study site

Table 4.2: Characteristics of the quantitative feedback collection and reporting system at each site

Table 4.3: Quantitative performance feedback metrics chosen by each study site

Table 4.4: Summary of key differences between sites

Table 5.1: Contextual features and barriers affecting usability of quantitative performance feedback at each site

Table 5.2: Site I - Barriers to using quantitative feedback for learning and quality improvement and potential corrective strategies

Table 5.3: Site II - Barriers to using quantitative feedback for learning and quality improvement and potential corrective strategies
List of Figures

Figure 1: Donabedian structure-process-outcome model of quality

Figure 2: Typology of organizational cultures as conceived by Westrum, Parker and Hudson

Figure 3: Schematic representation of the use of quantitative feedback for learning and quality improvement at each site.

Figure 4: Sample quantitative performance reports using mock data
List of Appendices

Appendix A - Descriptive and contextual data collection form

Appendix B - Interview guide

Appendix C – List of abbreviations
1 Chapter 1 Introduction

In 1999 the Institute of Medicine (IOM) released its seminal report, *To Err is Human*, which described the burden and consequences of adverse events and errors in healthcare settings (Kohn, Corrigan, & Donaldson, 1999). This document was part of a movement towards higher standards of safety and quality that involved healthcare providers, insurers and regulators. Improvement efforts in healthcare have borrowed strategies from other sectors such as continuous quality improvement (CQI) from manufacturing and the concept of safety culture from other high risk industries such as aviation and oil and gas exploration.

Surgical care is complex and over time, as technology is introduced and refined, surgical options are extended to older patients and patients with more comorbidities. Surgical patients are more likely to experience adverse events in-hospital than medical patients and half of these events are judged to be preventable (Brennan et al., 1991; Gawande, Thomas, Zinner, & Brennan, 1999).

The scope of anesthesiology has broadened beyond rendering patients insensate for surgery towards a practice of perioperative medicine which includes larger roles in preoperative assessment and optimization and postoperative management and pain. Improving anesthetic care has the potential to not only decrease anesthesia-related adverse events but to also improve overall surgical morbidity and mortality.

The ability to examine the effect of process and system changes on outcomes of anesthetic care requires the collection of data. As electronic health records, including Anesthesia Information Management Systems (AIMS), are becoming more prevalent, data availability is increasing and the cost of collection and analysis is decreasing compared to human abstracted data from clinical charts. It is important to assess strategies for using quantitative feedback in order to find the ones most likely to result in learning and quality improvement in anesthesiology care.

1.1 Background and rationale

Examples of programs that provide quantitative feedback to physicians for the purpose of improvement include structured practice audits (Ivers et al., 2012), and registry projects such as the New York State Cardiac Registries (Hannan, Cozzens, King, Walford, & Shah, 2012), and the National Surgical Quality Improvement Project (NSQIP) (Khuri et al., 1998). Electronic
medical records increase the ability to access data and it is anticipated that more clinical groups will become interested in using this data to provide feedback to their members. The effect of quantitative feedback on anesthesiology care quality and anesthesiologist learning has not been well described in the literature. In order to explore how quantitative feedback to anesthesiologists could be designed to increase effectiveness and usability, groups that were collecting data and providing quantitative clinical feedback to their members were sought for participation in a series of case studies.

Two groups with very different approaches were identified. These case studies were informed by a wide body of literature on performance measurement in medicine, feedback acceptance and use, adult pedagogy and organizational behavior. The goals of this study were to explore how quantitative performance feedback can be used to improve anesthetic practice, the contextual features that influence this process, and how the systems are viewed by those receiving the feedback. Using these two cases to explore these questions, and relating user experiences to their context will provide guidance to those designing quantitative feedback programs as they make decisions relevant to their specific goals and contexts.

1.2 Overview of thesis chapters

This thesis is organized into six chapters. Chapter two is a literature review that brings together several diverse topics for the reader. Quality and safety challenges in surgical care are reviewed along with the potential that anesthesiologists, as perioperative physicians, to make changes that will decrease human suffering. The reader is then offered grounding in the history of performance measurement as a quality improvement strategy in healthcare and the effect of changes in information technology and electronic health records (EHR) on this strategy. The review then considers the effect of performance feedback on learning in medicine by examining both individual and group clinical learning and organizational learning and culture change. Finally consideration is given to the theoretical and realized negative effects of performance measurement and feedback in the healthcare context.

Chapter three presents a description of, and rationale for, the use of a comparative case study format as a way of exploring the use of quantitative performance feedback by clinical anesthesiology providers.
Chapter four is descriptive and details the setting and characteristics of the anesthesia practice, and of the system used to collect, communicate and act on quantitative clinical feedback at each study site. The way that this data is used to promote learning and improve the quality of care is described and placed in perspective by considering its position within other quality improvement and quality assurance efforts at each site.

Chapter five builds on the descriptions in chapter four through analysis and discussion. The learning and quality improvement activities occurring at each site that are related to the availability of quantitative performance feedback are described. A discussion of how these activities are influenced by program features such as metric choice, presentation and communication follows. This chapter is interpretive and lays out how certain contextual elements influenced, and are influenced by, the performance measurement and reporting system present at each site.

The final section, chapter six, presents the study’s conclusions and ties together the interpretations of chapter five with the research questions posed by the literature review. It also explores areas for further study.
Chapter 2 – Literature Review

2.1 Outline

An investigation into the use of data systems to provide practice-specific feedback to anesthesiologists can draw on knowledge from quality improvement theory and practice, the history of data system use in anesthesiology, physician reflection and learning, and organizational behavior, learning and culture. First, the significance of errors and quality gaps in perioperative medicine and surgical services will be examined along with some current strategies to address them. This will be followed by a discussion of how integrated data systems and anesthesia information management systems (AIMS) have been used to collect process and outcome information and provide feedback to clinical anesthesiology providers. The process of clinician learning from performance feedback, including the effect of feedback structure, and the importance of self-regulation and reflection will be discussed. Group learning in response to performance feedback, and how this interacts with organizational culture, will be reviewed. The state of the evidence for the use of both individual and group performance feedback as a quality improvement strategy in clinical medicine will be examined along with a discussion of how system characteristics may influence both intended and unintended effects.

2.2 Perioperative complications are common and cause considerable suffering. Quality anesthesia care has the potential to decrease this burden.

An estimated 234 million operations occur globally each year (Weiser et al., 2008). Overall in the developed world, surgical mortality ranges from 0.4 – 0.8% and 3-17% of surgical patients experience complications (Gawande et al., 1999; Kable, Gibberd, & Spigelman, 2002). The mortality rate of non-ambulatory surgical cases is higher: in a review of 3.7 million of these cases in the Netherlands, perioperative mortality was found to be 1.85% (Noordzij et al., 2010).

Adverse events are not uncommon in a non-differentiated hospitalized population. Using a random sampling of patient charts, Brennan et al, found a 3.7% statewide incidence of adverse events in a mixed hospitalized population in New York. Adverse events were narrowly defined as, “an injury that was caused by medical management (rather than the underlying disease) and that prolonged the hospitalization, produced a disability at the time of discharge, or both.”
Adverse event rates in surgical populations tended to be higher than those in medical populations and the differences in rates between specialties was statistically significant, \( p < 0.0001 \) even though rates of negligence were comparable (Brennan et al., 1991). This suggests that surgical patients are at increased risk for adverse events compared to medical patients. This difference may be due to higher complexity of disease and associated interventions.

Using a similar definition, a chart review of 14 700 non-psychiatric patients in Colorado and Utah found an adverse event incidence of 3.0%. Two-thirds of these events were related to an operation or non-operative care by a surgeon and roughly half were judged to be preventable. Although the most commonly occurring surgical adverse event category was technique-related (24.2\% of surgical adverse events), a significant number of events were related to perioperative care such as surgical site infections, bleeding, cardiac events, and thrombotic and embolic events. Direct anesthetic injury accounted for 1.6\% of all surgical adverse events (Gawande et al., 1999). Using similar methodology, a study of only surgical patients in Australia found an adverse event rate of 21.9\%. When an adverse event occurred, it resulted in permanent disability 13\% of the time and death 4\% of the time. Again, roughly half the adverse events were judged to be preventable (Kable et al., 2002). Despite a greater emphasis on quality and safety in health care, the rates of adverse events found in a retrospective study of a stratified random selection of ten North Carolinian hospitals remained stable over the five year period (Landrigan et al., 2010) after the 1999 publication *To Err is Human*.

In a study of surgical patients in The Netherlands based on routine doctor-reporting over three years, adverse events occurred during 18.2\% of admissions. The definition of adverse event in this study included events that were attributable to the patient’s disease state and comorbidities along with iatrogenic events due solely to medical and surgical care. An adverse event was “an unintended and unwanted event or state occurring during or following medical care that is so harmful to a patient’s health that (adjustment of) treatment is required or that permanent damage results.” Operative treatment led to the most adverse events (42.5\%), and of all events, 20.9\% were judged to be preventable. The burden of adverse events was clinically and socially significant: 5.1\% of adverse events resulted in death, 3.1\% resulted in permanent damage or loss of function, and 21.6\% required an operation or reoperation to restore the patient to full function (Kievit, Kurkerink, & Marang-van de Mheen, 2010).
Five percent of the incidents that were reported to the Australian Anaesthetic Incident Monitoring Study (AIMS) occurred in the recovery room. Nearly one third of these incidents were serious enough to require patient transfer to a high-dependency unit or intensive care area. The AIMS database, which is based on voluntary self-reporting, may over-represent serious incidents (Kluger & Bullock, 2002). However, even minor recovery room incidents lead to increased resource utilization including nursing resources and increased length of stay (Bothner, Georgieff, & Schwilck, 1999).

Not only do postoperative complications cause suffering for patients, they also increase the cost of surgical care. Surgical complications were found to add an average of $10 000 per patient (Dimick, Weeks, Karia, Das, & Campbell, 2006) and hospitals in the highest complication quintile received payments from the Medicare system in the United States that ranged from $2436 to $5353 higher than hospitals in the lowest quintile for four common surgical procedures despite similar case mixes and risk profiles (Birkmeyer, Gust, Dimick, Birkmeyer, & Skinner, 2012).

Anesthesiologists have an opportunity to significantly impact the quality of perioperative care and intraoperative strategies aimed at reducing perioperative morbidity and mortality have been studied (Moonesinghe, Mythen, & Grocott, 2011). Examples include tight glycemic control (van den Berghe et al., 2001; Wiener, Wiener, & Larson, 2008), goal-directed fluid and hemodynamic therapy (Donati et al., 2007; Gan et al., 2002; Pearse et al., 2005; Poeze, Greve, & Ramsay, 2005; Sandham et al., 2003), and the use of neuraxial and regional techniques either on their own or combined with general anesthesia (Curatolo, 2010; Rigg, Jamrozik, & Myles, 2002; Rodgers et al., 2000; Wijeysundera, Beattie, Austin, Hux, & Laupacis, 2008). Attempts have been made to identify anesthetic processes that can influence postoperative pulmonary complications (Berg et al., 1997; Kilpatrick & Slinger, 2010; Pedersen, Vibly-Mogensen, & Ringsted, 1992; Popping, Elia, Marret, Remy, & Tramer, 2008; Rodgers et al., 2000; Sachdev & Napolitano, 2012; Tsui et al., 1997). Anesthetic management is an integral part of Enhanced Recovery After Surgery (ERAS) protocols which can reduce length of stay and postoperative morbidity (Doorley & Senagore, 2012; Kehlet & Wilmore, 2008).

Surgical site infections are a significant source of morbidity and mortality that occur after approximately 5% of surgeries and make up 20% of hospital acquired infections (Klevens et al.,
2007; Klevens et al., 2007; Leaper, 2010; Leaper, 2010). Prophylactic antibiotic administration prior to surgical incision is a shared surgical and anesthetic responsibility in many institutions. Anesthetic techniques that can influence SSI rates include hand washing (Loftus et al., 2011; Roy, Brull, & Eichhorn, 2011), maintenance of normothermia, use of regional anesthesia, administration of supplemental oxygen, glycemic control, restrictive fluid management, and use of blood conservation strategies (Gifford, Christelis, & Cheng, 2011).

Surgical patients experience considerable morbidity and risk for mortality during their treatment course. Optimal care by the surgical and anesthetic team can improve patient outcomes and there have been many initiatives within institutions and through multi-center and national quality improvement programs which aim to improve perioperative care.

2.3 How performance measurement has been used as part of a larger quality improvement strategy

The IOM defines quality as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Archer & Macario, 2006). Implicit in this definition is the need to measure both processes, which are the application of current professional knowledge, as well as outcomes. Performance measurement is the first step in the basic model of quality improvement - a cycle of observing outcomes, analyzing causation, making changes and re-observing (Dutton & DuKatz, 2011).

Drucker conceptualized “managing by objective” (Drucker, 2006) over 40 years ago. He viewed health care institutions as knowledge-based organizations whose front-line workers were repositories of specialized information and skills. Performance feedback in these organizations was seen as a way for management, who set goals and performance targets, to form a partnership with specialists in order to facilitate continuous learning. Sharing information and feedback with front line workers was seen as essential to improvement efforts (Drucker, 1991). From a quality improvement perspective, Berwick, building on earlier work by Deming and Shewhart, saw performance measurement as a part of his Plan-Do-Study-Act (PDSA) cycle as a way to identify areas for improvement (plan) and monitor improvement efforts (study) (Berwick, 1996). Both models suggest a shared responsibility for delivering quality care and imply a significant effect.
of organizational and cultural context on a physician’s ability to modify their practice (Friedrich, 2009).

Different stakeholders in healthcare approach the use of performance measurement with varying viewpoints and motives. Physician groups that have taken an interest in performance measurement and feedback have done so out of a professional desire to improve patient care and disseminate best practices (Friedrich, 2009). For example, the American College of Cardiology stressed the importance of performance monitoring and feedback in its “Guidelines Applied in Practice” project framework (Montoye, Eagle, Michigan ACC-GAP Investigators, ACC-GAP Steering Committee, & American College of Cardiology., 2005). Health care management professionals and payment organizations may stress the importance of “value” or buying better outcomes for fewer dollars. Clinicians have a strong tradition of prioritizing outcomes over cost and may distrust quality initiatives that prioritize efficiency.

Within the specialty of anesthesiology, there is increasing interest in pulling provider or group-specific feedback from data sets containing information on anesthetic process, patient risk factors and outcomes. This feedback could generate provider behavior change and inspire quality improvement projects. A 2011 white paper (Glance, Neuman, Martinez, Pauker, & Dutton, 2011) reviewed many of the issues surrounding performance measurement in perioperative care and anesthesiology within the Donabedian structure-process-outcome model of quality (Donabedian, 1986). In this model, the outcome of care is a function of a patient’s pre-existing state and risk factors, the structure in which care is delivered and the processes of care (Figure 1). Structure, process and outcome are all potential targets for measurement.
Examples of structural measurements that may be related to quality of care include organizational characteristics such as hospital size, case volume, the presence or absence of trainees, human resources elements such as the experience and credentialing of staff and the time of day when care was provided (Hurtado, Swift, & Corrigan, 2001). For example, in the early 1980s it was discovered that rotating anesthesia staff through long or difficult anesthetics (Cooper, Long, Newbower, & Philip, 1982) and providing team-based anesthetic care (Bechtoldt, 1981) resulted in improved patient outcomes and increased detection of intraoperative errors. Data from the National Surgical Quality Improvement Project (NSQIP) was used to show a significant increase in mortality after surgery performed on Fridays compared to earlier in the week (Zare, Itani, Schifftner, Henderson, & Khuri, 2007). An association between surgical volumes and improved outcomes has been demonstrated (Halm, Lee, & Chassin, 2002). For example, some of the outcome improvements that were made after public reporting on cardiac surgeries in New York State were due to the removal of low-volume practitioners (Hannan et al., 2012). It can be difficult to identify how to modify some structural elements to improve health care quality. Structural measures are often not actionable at the individual or group health care provider level but do provide information for higher-level decision making and program planning.
Process elements describe interactions between patients and health care providers including the frequency and content of information exchanges, diagnostic testing, and procedural and pharmacologic treatment. A health care process is an attractive target for measurement if there is strong evidence that it is an important determinant of clinically meaningful patient outcomes (Glance et al., 2011; Spertus et al., 2010). Many process measures do not need to be risk adjusted which simplifies collection and analysis. It is also gratifying to show improved adherence to a pre-determined “best-practice.” Process measures may be more sensitive to differences between providers and time periods compared to outcome measures (Mant & Hicks, 1995; Mant, 2001) making them more useful in guiding local practice. Since not every situation where a best-practice process is not followed will lead to a poor patient outcome or evidence of harm, process measurements may act as leading indicators or warnings of possible future adverse events.

One problem with process measurement is disagreement about what constitutes best practice as in many cases sufficient evidence is not available. In many cases process measures have been created through adaptation of clinical practice guidelines but this has introduced bias. Clinical practice guidelines help providers and patients improve decision making in specific clinical situations through distilling the evidence and providing recommendations. While they have helped disseminate new practices and decrease practice variation, they are not treatment protocols. The recommendations included in guidelines may not be appropriate for every patient and the quality of evidence underlying them is varied. A systematic appraisal of 22 clinical practice guidelines pertaining to perioperative care concluded that only half could be recommended for use in clinical practice (Barajas-Nava et al., 2010). The review also found that half of the guidelines were created without input from patients or patient advocates and often fail to consider value-based decision making.

The use of publicly reported process measures has resulted in dramatic improvement in adherence by American hospitals (Chassin, Loeb, Schmaltz, & Wachter, 2010) but empirically linking process to outcomes has been more difficult. Correlation was found between publicly reported hospital performance on the Center for Medicare and Medicaid Services (CMS) website and risk-adjusted mortality (Jha, Orav, Li, & Epstein, 2007; Werner & Bradlow, 2006). The Surgical Care Improvement Project (SCIP), a joint venture by the Center for Medicare and Medicaid Services (CMS) and the Joint Commission on the Accreditation of Hospital
Organizations (JCAHO) identified bundles of care designed to prevent surgical site infection, venous thromboembolism, cardiac complications and ventilator-associated pneumonia (Joint Council on the Accreditation of Hospital Organizations). The use of SCIP measures decreased the incidence of surgical site infection in one study (Stulberg et al., 2010) but other studies have failed to show a benefit (Awad, 2012). It is unclear if the SCIP measures are ineffective or if studies have been unable to measure the effect (Schwulst & Mazuski, 2012). An older review found only a weak link between process and quality care indicators and risk-adjusted mortality (J. W. Thomas & Hofer, 1998). In one randomized controlled trial, the use of continuous quality improvement strategies which included longitudinal and benchmarked process feedback, compared to no intervention, improved the use of, and adherence to, two process measures in patients undergoing coronary artery bypass grafting (CABG) (Ferguson et al., 2003).

Outcome measures refer to the effect of the health care system on the health status of patients and populations, for example, 30-day mortality and quality-adjusted life years. Outcome measurements are the most meaningful to patients (Krumholz, Normand, Spertus, Shahian, & Bradley, 2007) as they represent what Ernest Codman, a physician who pioneered evidence-based medicine at the turn of the century, deemed the “end result” of care (Codman, Reprint 1996; Kaska & Weinstein, 1998).

As attractive as they might be, outcome measures pose significant challenges to data collectors and analyzers. Patient outcomes are produced via complex interactions between the effectiveness of medical care, pre-existing conditions, risk factors, and chance (Iezzoni, 2003). Outcome definitions need to be rigorous in order to avoid ambiguity meaningful comparisons between institutions or providers requires risk adjustment. Risk adjustment models do not currently exist for many outcomes of interest and their development and validation is a resource intensive process (Spertus et al., 2010). Risk adjustment will likely never account for all of the potential variance because of unmeasured or unknown factors. The use of different models has been shown to influence the rankings of institutions and providers (Lilford, Mohammed, Spiegelhalter, & Thomson, 2004). However, for outcome measures designed for internal use rather than public reporting, risk adjustment may not be as important. Institutions may get useful information by comparing outcomes over time as long as they are prepared to assume that baseline population characteristics are stable and the introduction of potentially disruptive technologies is accounted for.
Risk adjustment models can be based on either administrative or clinical data. Clinical data is more expensive to acquire but is considered to be more accurate than administrative data. Many of the respected outcome reporting systems in operative care such as the Society of Thoracic Surgeons (STS), the Veterans Administration (VA) National Surgical Quality Improvement Program (NSQIP), and the American College of Surgeons (ACS) NSQIP and Trauma Quality Improvement Program (ACS TQIP) use clinical data (Nathens, Cryer, & Fildes, 2012).

There is some evidence that outcome measurement and feedback improves patient care from both observational (Hannan, Kilburn, Racz, Shields, & Chassin, 1994; Hannan, Sarrazin, Doran, & Rosenthal, 2003; Jha, Wright, & Perlin, 2007), and prospective cohort studies (B. L. Hall et al., 2009; Khuri et al., 1998; O’Connor et al., 1996). For example, the Veterans Health Administration used their electronic health record (EHR) system to provide clinicians with process and outcome feedback as a strategy to increase adherence to evidence-based guidelines for diabetic patients (Kupersmith et al., 2007). The mechanisms that underlie these benefits, whether the Hawthorne effect, provider behavioral change, targeted programs to improve specific indicators, a general improvement in medical science over time or some combination, has been more difficult to explicitly demonstrate within the context of research studies.

There are two main postulated pathways through which measurement can result in improvement (Berwick, James, & Coye, 2003). The first is selection, through which consumers, whether they are patients, referring physicians, payers or organizations, use publicly available performance measures to choose the most effective practitioner or center. Improvement would occur as poor performing entities make efforts to improve their rankings, or are removed from the market and as higher case volumes flow towards maximally performing entities. This mechanism may be minor as most evidence suggests that patients are unaware of available performance measures, have considerable difficulty understanding and interpreting them, and that both patients and referring physicians rarely use them to make decisions regarding institution or provider (Kaiser Family Foundation and Agency for Healthcare Research and Quality, 2000; Kaiser Family Foundation and Agency for Healthcare Research and Quality, 2004; Marshall & Shekelle, 2000; McIntyre, Rogers, & Heier, 2001; Schneider & Epstein, 1996). Although insurers have shown interest in using rankings to make reimbursement decisions, they have faced significant push back from physician and patient-rights groups including legal action and lobbying campaigns (Miller, Brennan, & Milstein, 2009).
The second pathway by which performance measurement could lead to improvement is through facilitating examination of processes and results. Once performance measures are collected, they can be compared to published or theoretical benchmarks and inspected for trends over time so that opportunities for improvement are identified. Teams can then use well established methodologies for planning and executing QI programs such as root-cause analysis, Six Sigma, Lean, PDSA and the principles of change leadership (Morrow, 2012). At the group or institution level this would manifest as changes in policy, work flow and procedure and at the individual clinician level through reflection, learning and behavioral change.

One example of a successful initiative using this approach is the project developed by Johns Hopkins University School of Medicine to decrease central line-associated bloodstream infections (CLABSIs). The program consisted of translating best evidence into practice, measurement of results, ongoing feedback to front-line staff and managers and activities designed improve teamwork and change culture. The project initially reduced CLABSIs by 66%, a change that has been sustained for many years. The program has been expanded across the United States and internationally (Sawyer et al., 2010).

Boston Children’s Hospital has created and expanded an innovative program that uses process and outcome measurement and feedback to standardize and improve care as well as generate new knowledge. Their Standardized Clinical Assessment and Management Plans (SCAMPs) were developed after recognition that strong evidence to inform the creation of clinical practice guidelines was lacking in several areas of pediatric cardiology care (Darst, Newburger, Resch, Rathod, & Lock, 2010). SCAMPs are care plans that are created through a process of literature review and expert clinician input and presented to clinicians in the form of decision trees that help guide care. Deviations from the suggested courses of action in the SCAMP, and their reasons, are recorded and analyzed at specified periodic intervals and this information is used to revise the SCAMP (Farias et al., 2012; Farias et al., 2013; Rathod et al., 2010). This process translates and incorporates in-house knowledge into a decision tree in a way that preserves respect for the autonomy and clinical expertise of the involved physicians. Providers respond favorably to SCAMPS (Farias et al., 2011) with high rates of adherence. SCAMP implementation has resulted in decreased practice variation and may lead to a reduction in the use of resources with no loss of care quality (Friedman et al., 2010).
Many programs gather and analyze process and outcome metrics. The innovative idea that forms the basis of SCAMPs was the decision to measure deviations and record the reasons behind them in order to refine and improve the tool. Feedback, in this example, is used to form a collective notion of “sound practice” and communicated back to the physicians through revised guidelines rather than directly.

Although performance feedback has the potential to be a powerful tool for effecting individual practice and system change, the concept is not without its detractors. Experiments with clinical provider feedback systems in the 1970s and 80s showed that they could be used to modify behavior but at the time the cost of collecting and analyzing data was thought to offset their benefits (Eisenberg & Williams, 1981). With the rise of electronic health records and data systems, more data at lower cost is available than ever before. The decreasing cost of data, in conjunction with calls for improved health care quality from the IOM, regulators, professional societies, insurers, and the public, seems to be driving increased use of performance measurement in healthcare.

2.4 The use of information technology for performance measurement and quality improvement in perioperative care

Electronic anesthesia records, or AIMS, contain information on the processes of anesthesia care and there are many projects underway to utilize this data to improve anesthetic care quality and safety. Penetration of AIMS into the North American market was estimated to be only 14.8% in 2011 but is rapidly increasing (Halbeis, Epstein, Macario, Pearl, & Shahian, 2008; Sandberg, 2008; Stonemetz, 2011). Initially, AIMS were seen as a way to improve the accuracy and legibility of the intraoperative record (Devitt, Rapanos, Kurrek, Cohen, & Shaw, 1999; Muravchick et al., 2008) but the recognition that data repositories can be used for quality assurance and patient safety purposes and used to support clinical research has been a major driving force for AIMS adoption (Halbeis et al., 2008). In the United States, AIMS with reporting functionality can aid anesthesia providers in qualifying for financial incentives (Dutton & DuKatz, 2011; Lai & Kheterpal, 2011; Springman, 2011). In addition to financial incentives, there is administrative pressure to collect data in order to comply with physician privileging standards set out by the Joint Commission. Both the Focused Professional Practice Evaluation (FPPE) and Ongoing Professional Practice Evaluation (OPPE) require collection and analysis of
provider outcome data (Dutton & DuKatz, 2011). AIMS data has been used to track resident educational experiences (Simpao et al., 2011), in what might be an increasing important feature as the specialty moves towards outcomes based training requirements (D. L. Brown, 2011).

Data on anesthetic processes collected by AIMS has been leveraged by those attempting to improve safety and quality of care through the use of clinical decision support and comparative outcomes research.

CDS systems have been shown to help anesthesia providers improve their process of care (Eden, Pizov, Toderis, Kantor, & Perel, 2009; Kooji, Klok, Hollmann, & Kai, 2008; Nair, Newman, Peterson, W, & Schwid, 2010; O'Reilly, Talsma, VanRiper, Kheterpal, & Burney, 2006; Wax et al., 2007). CDS prompts for prophylactic antibiotic administration, perioperative beta-blockade and maintenance of normothermia have now been built into commercially available AIMS systems (Rothman, Sandberg, & St Jacques, 2011). A systematic review on CDS systems from many different clinical fields concluded that while they are largely successful at modifying practitioner behavior, their effects on patient outcomes were understudied and results inconsistent (Garg et al., 2005). Schwann and colleagues studied the effect of real-time reminders to deliver prophylactic intraoperative antibiotics and showed a statistically significant decrease in surgical site infections (Schwann et al., 2011).

Well recognized obstacles to effective CDS use include organizational and cultural barriers, poor usability, advice that is not consistent or reliable, poor integration into existing workflow, inadequate computer literacy amongst staff, production pressure, and a high false-alarm rate (G. H. Kruger & Tremper, 2011). It has been suggested that information systems and CDS in the OR environment may distract from patient care (Hartzband & Groopman, 2008) and decrease situational awareness (Vigoda, Rothman, & Green, 2011). Providers fear that these systems will reduce their efficiency or result in unintended consequences (Ash, Sittig, Campbell, Guappone, & Dykstra, 2007; Eden et al., 2006; Han et al., 2005). Reminders work best when they occur at the time of decision (Kawamoto, Houlihan, Balas, & Loback, 2005) and are combined with an educational initiative or individual provider performance feedback (O'Reilly et al., 2006; Wanderer, Sandberg, & Ehrenfeld, 2011). Reminders have a greater treatment effect in the lowest performing group (Rusa, Klatil, Fu, & Swide, 2009) and their effect on behavior fades
with time (Demakis et al., 2000). Beneficial effects CDS are more long lasting if ways are found to integrate new behaviors seamlessly with workflow (Vigoda & Lubarsky, 2006).

There are several large scale attempts to use AIMS data for comparative outcome research. An early effort in this direction integrated AIMS data with the hospital mortality database to show that intraoperative adverse events were associated with mortality (Sanborn, Castro, Kuroda, & Thys, 1996). The National Anesthesia Comparative Outcome Registry (NACOR), spearheaded by the Anesthesia Quality Institute (AQI), aims to capture electronic data from all of the 25 million anesthetic and several million pain procedures that occur in the United States each year (Anesthesia Quality Institute, ). In addition to comparative outcomes data at the provider, group and institution levels, it will provide denominator data that will help analyze events from the closed claims database (Dutton & DuKatz, 2011). Currently, AIMS use is more prevalent in larger hospital and academic practices which introduces a bias into NACOR data. There is also a high degree of AIMS customization leading to variability in the data available between centers and difficulties with data translation and mapping.

The Multicenter Perioperative Outcomes Group (MPOG), founded in 2008 and based out of the University of Michigan, links perioperative data from 30 anesthesiology departments AIMS to the Social Security Death Master File and perioperative renal and myocardial injury outcomes defined by laboratory data (Kheterpal, Berman, Aziz, Jameson, & Tremper, 2010; Kheterpal, 2011). This data was designed to be used for observational research with an emphasis on patient-centric outcomes.

The Pediatric Regional Anesthesia Network (PRAN) gathers data on the practice patterns, complications and regional techniques used in the pediatric population. The database can be used for comparative outcomes research and to provide risk and safety data. In addition, participating sites have continuous access to their own data, and the current aggregate numbers for the entire dataset, which provides ongoing feedback about regional block utilization and outcomes indexed to that of all members (Polaner, Martin, & PRAN Investigators, 2012).

These systems could deliver provider- and group-specific feedback with respect to both process and outcomes and could provide a way to extend the work done to date with respect to CDS systems in perioperative medicine. Low-frequency, but highly consequential, anesthesia complications will likely only be captured by large scale multi-center data projects. Multi-group
data can provide information that would help establish benchmarks and allow departments to see how they compare with these benchmarks and with their peers. Although not a widespread strategy yet, AIMS data could be collected and analyzed to provide quantitative feedback at the individual and group levels to anesthesia providers. This feedback would be intended to result in reflection and be used in QI projects and could drive and disseminate organization learning through creation of, and changes to, clinical decision pathways, institutional process maps and standardized order sets. The effect of these changes would be monitored through ongoing feedback and monitoring of processes and outcomes.

2.5 Providing performance feedback as a pedagogical strategy for clinical anesthesia providers

The terms “self-assessment,” “feedback,” and “reflection” occur frequently and prominently in the medical education literature and it is worthwhile defining these concepts before looking at how formal quantitative feedback from a performance monitoring or quality system is, or could be used by, physicians as a pedagogical strategy.

Self-assessment has been broadly conceived as the process by which a learner judges whether or not they have reached some identified standard but is likely a more complex process made up of two distinct activities that may not have identical mechanisms. The first is the ability of an individual to identify weaknesses and deficiencies in their skill set while the second is the ability to identify strengths (Eva & Regehr, 2005). Self-assessment that occurs during clinical practice, or concurrent self-assessment, serves as a way for physicians to identify situations where they may require help in the form of information, consultation, other types of assistance or a change in plan. Self-assessment that occurs on the entirety of their practice or with hindsight, leads a physician to make a judgment about his or her performance and skills. This type of self-assessment enables a physician to set learning goals and seek appropriate learning and continuing medical education activities.

Different types of self-assessment appear somewhat analogous to Schön’s “reflection in practice” and “reflection on practice” (Schön, 1987). There are many different definitions and terms in the literature for reflective learning processes including reflective practice, mindfulness, critical reflection, self-regulation and reflective learning but they all share the idea of a cognitive process that “requires engaged examination of both self and situation, and that its goals are
learning and improved future personal and professional effectiveness” (Aronson, Niehaus, Hill-Sakurai, & O'Sullivan, 2012). While self-assessment may answer the question of, “how did I do,” reflection is primarily concerned with questions such as, “why did that turn out the way it did,” or “why am I performing at the level that I am?” In other words, reflection goes beyond trying to solve a problem; it is an attempt to understand a problem (Eva & Regehr, 2008).

Feedback is fundamentally different from evaluation. While evaluation is summative and provides a judgment, feedback is formative and is the, “information that a system uses to make adjustments in reaching a goal” (Ende, 1983). A definition of feedback was proposed after a review of the medical education literature: “Specific information about the comparison between a trainee’s observed performance and a standard, given with the intent to improve the trainee’s performance” (van de Ridder, Stokking, McGaghie, & ten Cate, 2008). This definition could be easily adapted through minor word substitution for use with practitioners who have completed their training. The standard for an independent practitioner could be one that is self-defined such as a practice or learning goal or one that is externally defined such as a published benchmark, group mean or median, or a standard associated with a reward system such as a pay-for-performance program, credentialing or privileging body.

In clinical practice, feedback occurs when a clinician’s actions or inaction is related to consequence in an insightful way. Learning occurs when this results in a behavioral change that alters performance. Feedback in medicine has traditionally occurred between undergraduate and postgraduate medical students and preceptors, been qualitative in nature and based on observation.

Quantitative feedback that consists of a provider’s adherence with process and/ or outcomes for their patients could result in learning if the feedback target is able to reflect on their behavior and pattern of practice and identify areas for improvement or modification. This will require a complex series of cognitive steps through which the clinician self-assesses, reflects, and self-monitors current and future behavior.

Self-regulating professionals are expected to be able assess their knowledge and practice in order to identify gaps and seek out appropriate learning opportunities. The assumption that they are able to self-assess is implicit in how continuing professional education programs are delivered and tracked for physicians in North America (Eva & Regehr, 2008). This expectation has been
formalized within post-graduate medical education in the United States and Canada. The US Accreditation Council for Graduate Medical Education (ACGME) requires education in practice-based learning and improvement (PBLI) so that new clinicians have the skills to “investigate and evaluate their care of patients, appraise and assimilate scientific evidence and continuously improve patient care based on constant self-evaluation and lifelong learning” (Accreditation Council for Graduate Medical Education). The anesthesiology residency program at Johns Hopkins found that a key challenge in implementing an educational program to equip residents with these skills was a lack of faculty mentors with knowledge and skills in practice-based learning (Schwengel et al., 2011). The Royal College of Physicians and Surgeons of Canada and the College of Family Physicians of Canada have structured their Maintenance of Certification and Mainpro ® programs in order to encourage self-assessment and reflection (Silver, Campbell, Marlow, & Sargeant, 2008).

While practice-based learning is undoubtedly a useful and necessary skill for clinicians, there are dangers inherent in a system that does not incorporate formal external feedback. The ability of clinicians and trainees within the health professions to self-assess has generally been found to be poor (Eva & Regehr, 2005; Gordon, 1991; Ward, Gruppen, & Regehr, 2002). In a seminal paper, Kruger and Dunning, demonstrated that low performers were less able to make accurate self-assessments and were more likely to overinflate their abilities (J. Kruger & Dunning, 1999). This phenomenon appears to occur because low performers lack the skill necessary to recognize their shortcomings (Ehrlinger, Johnson, Banner, Duning, & Kruger, 2008) and has been demonstrated in resident physicians (Hodges, Regehr, & Martin, 2001).

In a longitudinal study in the 1970s of resident physicians in an environment that provided almost no feedback from mentor and preceptor physicians, the residents developed systems of self-validation that were not well grounded and that contained a number of defenses that did not allow self-assessments to be altered when external expert feedback was available. These defenses included dismissing the source as not credible, dismissing the subject of the feedback as one of no importance, dismissing the feedback as not timely and therefore not reflective of current performance, or attributing negative feedback as resulting from a difference in style or personality clash (Bucher & Stelling, 1977). This study is suggestive of the possibility that a work environment for clinical anesthesia providers that does not provide frequent timely and relevant feedback may impede learning. A review of studies from the field of medical education
found that summative evaluation did not alter self-assessed performance, rather self-assessments correlated with generalized self-perceptions of ability (Gordon, 1991).

It is well established in the social psychology literature that most people consider themselves to be above average in ability across many diverse categories (Alicke, Klotz, Breitenbecher, Yurek, & Vredenburg, 1995; Pronin, Lin, & Ross, 2002). This is an illustration of self-serving bias. Individuals are also prone to attribution error – mistakenly attributing successes to their efforts and failures to situational factors beyond their control (Henriksen & Dayton, 2006). The implication is that external information must be sought in order to allow an individual to form an accurate picture of their abilities (Lichtenstein & Fischhoff, 1980). For health professionals, this implies the need for feedback:

“…the route to self-improvement is not through becoming a more accurate self-assessor, but through seeking out feedback from reliable and valid external sources (experts, self-administered tests, etc.), and then, according to the self-reflection literature, making a special effort to take the resulting feedback seriously rather than discounting it: to reflect rather than ruminate (Eva & Regehr, 2005).”

Eva & Regehr, in their survey of self-assessment literature from a variety of disciplines, defend their opinion that human cognition contains so many protective and insulating biases that accurate self-assessment is impossible without concerted and coordinated efforts to gather external feedback and evaluation through self-directed assessment seeking (Eva & Regehr, 2008). Despite this, many physicians are not receiving feedback. In a national survey of US physicians in 2003, only 33% of respondents reported receiving any feedback on the quality of their care (Audet, Doty, Shamasdin, & Schoenbaum, 2005).

Providing performance feedback to clinicians is supported by the principles of adult learning theory which holds that adult learners are self-directed, intrinsically motivated and most interested in problem-centered approaches (Kaufman, 2003). A constructivist approach to learning builds on prior experiences and knowledge. The cognitive dissonance created when there is a discrepancy between a clinician’s impression of their performance and data describing that performance contains the promise of a powerful motivator for change. This motivation is regarded as largely intrinsic, playing on a professional’s sense of achievement and desire to pursue excellence. Physician acceptance of systems that provide individual performance feedback increases when it is shown that those systems result in improved clinical outcomes.
For example, although initial opposition to the publication of individual performance data amongst cardiac surgeons in the UK was high, as the program became associated with improvements in mortality, the proportion of surgeons supporting the program increased significantly (Maytham & Dessaris, 2011).

The crucial link between feedback and behavioral change and the link between observation and learning is thought to be reflection (van de Wiel, Szegedi, & Weggeman, 2004). Reflection, as mentioned previously, is the process of trying to understand a problem. Examples would include questions such as why some patient outcomes were poor or why a physician’s feedback is lower than expected or why there is more variation than expected between individuals or over time. This is likely to require an organized deliberate approach and commitment. This type of experiential learning, which is based on the analysis of past results, was termed active learning by Simons (Simons, van der Linden, & Duffy, 2000) and deliberative learning by Eraut (Eraut, 2004). This may be more difficult for physicians when considering a range of cases within their practice compared to the immediacy of a need to learn while providing care for an unusual or difficult case.

There are a number of steps that must occur if individual feedback is to change provider behavior, and reflection is required at many of these steps (Brydges & Butler, 2012; Sargeant, Mann, van der Vleuten, Cees P., & Metsemakers, 2009). The provider must accept the feedback as an accurate reflection of their performance and accept responsibility for their performance (Myers et al., 2008). Providers must be able to set goals for improvement, determine which behaviors are most likely to result in improvement and have adequate power within the system in order to improve their performance. Some studies suggest that front line clinicians have considerable ability to improve quality via behavioral and practical systems changes (Porter & Teisberg, 2007) while other studies suggest that physicians may feel they lack the ability or resources to improve performance because of contextual or system challenges (Sargeant et al., 2009). Individual practitioners may not have enough power within their organization to make the system changes required to improve outcomes, no matter how much feedback they receive. Therefore, among the critical elements linking feedback and performance improvement are the resources needed to make improvements and the motivation to carry out the work required. There may be a balance of individual and system change, depending on the context, which would be required to maximize improvements in response to quantitative feedback.
Finally, providers must be able to self-monitor and have access to on-going feedback so that they can adjust their objectives and improvement and learning strategies appropriately. These learning activities take place within layers of context which may support or impede the process (Brydges & Butler, 2012).

Ajzen, in his theory of planned behavior, linked an individual’s attitude towards a behavior, their perceived behavioral control and subjective norms regarding the behavior, to their intention to perform the behavior (Ajzen, 1991). Behavioral intent has been found to be a strong predictor of actual behavior. Performance feedback has the potential to influence behavioral intent by shifting attitudes through linking clinical decisions and outcomes and by altering the perception of group norms. Communicating an individual’s performance in comparison to other members of a peer group or a group average is one way of altering the perception of group norms.

The concept of perceived self-efficacy has been used to help explain differing levels of career pursuit and human responses to failure (Bandura, 1982). Early success and observation of peer success both increase the self-efficacy of adult learners (Kaufman, 2003). Performance feedback that includes comparison to a peer group could increase the self-efficacy of an individual who self-identifies with other members of that peer group even if their performance is below average by comparison as it provides an example of what results are achievable.

Multi-source or 360-degree feedback (MSF) programs provide physicians with information on their clinical practice gathered from self-assessment, and questionnaires administered to peers, coworkers and patients. Physician ratings are generally given in the form of averages and standard deviations. The majority of physicians exposed to MSF assessments report contemplating behavioral change (W. Hall et al., 1999) and in one study, two thirds of physicians did initiate a behavioral change in response to their feedback (Fidler, Lockyer, Toews, & Violato, 1999). Factors that have been associated with an increased chance of behavioral change include personal characteristics of the physician (older age, female gender, non-surgical specialty), lower ratings on assessments, assessments from colleagues and patients and the availability of a mentor to stimulate reflection. A study of 28 family physicians participating in a MSF program showed that reflection occurred when feedback was incongruent with a physician’s self-perception. Reflection on negative feedback was seen as a difficult and emotionally charged activity and physicians in this study expressed a desire to have access to
facilitators to assist in the process (Sargeant et al., 2009). The perceived quality of mentoring was the most important motivational factor for intended practice change in response to MSF in one study (K. Overeem et al., 2012). This same group in an earlier qualitative study concluded that for MSF to be useful, it needed to be combined with goals, mentoring and structured follow-up (K. Overeem et al., 2009).

MSF differs from the type of quantitative feedback considered in this thesis in that it is not based on process and outcome metrics but from the impressions of oneself, colleagues, patients and coworkers. To date there has been no comparison between the effects of MSF and feedback from “harder” clinical outcomes on physician learning, behavioral change or improvement. The success of programs that provide quantitative performance feedback may be dependent on having systems in place that support learning. Whether goal-setting, mentoring and structured follow-up are as important to the use of process and outcome feedback as they are to MSF is an interesting question that remains unanswered.

2.6 Would performance measurement influence organizational culture to promote learning and safety?

Much of what individual providers can accomplish is influenced by context. Improvement in quality and safety cannot occur in the absence of organizational learning (Rivard, Rosen, & Carroll, 2006) and a supportive climate. In high risk industries such as aviation and oil and gas exploration, organizations that have a strong safety culture experience fewer failures (Hudson, 2003) and there has been significant interest in measuring, studying and supporting a safety culture in health care. The use of tools such as checklists and safety management systems has also been extrapolated from these industries to the health care setting.

According to Reason, a culture of safety is one that is informed, just, flexible and learning (J. T. Reason, 1998). A recent review of safety culture and climate in health care found that a common understanding of safety culture as a concept was lacking but that there were frequently cited elements common across studies. These included leadership commitment, open communication, trust, organizational learning, teamwork, non-punitive responses to reported adverse events and a shared belief at all levels in the importance of safety (Halligan & Zecevic, 2011).
Although quality and safety in health care may coexist, and indeed physicians, patients and payers hope that they do, they are not perfectly synonymous. Quality can be understood from several different perspectives (Garvin, 1984) with different stakeholders in the perioperative process emphasizing different aspects of quality. For example, safe care that occurs with minimal regard for human dignity is unlikely to be regarded as high quality by patients and their families.

Organizational learning requires errors and problems to be detected and communicated so that corrections can be made (Argyris & Schon, 1978). There is likely a common pathway through which organizations learn to provide safer care and through which they learn to provide care of higher quality: identification of a problem, research into possible solutions, dissemination of a defined better practice through education and process changes, monitoring of outcomes and refinement of strategy. These strategies include variations on the PDSA cycle.

Westrum developed a typology of three culture types: pathological, bureaucratic and generative (Westrum, 1991; Westrum, 2004) which was extended by Parker and Hudson to a five level model through the addition of reactive and proactive cultures (D. Parker & Hudson, 2001). In this model (Figure 2), organizations may mature from a pathological culture, which is preoccupied with power struggles, through several stages to an ideal cultural form – the generative culture. Generative cultures are highly cooperative and focused on performance and organizational goals. They embrace failure as a way to generate inquiry and learning. Case studies set in aviation, nuclear power and medicine support a relationship between this typology and safety (Westrum, 2004). The model has been used to create tools to assess safety culture in medicine (Kirk, Parker, Claridge, Esmail, & Marshall, 2007) and the oil and gas industry (D. Parker, Lawrie, & Hudson, 2006).

The cultural aspects that are associated with a strong safety culture are likely to reinforce social norms in an organization that would strengthen the quality of care. Parallels can be drawn between an organizational culture that supports continuous learning and improvement and the concept of “safety culture.” Some studies have suggested that organizational cultures that value teamwork and reward innovation have a higher capacity to improve compared to those that are predominantly hierarchical and bureaucratic (Shortell et al., 2000; Strasser, Smits, Falconer, Herrin, & Bowen, 2002). These same cultural aspects were found to correlate with attitudes
towards safety and safety climate in a survey of 100 adult and pediatric ICUs (Speroff et al., 2010).

Westrum’s typology uses the ways in which information flows in an organization as the basis for cultural assessment and predicts that increasing levels of “informedness” and trust between an organization’s members are associated with greater maturity of culture. Culture is felt to be heavily influenced by the way leaders shape communication, cooperation and their favored approaches to problem solving (Westrum, 2004). The IOM has called increased openness and sharing of information, particularly around failures, an essential component of building a safety culture in health care (Institute of Medicine, 2003). In health care, the flow of information surrounding failures is largely based on voluntary reporting and fails to capture a significant number of events (Cullen et al., 1995; Sari, Sheldon, Cracknell, & Turnbull, 2007). The power of a performance measurement system to capture failure events is increased by its non-voluntary nature. The use of leading and lagging indicators around identified barriers to potential harm or negative outcomes within the measurement system can give early notice of weaknesses in preventive capacity (Broadribb, Boyle, & Tanzi, 2009). Failure to follow procedures has been identified as a major contributor to error in the rail (Lawton, 1998) and aviation (Helmrich, 2000) industries and these studies also showed that those that commit procedural violations are more likely to commit other types of errors. Lawton and Parker went on to suggest that procedural violations were associated with attitudes and other social and psychological variables while other types of errors were associated with lack of skill or poor information processing (Lawton & Parker, 1998).
In the early days of performance measurement, greater emphasis was placed on providing information to managers, executives, payers and government agencies compared to feeding information back to front line staff. In Westrum’s generative culture, individuals are described as being aware of the variables that affect their efforts and results, aligned with the mission of improving outcomes and having a sense of ownership. This leads to innovation, empowerment and cooperation. High reliability organizations characteristically have structures such as reporting and information systems that encourage mindfulness in frontline staff (Friedrich, 2009). Performance measurement could support an organizational culture geared towards safety and learning if the information is effectively shared with health care providers.

When information from the performance measurement system is shared with front line workers, it has the potential to shift attitudes, decrease indifference to process and improve outcomes and
prevent other types of errors. When physicians and nurses are shown performance data, they are more likely to support quality improvement efforts and more likely to demonstrate improved performance (Amin, Fiorta, & Niesen, 2007; Donaldson, Brown, & Aydin, 2005; Mangram et al., 1999; Poniatowski, Stanley, & Youngberg, 2005). Impactful, low resource, innovations can come from front line workers so it is crucial to have them engaged with performance improvement efforts (Lee, 2010). Individuals tend to identify with, and modify their behavior to fit in with the norms of their team or work group (Hofmann & Stetzer, 1996; Zohar, 2000) and group norms were found to have a strong influence on both intentions to violate and actual violations in a study of armed forces aircraft maintainers and maintenance staff (Fogarty & Shaw, 2010). If front-line health workers see data that shows their performance in comparison to that of their peers, those with poor performance may find their attitudes towards certain tasks change as they attempt to increase their conformance with group norms.

Benchmarking and comparing data across peers and communicating those comparisons is likely to increase self-efficacy. Individuals who see their peers succeed are more likely to believe that they are personally capable of an equivalent degree of success (Kaufman, 2003). A common method of benchmarking is to use the mean or median for the group. However, this makes it possible to become complacent about performance in a very low performing group or institution. Benchmarking the data to an agreed upon standard as well as the mean gives individuals and groups important feedback in this regard (Cook, Coupe, & Ku, 2012).

The importance of leadership in creating and setting the tone for organizational culture, especially through responses to negative information, has also been emphasized (Edmondson, 1996; Westrum, 2004). Edmondson’s studies relating nursing supervision styles to the degree to which problems, errors and near misses are reported (Edmondson, 1996) and the effect of leadership on team learning (Edmondson, 2003) support the prominence leadership is given in Westrum, Parker and Hudson’s model of cultural maturation. Schein has suggested a model by which leaders can plan and manage cultural change (Schein, 2004) that begins with a period of disequilibrium or unfreezing in which data that is contradictory to organizational goals or vision causes anxiety. This anxiety, in a setting of sufficient psychological safety, allows learning to occur. Leadership can play a significant role in setting the cultural direction during the implementation of performance measurement in health care settings. Performance measures provide the contradictory data required for change; leaders provide a setting of psychological
safety that reduces learning anxiety in order to allow change to occur (Schein, 2004). Their first task will be making it very clear how the measures will be used, who will have access to them and what level of performance, if any, will have actionable consequences (Edmondson, 2011).

Leaders and champions for performance measurement and QI could be identified at multiple levels within the organization. At the micro level, clinical providers who act as champions could influence their colleagues’ attitudes towards and engagement with performance feedback. At the team or group level leaders could be instrumental during the implementation of improvement programs and at higher levels leaders may function by providing resources and incentives. Health care providers will change their behavior in response to performance measurement but, as they experience significant daily work demands, unless direction and resources are provided by leadership, they are likely to approach the problem using quick fixes, or what Edmondson has dubbed first order analysis (Edmondson, 1996). Larger and more reliable improvements will be generated by system and workflow changes (Deming, 1986).

Adler studied the performance improvement efforts of seven children’s hospitals to create a conception of an organization’s performance improvement capability (PIC) (Adler et al., 2003). The PIC is composed of an organization’s skills, systems, structure, strategy and culture. In this model, small improvements in PIC are easily accomplished with changes in skills and systems but larger changes require alterations in structure, strategy and culture.

A performance measurement system that uses performance feedback could increase PIC, however, based on their studies, improvements would be expected to be modest without investments in skills, structure, strategy and culture. Modest improvements that can be realized a short time after beginning a program of performance measurement are “early wins” that can be leveraged to increase the excitement and commitment of front-line staff (Edmondson, 2011).

Poor performance, whether on a process or outcome measure, could be considered a failure. Edmondson writes that the cultural dimensions that support safety are the same as those that support organizational learning from failures: open communication, teamwork, trust, and a non-punitive approach towards the bearers of bad news. Hospitals are complex systems where there is a great deal of uncertainty in work and where small process errors are likely to be unavoidable. A culture of psychological safety that counteracts the human tendency to assign blame when an error is identified is necessary for an organization to develop a learning culture around process
failures. An organization with a learning culture would constantly report failures of all sizes, systematically analyze them and search for opportunities to experiment with alternative approaches. Analysis of failures, even in hospitals and health care organizations where the consequences can be severe morbidity or mortality, is often cursory. Interdisciplinary teams may be helpful in analyzing failures when events in multiple areas, departments, disciplines or levels of the organization may have contributed to the outcome (Edmondson, 2011).

Edmondson has identified two organizational issues that prevent hospitals from learning from error. The first involves interpersonal factors that limit individuals from speaking out or challenging the status quo. This concern could be minimized if performance measurement using electronic data allows process errors and patterns of poor outcome to be reported automatically without any special human effort. The second is a culture that favors quick solutions or first order solutions at the expense of root cause analysis and problem solving (Edmondson, 2004). Even if information technology makes it easier to collect the data required for performance measurements it is unlikely that significant improvements in patient care will be made if attention and resources are not dedicated to analysis and problem solving.

One type of first order problem solving is known as the “work-around,” which is a deviation from usual process in order to bypass a block in work flow. In a work place with a climate of psychological safety and teamwork, front-line staff would more commonly bring work-arounds and process problems to the attention of their group and managers. Workplaces that have a strong orientation towards continuous quality improvement and that empower their employees to participate in, and influence, change would also be more likely to identify actual or potential work-arounds and a greater degree of second- and third-order problem solving. Partial support for this hypothesis was found in an exploratory study of work-arounds and climate within a cancer registry program (Halbesleben & Rathert, 2008).

Chuang et al developed a multi-level model of learning from preventable adverse events in health care (Chuang, Ginsburg, & Berta, 2007). Poor performance could be considered a “preventable adverse event,” from the perspective of the feedback target whether that is an individual clinician, or a group. The design and structure of feedback may affect several key elements of their model such as the perceived salience and causes of the failure as well as the attribution process. Whether an organization or group decides to engage in performance
feedback, and how that feedback is delivered, are both a reflection on organizational culture and leadership and these, in turn, may be reciprocally affected as a consequence of ongoing performance measurement.

Front line health professionals who are not given resource support by leaders to work on QI may start to view their performance metrics as being non-actionable. This may be especially true for performance metrics that have group or shared attribution where improvements are likely to require group solutions. The resources offered to support this work include dedicated time, extra training in epidemiology, process analysis and quality improvement (Adler et al., 2003), and mechanisms that improve communication. The literature suggests that structures that support team learning are important for organizational learning (Edmondson, 2002) and that team learning requires inspiration and organization from leadership (Dukerich, Golden, & Shortell, 2002). Group learning projects, where multiple members work together to shift attitudes and develop new ways of working, decrease learning anxiety by decreasing individuals’ fears of losing group membership during periods of change (Schein, 2004).

Teams of health care workers that are engaging in quality improvement projects are likely to require tools or training in effective communication. Dysfunctional communication is common in many areas of medicine and has been described in the physician-nurse, attending-resident and physician-physician relationships (Goldman, Lee, & Rudd, 1983; Makary et al., 2006).

Some researchers argue that individual performance measurement may destroy team-work and result in anxiety and that this would be minimized if the unit of measurement is the team level. (Bowman, 1994; Deming, 1986; Prince, 1994; Çiçek, Köksal, & Özdemirel, 2005). These writers have focused on individual performance appraisals conducted by human resources that have implications for remuneration and career advancement. Individual quantitative performance feedback that is confidentially provided and not associated with extrinsic motivators may not have the same effect on teamwork although this remains speculative. In their proposed model for providing teams with performance measurement, Çiçek et al., discouraged the attachment of extrinsic motivators on the grounds that this would lead to distortion of the measurement process and have deleterious effects on poorly functioning teams (Çiçek et al., 2005).
The idea that a program of performance measurement requires a robust learning culture for maximal improvements to be realized has attractive face-validity. Performance measurement that occurs without supportive cultural underpinnings may elicit minimal compliance, fear of judgment or reprisal in front-line clinicians and gaming or manipulation of metrics.

2.7 Performance Measurement and Feedback can Result in Improved Care but Feedback Structure Matters

Without practice change increases in medical knowledge will never translate into improved quality of care and outcomes for patients - yet practice change is difficult to come by. A comprehensive review found that continuing medical education (CME) activities had, at best, a modest effect on practice (Davis, Thomson, Oxman, & Haynes, 1995). It is no longer controversial that measuring process performance can change physician practice (Andrews, Tingen, Waller, & Harper, 2001; Eisenberg & Williams, 1981; Garber, 2005; Lüders et al., 2010) and this principle has been used successfully with anesthesia providers (Frenzel, Kee, Ensor, Ridel, & Ruiz, 2010).

A systematic review of the effects of assessment and feedback on physicians’ performance identified 41 studies. Feedback alone resulted in a positive change in physician performance in 74% of these studies and feedback combined with another initiative, such as reminders or educational programs, resulted in a positive effect 77% of the time. Feedback was more likely to have a positive effect if it was delivered by an authoritative source, such as an administrative unit or professional group as opposed to a research team, and if it was delivered over extended periods of time (Veloski, Boex, Grasberger, Evans, & Wolfson, 2006). A Cochrane review of audit and feedback given to health professionals, including trainees, also showed an overall positive effect after examination of 140 studies. The effect size varied widely from no effect to strong effect in this group of very heterogeneous studies. The effect size was greater if performance baseline was low, the feedback came from a supervisor or colleague, was given more than once, given in both written and verbal form and if it included an action plan or target (Ivers et al., 2012).

In some studies, personal feedback did not result in a sustained behavior change unless feedback was on-going (Body, Fanikos, DelPiero, Phillip, & Segal, 1999). It appears that large quantities of feedback may be detrimental to behavior modification efforts. “Individualized feedback is
successful at changing behavior and maintaining that behavior change as long as those efforts are not supplanted by another and another feedback to the same practitioners” a phenomenon that has been termed “feedback pushback,” (Lubarsky, Glass, & Ginsberg, 1997; Vigoda, Gencorelli, & Lubarsky, 2006). If practitioners are given performance feedback referenced to a collective average score, changes in behavior could be explained by regression to the mean. It has been suggested that those scoring above the mean would have no incentive to further improve their practice (Windokun & Windokun, 2011).

The American College of Cardiology Foundation and American Heart Association were among the first clinician groups to embrace and implement performance measurement. Their Task Force on Performance Measures has developed a detailed methodology stating that, “measures must also be valid, reliable, actionable and measurable, and they must address a demonstrable gap in care” (Spertus, Eagle, Krumholz, Mitchell, & Normand, 2005). Health professionals and administrators need to be confident that their chosen metrics can be used to improve the health of patients. This includes a process by which metrics can be evaluated for continued relevance, cost-effectiveness and unintended consequences to ensure that they continue to meet the needs of the organization (Spertus et al., 2010).

Consideration should be given to creating metrics that provide a balanced view of quality according to the priorities of different stakeholders. The “users” of anesthetic practice broadly encompass four groups: patients, anesthesiologists, surgeons, and the health care organization (McIntosh & Macario, 2009). Fung and colleagues found that patients prioritize empathy and communication as markers of quality while anesthesiologists value clinical outcomes and efficiency (Fung & Cohen, 2001).

More work needs to be done to create measures for care that involve patient transitions between providers, care areas and even health systems and that capture shared accountability (Fisher, Staiger, Bynum, & Gottlieb, 2007; Spertus et al., 2010). Given the complexities of modern medical care and widespread use of a team-based approach, it is difficult to assign outcome measures to any one provider (Myers et al., 2008) and process measures may rely more heavily on the system and team environment than has previously been realized. With shared accountability comes the recognition that both individuals and groups must be able to act on the information provided by a measure and that structures to support such work must be put in place.
Reliability and measurability are objective characteristics that relate to the technical aspects of data collection and analysis. Validity may be somewhat subjective. Efforts to improve the validity of performance measures in the eyes of physicians have included using only measures for processes (tests or treatments) for which strong scientific evidence exists and allowing exceptions on the basis of clinical, patient or system contraindications (Spertus et al., 2010). Physicians have been quick to criticize measures and are unlikely to engage meaningfully with a performance measurement program that does not choose high quality measures. When physicians don’t see performance measures as valid they will resist mandatory programs and drop out of, or refuse to join voluntary ones (Greene, Beckman, & Mahoney, 2008).

The choice and construction of measures is likely to have consequences on both professional motivation and quality improvement efforts. The fascination that physicians as a group have with data has been successfully used to gain their support for improvement efforts (Lee, 2010). Reason felt that a critical component of safety culture was willing participation by workers in a safety information system (J. Reason, 1997; J. T. Reason, 1998) and the same may be true for performance measurement systems and learning culture.

Performance measurement and feedback can be designed to take advantage of extrinsic and intrinsic motivators for physicians and hospital management. Trisolini recently reviewed extrinsic and intrinsic motivators for physician behavior (Trisolini, 2011). Programs that reward performance, either through P4P initiatives or public reporting, appeal to extrinsic motivations such as financial considerations and recognition. These extrinsic motivators work well for rote tasks. For professionals engaged in complex tasks, intrinsic motivators such as feelings of accomplishment or pride, autonomy, respect and collegial relationships may be more powerful. Intrinsic rewards were found by Pink to be more important for incenting problem solving, creativity and the management of complexity (Pink, 2011). There is some suggestion that extrinsic rewards can have a negative impact on physician motivation (Wynia, 2009).

“Reward and punishment strategies do not produce knowledge; they produce fear and anxiety often leading to distortion of the data or the process (Lilford et al., 2004).”

The ACC/ AHA first conceived their performance measurement program as a way to harness intrinsic motivation for physician-designed QI initiatives but performance measurement has been increasingly used in ways that harness extrinsic motivators through a focus on reimbursement
schemes and public reporting and accountability programs. Managed care organizations have also taken an interest in performance measurement and quality reporting as a way to gain competitive advantage and market share (McIntyre et al., 2001). There has been much discussion about the value of publicly reporting scores on performance measures. One randomized controlled trial showed that public reporting of hospital performance resulted in faster change and improvement efforts compared to private feedback (Hibbard, Stockard, & Tusler, 2003). Public reporting increased motivation for change amongst hospital top management teams (Galvin & McGlynn, 2003). Hospitals have used performance measurement and benchmarking with peer institutions in an attempt to improve their public report cards (D. S. Brown, Aydin, & Donaldson, 2008). The concept that underlies public reporting is that market pressures and individual pride will be able to drive health care improvement. However, only modest improvements have resulted from public reporting (Berwick et al., 2003; Hibbard et al., 2003; Letherman, Hibbard, & McGlynn, 2003; Marshall & Shekelle, 2000). This may be due to institutions urgently responding to measures in an effort to maintain the good will of the consumer with the resulting efforts being poorly thought out, poorly executed or merely symbolic. Institutions may also believe that it is adequate to “hide in the mean.” A survey of hospital CEOs in California reacting to performance data in a setting where rewards or punishments were offered showed that the majority did not believe action was required unless they fell significantly below the mean (Luce et al., 1996).

The act of public reporting of performance may not be well aligned with the goal of quality improvement although it is aligned with the goal of increasing accountability and transparency in health care. Health care providers that are subject to publicly reported performance measurement may be motivated to change their behavior on the basis of positive pressures, such as pride, or negative pressures, such as shame, or unable to change their behavior at all due to system constraints. Pride is likely to work as a motivator only if the measure is seen as a legitimate reflection of performance. Otherwise anger is the likely result (K. W. Hall, 2010).

2.8 Unintended Effects of Performance Measurement and Feedback

When performance measurements, which are largely still imperfect, are used by outside agencies to apply rewards such as financial gain or freedoms) or sanctions (such as star ratings) to
institutions, there is danger that data will be misinterpreted or over interpreted. When this is perceived as unfair by institutions and individuals, manipulation and gaming may result (Lilford et al., 2004). A study of recent initiatives to improve health care in the United Kingdom concluded that paying for performance did not incent physicians to work with hospitals and payers on QI projects but that this cooperation was an important aspect of health system transformation (Doran & Roland, 2010). When performance data was shared with members of a group practice, improvements were realized even in the absence of external reward suggesting that internal motivators are important in this setting (Perednia, 2011).

The use of negative extrinsic motivators is associated with less mature safety cultures (Edmondson, 1996; Edmondson, 2011; Westrum, 2004) and may increase attempts to manipulate data (Spertus et al., 2005). Individual provider performance data that is kept private will distance providers from feelings of blame and shame (Werner & Asch, 2005). In contrast, stronger safety and learning cultures emphasize the importance of a shared vision, open communication and teamwork (Garside, 1999; Westrum, 2004), characteristics that support intrinsic motivation. Learning cultures are described as rewarding risk and innovation, although whether the rewards appeal to intrinsic or extrinsic motivators is unspecified (Garside, 1999).

One of the unintended consequences of using the frequency of a clinical test or treatment as the basis of a performance measure is overuse and inappropriate application. Physicians may apply tests and treatments in ways that would normally be considered clinically inappropriate in order to fulfill process measures and may also ignore or downplay the role of patient autonomy in choosing direction and intensity of care. If process measures are time-sensitive, the process may be applied prior to establishing a firm diagnosis in order to fulfill the measure. Performance measurement has resulted in overuse of antibiotics (Wachter, Flanders, Fee, & Pronovost, 2008) and premature and costly activation of specialist teams (Larson et al., 2007). A study of colorectal screening in the VA health system found that what looked like poor performance on the measure was in fact appropriate care as 90% of patients that were not screened either declined the test, had significant co-morbid disease that increased the risk of and decreased the value of the test, or failed to present for the test or appointments (Walter, Davidowitz, Heineken, & Covinsky, 2004). It is widely recognized that outcome measures need to be risk adjusted but the example above shows that process measures may need risk adjustment too aid interpretation.
as higher illness burden can be a legitimate reason for choosing not to pursue certain tests and treatment.

When performance monitoring is implemented, problems and errors may seem to have increased (Glance et al., 2011) compared to prior experience or data from audits. Health care culture has long supported hiding errors and problems (Edmondson, 1996; Edmondson, 2011). This may lead to a distrust of the system through the phenomenon of “blaming the messenger” or be misinterpreted as a true worsening in performance. There is also the potential for data to be released before it is understood or before sufficient quantities for risk-adjustment and analysis are collected. Leaders should be prepared for this phenomenon and actively seek to manage it by educating and adjusting the expectations of health care providers. A perception of early failure can decrease self-efficacy (Kaufman, 2003) and negatively impact attitudes towards the program. The fear of temporary incompetence is a component of learning anxiety (Schein, 2004) and an unveiling of errors and problems feeds this fear. The temporary incompetence that occurs as new skills and routines are learned and incorporated into practice can also have a direct negative influence on patient care and OR functioning and strategies to mitigate these negative effects need to be in place.

There are several examples of gaming or manipulation of performance metrics within the medical literature. When New York State began publishing mortality data for coronary-artery bypass grafting, improvements resulted. However, Green and Wintfeld found that 41% of this improvement was due to data manipulation such as inflating risk factors or changing definitions (Green & Wintfeld, 1995). Some surgeons would modify the procedure performed on patients they saw as high-risk to one that was not captured so that these patients would not influence their ranking (Carey et al., 1998). Public reporting of outcomes after percutaneous coronary intervention in several states decreased the proportion of patients that received this procedure compared to states with no reporting (Joynt, Blumenthal, Orav, Resnic, & Jha, 2012). Critically ill patients may be subjected to care that is futile in order to prolong their lives past an arbitrary reporting end point such as 30-day mortality (Shahian et al., 2011).

Providers often perceive that they are not given enough credit by risk adjustment models (Schneider & Epstein, 1996) and some react by avoiding patients they see as high risk (Burack, Impellizzeri, Homel, & Cunningham, 1999; Narins, Dozier, Ling, & Zareba, 2005; Omoigui et
al., 1996; Schneider & Epstein, 1996) including patients from marginalized or vulnerable groups (Werner, Asch, & Polsky, 2005). The overall effect on the health system from this avoidance can be an increase in health expenditures and lower overall health outcomes for patients (Dranove, Kessler, McClellan, & Satterthwaite, 2003). Physicians who do not trust the measures used, may try to improve their performance through “dilution,” by applying procedures and tests to healthier patients which will result in false “improvements” in the metrics over time (Hayward & Kent, 2008; Werner & Asch, 2005).

The system in New York involved public reporting of outcomes and was based on a system of extrinsic reward and punishment via changes in peer and public perception of the involved physicians and potentially changes in referral and operating volumes. By contrast the Veterans Affairs National Surgical Risk Improvement Program was designed on no-fault-no-judgment principles and was able to achieve improvements in patient outcomes without changes in case mix or risk profile (Khuri, Daley, & Henderson, 2002) and fewer attempts to game or manipulate metrics than what was observed in relation to the New York CABG registry. Measures of appropriateness of care may help to de-incentivize gaming behaviors and structured audits can identify attempts to manipulate data.

Another unintended effect of performance measurement could be the development “tunnel vision” (Shahian et al., 2011) where providers focus their efforts on improving certain measures to the detriment of other measured or unmeasured aspects of patient care (Cook et al., 2012). A greater focus on adhering to process measures rather than care as a whole may result in a misallocation of healthcare resources or minimal effort. The use of multi-dimensional measures such as radar diagrams or performance polygons which visually display several different, complimentary measures simultaneously may make it less likely that efforts to improve certain aspects of care will result in worsening quality of other aspects, including unmeasured ones (D. S. Brown et al., 2008; Cook et al., 2012).

Individualized feedback and performance measurement may lead to a loss of trust and a culture where providers feel spied on and judged and perceive threats to their professional autonomy (Archer & Macario, 2006; Sheldon, 2005; Vigoda et al., 2011). This can occur when the distinction between feedback and evaluation is blurred. It has been recognized that feedback is more effective when delivered with sensitivity and when techniques to limit emotional reactions
to feedback are employed (Ende, 1983). Providers are more likely to have negative reactions towards quantitative feedback when performance measures are publicly reported (Glance et al., 2011). In order to support trust between members of a health care organization, it should be very clear to all parties what the purpose of measurement is and consequences stemming from measurement should be consistent with the program’s intent. If the ultimate goal of a performance measurement system is to increase the quality and safety of health care delivery, that goal should be supported by its design, implementation and use.

A proposed consequence of performance reporting is that attempts to reduce variability in care processes could stifle innovation and creativity which may result in a reduction in the acquisition of new knowledge (Riskin & Campagna, 2009). Performance measurement in healthcare has been deemed “utilitarian” in that once measures are chosen, they are a predefined standard and the act of measurement becomes a tool by which to affect behavior and process change in order to meet this standard. Learning and discovery become secondary goals, if they occur at all (McMaster, 2004).

There is also a concern that performance feedback could contribute to what has been termed the “deprofessionalization” of medicine (Cockerham, 2007). Physicians have historically determined their own standards of training and licensing and set up bodies to enforce these standards and level sanctions when necessary. The increasing amount of oversight and regulation being imposed upon medical work by governments, insurers, corporations, patients and advocacy organizations is eroding this traditional professional autonomy. Performance measurement, if imposed by a non-physician group such as hospital administration or a large insurer, may further damage this sense of professionalism. On the other hand, performance measurement that is designed and implemented by physicians may increase group cohesion and professionalism (Trisolini, 2011).

2.9 Summary

The morbidity and mortality burden of surgical treatment remains high and several contemporary studies suggest that 20-50% of adverse events experienced by this population are preventable (Gawande et al., 1999; Kable et al., 2002; Kievit et al., 2010). Large scale efforts promoted by national surgical and anesthesiology bodies to gather process and outcome data in order to inform and monitor quality improvement efforts are underway. Payers, such as large medical
insurance providers, CMS, and some provincial Ministries of Health in Canada have implemented, or are planning on implementing, pay for performance schemes in an effort to improve quality and decrease health care costs. In some centers smaller scale efforts to provide individual anesthesiology providers with performance feedback are underway. These efforts are supported by the availability of low cost data through increasingly sophisticated AIMS and EHRs.

There has been quite a bit of attention regarding how to gather data to provide performance measurement to anesthesiologists but considerably less attention on the processes of reflection and clinician and organizational learning and potential enabling elements and barriers within organizational cultures.

Table 1 lays out some of the factors identified during this literature review that might influence the development, use and outcomes of performance measurement in anesthesiology and perioperative care at the organizational, team and individual levels. Understanding these factors and how they influence the use of performance measures in different settings is important because this knowledge can be used to guide the creation and implementation of performance measurement programs. Tailoring these programs to context will maximize effectiveness and limit unintended adverse consequences.

There are a number of critical challenges in the use of performance measures that will help determine their impact on practice and outcomes.

Firstly, there is potential for the different stakeholders to have differing motivations for engaging in such an effort; such motivations may be aligned or in conflict. These motivations are likely to inform the intent and design of the program. Performance measurement undertaken in order to satisfy an organizational accountability, is likely to look different than performance measurement undertaken for to support clinician and team learning. Programs created by the front line, even if efficacious for improving care, may not survive funding and resource uncertainty if they do not address the motivations, priorities and strategy of the organization. Programs created by management may not find traction amongst front line staff and clinical teams if they do not align with clinical concerns.
Leadership is likely to be important at all levels. At the board and executive levels, leadership chooses and prioritizes goals, allocates funding and resources and sets the tone and culture of an organization. Team leaders both influence team culture and norms, and act as liaisons between team members and higher levels of management. Informal “champions” and early adopters can strongly influence their peers. Clinicians show a type of leadership when they choose to assume responsibility for patient outcomes and for improving care.

Clear communication at a number of different levels will optimize the use of a feedback program. If the intent and nature of the program is not communicated to the subjects of measurement, buy-in and motivation may suffer. Feedback method and presentation is likely to influence both acceptance and use. Sharing information and experiences between peers, among coworkers, and team members is necessary for individual learning and dissemination of insights and new routines. Communication between teams and individuals and management is required to allow allocation of necessary and appropriate resources and recognition of improvement efforts – both of which are necessary for on-going work in this area. Breakdowns in communication could lead to wasted time and resources, misunderstandings, a worsening of organizational culture, the inability to improve care that involves multiple specialties, teams that are not cohesive or effective and gaming or manipulation of metrics.

Ideally a performance measurement system would lead to learning on multiple levels – organizational, team and individual. Organizational learning is associated with system changes and the creation of artifacts such as new routines, protocols, work-flows, standardized order sets, CDS and clinical decision trees. Such learning requires a planned approach to use of performance information, structures that support the design and implementation of QI projects and other dedicated resources.

The steps to team and individual learning from feedback may be similar and involve feedback acceptance, reflection, assessment or self-assessment and then the creativity, flexibility, time and power to experiment with solutions. Motivation, the availability of resources and organizational culture may facilitate or be barriers to this process.
Table 2.1: Factors at the organizational, team and individual levels that may have an impact on the development, use and efficacy of performance measurement in anesthetic care.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Organization/ Management</th>
<th>Team</th>
<th>Individual Clinical Anesthesia Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivations</strong></td>
<td>Transparency</td>
<td>Intrinsic</td>
<td>Intrinsic</td>
</tr>
<tr>
<td></td>
<td>Accountability</td>
<td>Extrinsic</td>
<td>Extrinsic</td>
</tr>
<tr>
<td></td>
<td>Regulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Market share</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Leadership</strong></td>
<td>Sets goals, strategy,</td>
<td>Team and group leaders to provide support,</td>
<td>Individual champions/ early adopters</td>
</tr>
<tr>
<td></td>
<td>tone and intent</td>
<td>leverage motivators, monitor effects and adjust</td>
<td>Top-down vs. bottom-up approaches</td>
</tr>
<tr>
<td></td>
<td>Define and communicate</td>
<td>measurement and feedback strategies accordingly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>extrinsic motivational</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>strategies</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Information about program</td>
<td>Team cohesion and functioning</td>
<td>Peer to peer (formal and informal structures)</td>
</tr>
<tr>
<td></td>
<td>(purpose and uses)</td>
<td>Interdisciplinary cooperation</td>
<td>To and from leadership</td>
</tr>
<tr>
<td></td>
<td>Information from program</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>to inform decision making</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning</strong></td>
<td>System change</td>
<td>Feedback acceptance</td>
<td>Feedback acceptance</td>
</tr>
<tr>
<td></td>
<td>Artifacts (order sets,</td>
<td>Assessment</td>
<td>Self-assessment skills</td>
</tr>
<tr>
<td></td>
<td>decision trees, CDS, etc.)</td>
<td>Reflection: analysis of failures, outcomes</td>
<td>Reflection: analysis of failures, outcomes,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Team power/ autonomy</td>
<td>outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Creativity and innovation</td>
<td>Self-efficacy, autonomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First vs higher order problem solving</td>
<td>Creativity and innovation</td>
</tr>
<tr>
<td><strong>Resources and Barriers</strong></td>
<td>Resource allocation</td>
<td>Non-clinical time</td>
<td>Non-clinical time</td>
</tr>
<tr>
<td></td>
<td>IT support</td>
<td>Autonomy</td>
<td>Autonomy</td>
</tr>
<tr>
<td></td>
<td>QI infrastructure</td>
<td>Access to QI, KT, resources</td>
<td>Mentorship</td>
</tr>
<tr>
<td><strong>Culture</strong></td>
<td>Sets tone, expectations</td>
<td>Group norms</td>
<td>Conformity with group norms and expectations</td>
</tr>
</tbody>
</table>

2.10 Purpose and Research Questions

Although performance measurement and feedback is becoming increasingly discussed within medicine and health policy, there have been very few studies of its use within the field of anesthesiology. Through a series of case studies, the use of performance feedback in anesthesiology will be explored. The goal is to provide a rich description of the quantitative performance feedback programs in use by anesthesiologists and then investigate features that support or impede usability.
• How is quantitative feedback used to support learning and quality improvement activities and what design features of the feedback programs have been found to be helpful in this regard? These include program design decisions such as level of measurement (group or individual), choice of measures, use of electronic data and systems, analysis and presentation of measures, and structures and resources supporting the measurement program. Do those who have had experiences with quantitative clinical feedback have suggestions that would make the programs more effective? Are there different effects at the individual and group levels of learning?

• How do anesthesia providers integrate quantitative feedback into their strategies for self-assessment and continuing education?

• How are systems that provide quantitative feedback perceived by the providers and what features lead to those perceptions?

• What contextual features support or hinder learning and change efforts in response to quantitative performance feedback and do these elements have varying effects at the individual team and organizational levels?

The literature review has suggested some contextual features that are likely to affect the use of quantitative performance feedback. These include organizational culture, leadership, communication, use of different motivators, availability of resources and presence of barriers.

It is hoped that an in-depth look at how these systems interact with their context will aid in understanding the potential and limitations of quantitative feedback for improving quality and safety within perioperative care. This study has implications for anesthesiology providers who are subject to individual measurement as part of regulatory and remuneration schemes and groups who are contemplating the addition of quantitative performance feedback to their strategy for quality improvement.
Chapter 3 - Research Procedures

I chose to conduct multiple and comparative case studies of centers where anesthesia groups are using performance feedback as part of their quality improvement strategy. Case studies allow for rich descriptions of the settings and interventions. The goals were to explore ways that performance feedback can be used to improve anesthesia practice, gather multi-level perceptions of such a system, identify contextual features that may facilitate or hinder use and describe how introduction of the system may have affected group functioning and quality improvement efforts.

3.1 The role of case study research in understanding quality improvement efforts

The literature contains many studies of the effect of performance measurement and feedback on quality improvement in clinical medicine. A recent Cochrane review found that, overall, audit and feedback to physicians resulted in small improvements but effect size was variable. Exploratory regression analysis was used to show that the format in which the feedback was delivered, when it was delivered and how often and by whom, appeared to be important determinants of “treatment effect.” The authors concluded that it would be valuable to design future studies on feedback in a way that would allow these contextual elements to be directly compared (Ivers et al., 2012).

Few studies have specifically sought to identify contextual features which promote feedback usability or attempt to describe their relationships with each other. This may be because of a bias in the medical literature towards the quantitative paradigm. Some writers have even gone so far as to suggest that the mechanism by which performance feedback results in improved care is unimportant as long as improvements result (Glance et al., 2011).

Such a view is overly simplistic and neglects some fundamental practical considerations. Innovations, even ones that have been shown to be effective, can succeed or fail based on the way they are implemented and the context or environment to which they are exposed (Berwick, 2008; Davidoff, 2009; Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004; Pawson, Greenhalch, Harvey, & Walshe, 2004). Health systems need to be concerned about the cost of the improvements they are buying compared to other strategies including costs associated with implementation, ongoing support, upgrades and longevity. Ignoring context could also result in
unintended and negative effects from the implementation of the innovation. Outside of changes to health outcomes, introducing practice performance feedback into a group of anesthesiologists may change group functioning and quality improvement efforts positively or negatively. This process of change is not captured by empirical studies of feedback’s treatment effect which attempt to remove contextual elements through controlling experimental conditions.

Case studies are a powerful methodology for identifying contextual elements within an organization and for understanding the structures and processes that underlie change (Baker, 2011). “A case study is an empirical inquiry that:

- Investigates a contemporary phenomenon in depth and within its real life context, especially when
- The boundaries between phenomenon and context are not clearly evident” (p 18) (Yin, 2009).

Through interviews, document analysis, observations and other methods, a rich and detailed picture can be created that can be used for theory development (Edmondson & McManus, 2007; Eisenhardt & Graebner, 2007). While experimental methods excel at determining if something “works” they generally fall short with respect to explaining “how” or “why” something works. For these questions, case studies and other qualitative methods may be superior (Greenhalgh, Russell, & Swinglehurst, 2005; Vandenbroucke, 2008; Yin, 2009).

Important contextual elements could be identified through drawing on the literature of feedback both in medicine and in other industries but this would require a significant amount of extrapolation and should serve as a guide only. Medical training is an intense period of socialization where group norms are enforced. These group norms may not be present in other industries and may not even be constant across different fields and specialties within medicine.

These case studies of the use of performance feedback in anesthesiology gathered different perspectives on this phenomenon through the inclusion of multiple data sources – interviews, documents, examination of archival material and observation. Inclusion of more than one case allows for identification of patterns across different settings and each case may reveal different and contrasting aspects of the interaction between context and the feedback system. Identification of contextual elements that impact feedback use will guide future experiments
designed to test which contextual elements are most important for maximizing beneficial effects and provide guidance to other groups planning to introduce quantitative feedback into their settings.

3.2 Data Collection

3.2.1 Sampling

This study used multi-level, or stratified, purposeful sampling (Collins, 2010; Patton, 2002; Teddlie & Yu, 2007). Anesthesia groups that are using group performance measurement for quality improvement or are experienced in providing individual quantitative performance feedback to clinical anesthesiology providers were sought. The strategy for identifying sites included word-of-mouth networking, a search of the grey literature for promotional material and news items regarding medical centers with innovative approaches to quality improvement, and through discussion with the director of the Anesthesia Quality Institute (http://www.aqihq.org/). There were no a priori requirements for site selection other than that the site must be providing quantitative clinical feedback to anesthesiology providers for the purposes of learning or quality improvement.

Once potential sites were identified, the chair of the department at each site was contacted in order to determine if the site’s quality improvement program and system of performance measurement represented a case of interest and to determine site willingness to participate in the study. Sites that were using performance measurement solely for the purposes of quality assurance, human resources management or to fulfill the requirements of a regulatory body or payer were not pursued.

This sampling strategy has both risks and benefits. Purposive sampling allows the identification and inclusion of cases of great interest but may not yield data that is representative of the phenomena. The broad goal of this study is to identify elements within the cases that will be important for further study and for constructing a framework that may help guide further research in this area. Extreme or deviant case sampling would also have been a reasonable strategy (Patton, 2002) but in this case there were issues with access. Firstly the totality of cases was unknown. Even if all anesthesia departments in North America that were using quantitative performance feedback were identified, it would be very difficult to rank them on the basis of
success as there is no objective standard in this area and success and failure are likely to be largely group-defined. Anesthesia groups may also be reluctant to draw negative attention to a failed performance feedback program and this would limit access to cases which have experienced implementation failure.

Contact was made with three anesthesiology groups and one commercial provider of anesthesiology services which staffs upwards of 10 centers across the United States. Two of the sites agreed to participate. There were marked differences between these two sites with respect to program design, resource availability and organization of health services which offered the potential to explore the interaction of contextual elements with program design and implementation.

The first case is a university-affiliated tertiary care center located in Canada. This site has been collecting and distributing provider-level performance feedback for over ten years. The feedback program was conceived and designed by the anesthesia group and has not received much attention or resources from management.

The second site is an academic anesthesia practice located in the United States of America. This site collects clinical data using a predefined sampling strategy from patient charts, direct observation of care and patient questionnaires. The data is analyzed and presented in the form of aggregate rates for the entire group. The program has also been running for over 10 years. The aggregate data is used in a well-resourced program of quality improvement which is strongly supported by upper management.

Within each case study, interview subjects were selected through purposeful sampling. This sampling strategy was chosen in order to maximize researcher access to insights and information relevant to the use of performance feedback in each setting. Targeted selection of key informants was felt to be a better method of sampling than random or volunteer selection as these individuals were more likely to have the information of interest and the likelihood of obtaining key insights was increased. The department head, QI director, and members of QI working groups or teams as well as front-line clinical anesthesia providers were recruited. Efforts were made to include an “outsider” perspective by searching for individuals external to the department of anesthesia who were familiar with the intent and use of the system. In the case of site I, this proved to be impossible as the program was not known outside of the anesthesiology department.
In the case of site II, the anesthesia program was embedded within a larger QI program throughout the institution with the same foundational principles. These cut across all clinical groups and levels of management so that no “outsider” perspectives were available.

At each site, the department chair served as the initial contact between the site and the research team and acted as a “gate-keeper,” to identify individuals for interviews and also suggesting others with interesting perspectives to serve as informants. This introduces the potential for individuals with negative or extreme opinions to be insulated from the study. In order to minimize the sampling bias that a gate-keeper may cause, key interviewees that are likely to have the most information were identified by their roles prior to discussion with department chairs. Snowball and opportunistic sampling was used during site visits to identify additional interview subjects, especially subjects which may have differing viewpoints and those who had no professional or academic ties to performance measurement programs.

The perceptions and experiences of anesthesia providers who have no special roles with respect to performance measurement or quality improvement but who nevertheless interact with it through the course of fulfilling their duties were of great interest. It was desirable to bring a strong provider and end-user-centric perspective to these case studies as the key questions revolved around “usability,” and impact. While the collection and reporting of practice information may be mandated by insurers, administrators or government, the providers are the subjects of measurement and generally the program’s intended targets. After all, it is their actions and their care that will be scrutinized, potentially judged and that will be subject to programs and efforts to improve it. Their experiences and attitudes towards the system are expected to be highly influential in terms of its success or failure. All clinical anesthesia providers at each site were invited to participate in the study via departmental email and additional subjects were recruited opportunistically during site visits.

It was anticipated that few individuals would volunteer due to the demands of clinical work and lack of remuneration and that those that do may hold stronger opinions and views that deviate further from the norm. This was felt to increase the potential for interesting information and tensions to emerge that would be in line with the exploratory purpose of this study.

If new questions emerged after analysis of the initial interviews, further data collection was planned via contacting the original interview subjects for clarification and follow-up questioning.
3.2.2 Site Visits

Site visits were undertaken to allow a detailed description of each site and its performance measurement system and quality improvement strategy to be formed through observation, informal discussion and document analysis.

The first site was visited June 3rd to 5th, 2013. Prior to the visit, correspondence occurred with both the department chair and the department’s administrative assistant in order to schedule interviews with key informants, inform the group of the researcher’s presence, allow documents to be retrieved and made accessible in advance, and identify informants from other departments. Three days were spent completing interviews, touring the facilities, observing the artifacts of region-wide QI initiatives, the process of data collection and reporting in the medical records department and on-site opportunistic sampling and recruitment of additional informants.

While data obtained from the purposeful sampling of individuals in leadership positions was considered complete when all such individuals had been interviewed, sampling of front line anesthesia providers was considered complete when saturation was achieved. Eight subjects, all practicing anesthesiologists, were recruited from this study site and their experiences reacting to and using their performance feedback were similar. In this case new information was not emerging from the final two interviews and it was felt that further attempts to recruit participants or a second site visit would not add substantial information.

A physical visit to the second site could not be undertaken because of travel restrictions due to a health concern within the timelines required for this project. Individuals from the second site were interviewed by telephone between September 23rd and October 4th, 2013. Opportunistic sampling could not be used in this case and attempts to use snowball sampling did not yield further interview subjects. A variety of perspectives were still obtained. Interview subjects from Site II had all at some point been involved with gathering data or designing quality improvement projects that used the data. The perspectives of those with no special connection to the quality data system may have differed but these individuals were inaccessible at the time of the study. Documents and information about the practice and setting were obtained from the institution’s web page and promotional material and also from the chair of the department and QI director.
3.2.3 Collection of data from documents and administration

Administrative data was used to gather information on the setting and type of organization including its size, number of clinical anesthesia providers, work sites, operating rooms (ORs) and out-of-OR service delivery locations as well as the scope and nature of anesthetic practice. The department chair at each site provided a description of how anesthetic practice is structured and organized. The quality improvement coordinator or director at each site provided information on QI and performance data system organization including details on data sources, collection methods, data cleaning, analysis and reporting. Staff members from the department of Medical Records at the first site were extremely helpful with respect to forming a clear picture of data collection, storage and analysis as they performed this function for the anesthesiology group.

Documents relating to the policies and procedures for performance measurement and feedback were collected. Archived documents were obtained to help determine changes in the program over time. Copies of the performance reports generated and distributed at each site were obtained for comparison.

Archival material, including census data and city and provincial or state government websites, was used to form a description of the community served by each anesthesiology group.

The form used to collect this information is included as Appendix A.

3.2.4 Collection of Semi-Structured Interview Data

Individual semi-structured interviews occurred with the department chair, quality improvement director or equivalent and members of the anesthesia quality improvement team and clinical anesthesia providers. These interviews lasted between 20 and 60 minutes and were audiotaped. All interviews were conducted by the same researcher (AB) and followed the interview guide in Appendix B. The goal was to elicit the different perceptions, perspectives and experiences of people within different professional roles with respect to the use of performance data for quality improvement and organizational learning. Individuals with different roles were asked the same initial questions. Additional questions were asked depending on their responses and experience which allowed new insights and ideas to be immediately pursued. This flexibility in interview structure allowed exploration, probing and pursuit of the perspectives and experiences that differed between informants.
Interview subjects were asked to identify other individuals that may have useful insights to share and these individuals were contacted during the site visit and invited to participate.

Overall, 15 in-depth interviews were conducted, eight at site I, and seven at the site II. They were conducted at a time that was convenient for the participants.

At site I, interviews with individuals holding administrative positions occurred during the work day while clinicians without supported non-clinical time generously agreed to meet either prior to or after they were released from clinical duties. Clinicians that were opportunistically sampled were interviewed during breaks in their clinical duties during the work day. This necessitated deviation from the script and prioritizing lines of inquiry due to limitations on available time. All interviews at this site occurred in meeting rooms, offices and boardrooms located on the hospital campus. These interviews were recorded using the voice memo application included on the iPhone 4 and immediately transferred to a laptop and password protected.

Interviews with individuals from site II were conducted by telephone through Skype (www.skype.com). These were scheduled outside of times where the participants had immediate clinical duties, generally during a non-clinical administration day supported by the institution. In one case, an individual was contacted at home as they did not have access to non-clinical time. These interviews were recorded using Callburner (www.callburner.com) and the recordings were password protected and stored on a secure laptop in a locked location.

3.3 Data Transcription, Management and Analysis

3.3.1 Interview Data Transcription and Management

All interviews were transcribed by the same individual (AB) as soon as possible after the interview. The process of transcription is one of selection as it is impossible to completely render all of the information on an audio recording into typed text (Davidson, 2009). Selection is informed by theory (Ochs, 1979). Bucholtz described two main types of transcription, naturalized and denaturalized. Naturalized transcripts contain elements not present in spoken communication such as commas, full stops and the addition of paragraphs whereas denaturalized transcripts do not contain these features and preserve the “ums,” “errs,” and pauses of verbal communication (Bucholtz, 2000). Oliver et al. have suggested a transcription continuum that
also ranges from naturalism to denaturalism but define the terms differently. Naturalistic transcripts preserve “as much information as possible” whereas denaturalized transcripts omit accents, pauses, and other “idiosyncratic elements of speech.” They suggest that denaturalized transcripts emphasize meanings over delivery and are suited to the analytic methods of grounded theory and critical discourse analysis (Oliver, Serovich, & Mason, 2005).

The transcriptions created for this study fall on the denaturalized side of Oliver’s continuum. The meaning and intent of the participant’s communication was felt to be more important than the way it was delivered. Transcriptions were verbatim without correction for grammar, smoothing of disjointed thoughts or omission of repetition but punctuation, stops and paragraphs were added. The addition of punctuation and paragraphs was based on pauses and transitions between different ideas in the participant’s speech. These decisions formed the first elements of transcript analysis as they involved separating speech into different ideas. Significant pauses, involuntary sounds and sounds associated with emotional reactions such as laughter were preserved as it was felt that these might flag responses about which the participant felt strongly or where they were making efforts to edit or censor their thoughts.

As all interviews and transcriptions were conducted by a single person, consistency could be achieved in the way that the material was handled.

All interview participants were offered the chance to review their completed transcripts in order to provide corrections, clarification in those instances where they felt their words did not accurately reflect their impressions and the opportunity to strike passages. Participants then released their transcripts to the researcher.

Where quotations from participants have been used to illustrate study findings, these quotations have been further denaturalized with grammatical adjustments and insertion of punctuation to minimize distractions for readers and to disguise participants whose characteristic turns of phrase or grammatical errors may have led to their identification.

3.3.2 Case Database

All information collected for each case was kept in a case database which was stored on a secure, password protected, laptop.
The site I database consisted of the following:

- Audio recordings of interviews
- Interview transcripts (clean), some including notes added by informants as part of the release process.
- Interview transcripts with hand coding and researcher notes (scanned documents)
- De-identified copies of all email correspondence with site members
- Electronic documents from medical records detailing the exact procedures to be used for data collection and coding
- A scanned copy of the most recent quality indicator report
- Scanned copies of field notes
- Scanned copies of other researcher notes – for example notes made during preliminary phone calls
- Archival data: screen shots of census data and hospital and governmental websites used to develop the site description

The site II database consisted of the following:

- Audio recordings of interviews
- Interview transcripts (clean), some including notes added by informants as part of the release process.
- Interview transcripts with hand coding and researcher notes (scanned documents)
- De-identified copies of all email correspondence with site members
- Scanned copies of field notes
- Scanned copies of other researcher notes – for example notes made during preliminary phone calls
- Archival data: screen shots of census data and hospital and governmental websites used to develop the site description

3.3.3 Analysis

The analysis began by considering each case study separately prior to comparing and contrasting. This structure was chosen for efficiency as the case studies were conducted sequentially. A significant amount of analysis of the first case was completed prior to the second site visit.
For each case, a description of the setting and characteristics of anesthesia practice was generated from data gleaned through observation, documents, archival records such as census data and the institution’s web site(s), and conversations with individuals in administrative roles as well as the interviews. The purpose was to develop a picture of the community the group serves, the resources available in the setting for accomplishing the work, and how the group organizes anesthesia services in order to meet its obligations. Study sites were described and compared in order to highlight similarities and differences. This contextual information helps the reader form a picture of the setting and anesthetic practice of each site. Several interviewees directly remarked on how different elements of their context influenced their ability to engage and interact with performance feedback.

Interview transcripts were analyzed through a general inductive approach (D. R. Thomas, 2006). Transcripts were coded by hand and a code book was developed. Codes were then consolidated and organized into themes. Each case was coded individually. The themes that were generated during the analysis of the first case were examined to see if they would also fit the second case or if different themes were more appropriate. During the stage of comparing and contrasting the cases, themes that crossed cases were developed and some themes that remained isolated to each case were retained.

A case narrative was prepared and discussed by the research team for each case. The team discussed categories and themes, and referred to the original material in the case study database where there were questions and disagreement and when plausible alternative explanations emerged.

A mix of techniques, as described by Yin, was used to analyze the case data (Yin, 2009). The general strategy mixed elements of relying on theory, developing a descriptive framework and looking for rival explanations.

Interview subjects provided a number of divergent perspectives of the same phenomenon and I endeavored to grant these different perspectives equal weighting in the analysis. It is possible that some of these perspectives are restrained by the individual’s leadership roles and an expectation of program promotion within them. For example, it may be that the director or lead of quality improvement may speak about the program more positively and show more
enthusiasm as a function of their leadership position while those that have no special role, but interact with the system, may have less incentive to depict their experiences in a positive light.

3.4 Researcher Reflexivity

As a practicing anesthesiologist researching the perceptions and experiences of other practicing anesthesiologists, I tried to be as cognizant as possible of my preconceptions and biases as well as how shared educational and clinical experiences would affect establishing rapport with interview subjects and may influence the types of information that surfaced.

I have no personal experience with receiving or using quantitative performance feedback. During my medical training I received verbal and written feedback from preceptors and verbal and video feedback after sessions in a high fidelity simulator. As a practicing physician I receive verbal feedback from anesthesia and nursing colleagues, patients, and my department chair during the annual review process. I became interested in the use of quantitative practice feedback because of a personal desire for more information on my practice coupled with limited knowledge of how to best collect, structure, and present that feedback.

With my interview subjects I have a shared background and knowledge from general medical training and specialized training in anesthesiology. My subjects and I have undergone a similar socialization process within medical and specialty training, have had similar clinical experiences, and share similar concerns about our anesthetic practices. This became apparent early on during conduct of the interviews and, as I suspected, was both a help and a hindrance. Rapport with my subjects was easy to establish based on shared background and experiences but their curiosity about my motivations, perceptions and future plans threatened to bleed into and contaminate the interview data. I needed to be mindful during the process in terms of what I shared and how I phrased questions and the examples I used to further illustrate questions that were confusing to participants. I deferred questions about my practice and future plans towards the end but never refused to answer as I had concerns that would ruin rapport and be interpreted as non-collegial or deceptive.

Subjects often communicated through clinical anecdotes and descriptions and because these clinical scenarios and their ramifications were familiar to me, I did not have to interrupt the interview asking for clarification or explanation the way a non-anesthesiologist interviewer
would. However, this may also have introduced an element of manufactured meaning on my part. This potential bias, which would weaken internal validity, made member checking and case review an important part of my analytic approach to the case study material.

3.5 Ethical Considerations and Consent

The study protocol was approved by the University of Toronto’s Health Sciences Research Ethics Board (REB), and the REB at site I. The REB at site II advised that their approval was not required. All participants understood that their involvement was voluntary, that they may withdraw at any time and that they may chose not to answer any particular question. The step of requesting transcript release gave participants an opportunity to add clarifying material or remove of information that they felt might pose some personal risk and to confirm that the level of de-identification was satisfactory.

All data collection occurred at times when anesthesiologists had no immediate clinical duties so that work in the perioperative areas was not disrupted and there would be no effect on patients or their families. Informed consent, both written and verbal, was obtained from all study subjects for the interview and for audio recording and all subjects had the opportunity to drop out of the study at any point without repercussions.

The privacy of participants was protected through de-identification. Each participant was assigned a three digit identifier through an on-line random number generator\(^1\). The record of participants, their contact information and assigned number was kept in an encrypted and password protected file on a secure computer. No third party was involved in transcription. Hard copies of transcripts were kept in a locked cabinet and will be destroyed upon study completion. Electronic copies of transcripts and digital audio recordings were kept in secure electronic storage as described in the Case Database section.

All interviews were conducted privately and information from interviews was not shared with other participants.

\(^1\) [http://www.random.org/](http://www.random.org/)
3.6 Judging the Quality of Case Study Research Design

Four tests that are commonly applied to judging the quality of studies in the social sciences, including case studies, are:

1. Construct validity
2. Internal validity
3. External validity

Construct validity refers to the ability of the researches to sufficiently operationalize concepts in order to decrease subjectivity in measurement. The concepts identified during the literature review stage and in developing logic models for individual and group practice feedback and quality improvement work in anesthesia formed the basis of the study objectives and questions.

Suggested methods to increase construct validity include data triangulation, review of transcripts and drafts by peers and key informants and clear explanations of the chain of evidence and data collection and analysis procedures (Baker, 2011; Yin, 2009). Multiple sources and kinds of data were during the conduction of each case study which allows for triangulation. For example, both informant perceptions and experiences and documents detailing data collection procedures can speak to the construct of feedback validity and reliability.

I created a narrative for each case, referencing all the collected data sources and drawing heavily from transcribed interviews to provide illustrative quotes. This formed the basis of discussions amongst the research team whose members brought different experiences, backgrounds and levels of expertise to the exercise. This diversity helped to decrease the potential myopia and biases I might have brought to the material.

Internal validity is weakened when the researcher makes incorrect inferences. In a case study, inferences are made whenever an event cannot be directly observed, for example when analyzing interview data and looking at historical events. Strategies for improving internal validity in case study research include considering rival explanations, looking for converging and diverging evidence, pattern matching and the use of logic models ((Yin, 2009). This study used a combination of these approaches.
A study with external validity has findings that can be generalized to other cases. While empirical research may have statistical generalizability, where findings shown in a sample may be inferred to, or generalized, to a population, case studies have analytic generalizability where findings are generalized, not to other cases, but to theory. In multiple-case study designs, replication occurs when multiple cases support the same theory.

Two very different cases have been selected here to look at the usability of practice performance feedback for clinical anesthesiology providers. The two sites that were chosen have different strategies for collecting, analyzing and using information and occur within dissimilar contexts. Should the information from these cases, despite their contextual differences, support the same theory, this replication would provide strong evidence. Should the cases support different theories, then the contextual differences may be acting as independent variables and altering outcomes or the theories may be contextually-bound. The case studies were planned with the requirement to collect richly detailed contextual information because of the predetermined importance of this towards analytic generalizability.

Reliability speaks to the ability of research to be replicated by outside investigators. That is, if these case studies were completed by another investigator, using the procedures described here, that the findings would be similar. The case study protocols have been described here in depth, including selection, sampling and participant recruitment and more information is available in the appendices which include guides for collecting information from observation, interactions individuals at each site, interview guides and required contextual information. The information collected from all sources has been collected into an electronic case study database which contains emails and other correspondence, scanned and electronic documents, field notes, audio recordings, transcripts, investigator notes and narratives as described previously.

One of the suggested strategies for enhancing the reliability of case study reports is for researchers to include the site’s actual name (Baker, 2011). However, the anesthesia community is small and the identity of individuals in leadership roles within departments is well known by other members of the community. Identifying the study sites by name will effectively de-identify a significant number of study participants.
4 Chapter 4 – Descriptions of the sites and their quality and data systems

The two anesthesiology groups studied have developed different approaches to providing quantitative practice feedback to their members. In order to explore how these different approaches and their contexts influence the utility of the data for learning and quality improvement, a detailed description of each study site and measurement strategy is provided.

4.1 Descriptions of Study Sites

Site I is situated in a medium sized city and is one of two major referral centers within a Canadian province. The majority of anesthetic care in this city occurs at two hospital sites that both provide a mix of in-patient and ambulatory surgeries. Data from these two centers is collected and pooled for quality reporting. Between them, these two hospitals have approximately 600 beds of which just over one-half are acute care medical and surgical beds. One hospital provides more ambulatory surgery and handles community-level pediatric surgery while the other is a trauma center that provides specialized surgical services such as neurosurgery, cardiothoracic surgery, vascular surgery, interventional radiology, electrophysiology and labour and delivery. The anesthesiology group also provides services for one ambulatory operating room a week at a women’s health center and three ambulatory operating rooms a day at private facilities that have received contracts from the publicly funded surgical program. Data from these external sites are not captured by the feedback system.

The second study site is an anesthesiology group located within a not-for-profit private hospital which is university affiliated and a recognized research institute within the United States of America. The group provides services for the 600 bed main hospital site and for an ambulatory surgical program at a secondary hospital. The hospital provides a full range of subspecialty anesthesia and surgical services and draws cases from a geographical area of 22 million. The hospital also provides highly specialized services that receive national and international referrals. There are at least three other hospitals within this geographical area but these do not provide the same degree of complex and specialty services. Table 2 provides a quick snapshot of the setting of each study site.
Table 4.1: Major Characteristics of Each Study Site

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Site I</th>
<th>Site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health institution type</td>
<td>A private, not-for-profit, tertiary center funded by the Ministry of Health.</td>
<td>A private, not-for-profit, tertiary and quaternary center.</td>
</tr>
<tr>
<td>Location</td>
<td>Canada, city of &lt; 1 million</td>
<td>United States, city of &gt; 2 million</td>
</tr>
<tr>
<td>Referral base</td>
<td>Provincial</td>
<td>Regional National and global for certain specialized programs</td>
</tr>
</tbody>
</table>
| Facilities where anesthesia services are provided | Two hospitals  
Women’s health center  
Two free standing surgical clinics | Main hospital  
Secondary hospital with ambulatory surgical program |
| Competition                          | Negligible                                                             | Three nearby hospitals that compete for some, but not all case types   |
| Facilities where data is collected   | Two hospitals                                                         | Main hospital and secondary center                                    |
| Case volume and mix where data is collected | ~24 000 surgical cases per year  
40% ambulatory | ~35 000 surgical cases per year  
75% ambulatory |
| Anesthesiologist remuneration        | Fee for service                                                       | Salary plus variable compensation                                      |
| Anesthesiologist group size and provider type | 30 physicians  
40 CRNAs                       | 50 physicians  
40 CRNAs                              |
| Learners in the ORs                  | Medical students from local satellite medical school  
Resident physicians from a training program in a neighboring city | Residents  
Fellows  
Advanced fellows |

At Site I, most anesthesia care is provided by a staff anesthesiologist working alone. Because the main medical school is in a neighboring city, the presence of medical students and residents in the ORs is sporadic. There is a local satellite medical school that sends learners to the site for anesthesiology practicum training. At Site II, the 50 specialty trained physician anesthesiologists supervise CRNAs, fellows and residents who provide significant amounts of hands-on anesthetic care. The concurrency ratio for supervision of CRNAs and fellows at Site II is 1.7.

4.2 Resources for Quality Improvement and Provider Learning/CME

Site I receives limited resources from the institution for quality improvement and learning. The chair of the department is responsible for ensuring quality of care and is supported by a stipend of 20% of a full-time equivalent. The department names a quality improvement coordinator who
also receives a stipend. This individual is expected to prepare a quarterly report describing the department’s QI efforts to the Senior Medical Officer and complete one quality improvement project per year. Extra training in the areas of physician professional development, leadership development and quality improvement is available from the health authority and province through online, self-study, and didactic delivery methods. Depending on the program, this may be complementary or require a fee. There is a limited amount of funding available to physicians attending these programs and limited relief from clinical duties to do so. The department spends one hour a week in meetings of which one meeting per month is a quality improvement round. This time is uncompensated. The quality improvement round at Site I, despite its name, follows the traditional format of a critical incident or morbidity and mortality round. During these rounds cases are presented that resulted in actual or near-miss morbidity or mortality and potential contributing factors and corrective actions are discussed by the group.

The health region, in an attempt to increase the emphasis and visibility of quality and safety efforts is soon to require every department and ward to collect and report metrics. These metrics must fit into the Quality-Cost-Delivery-Safety-Morale (QCDSM) framework adapted from Lean methodology. The information is expected to be widely shared in order to inform strategic decisions at multiple levels including the Ministry of Health, health regions, institutions and departments. The department of anesthesia plans to modify their data program to fulfill this requirement although it is not yet known what changes will be made or what effect this will have on the primary intention and function of the original program.

Site II positions itself in the market by attempting to create and maintain a reputation of being the “best at improving.” Quality programs are pervasive, well-resourced and supported by the institution’s executive team and board of trustees as part of the institution’s strategic plan. The institution has a quality and safety position at the vice president level and at least four senior leaders that are involved in quality and safety.

In order to support the board’s strategic vision, site II launched a center for improvement several years ago which provides education programs, research and statistical support and access to experts in fields such as knowledge translation, health services research, health policy and evidence-based decision making. This center runs a six month, part-time, course in quality improvement science. The department of anesthesiology has sent approximately 15 members,
including some CRNAs, to this course. Each participant plans and executes a QI project as part of their course requirements. There is also an advanced course aimed at giving faculty the skills to plan quality improvement projects with rigorous methodology suitable for publication in academic journals. A few physician members of the department of anesthesiology have completed this training.

The department of anesthesiology has set aside time to meet as a group to discuss quality data and QI initiatives weekly and monthly. As the providers are salaried, this time is compensated. Several members of the department are provided with non-clinical time to work on QI projects and maintain the data gathering system.

Quality improvement priorities are shifting at Site II. Firstly, there is an increasing focus on value projects – delivering quality care at a lower cost. The institution has had success improving measured aspects of care and there is recognition that there are diminishing returns with respect to the amount of resources required to lower an already low failure rate. Secondly, the focus on quality improvement has led to duplication of quality projects and initiatives between divisions and departments. The institution has started to take steps to increase interdepartmental rationalization of projects such as bringing together quality leaders from different divisions in order to encourage collaboration and break down silos. Thirdly, a large project currently underway aims to improve the entire experience from the patient’s perspective. This initiative arises partially from the external pressure of the Center for Medicare Services’ (CMS) pay for performance scheme that is related to experience and satisfaction scores, and partially from the institution’s desire to promote itself and increase market share.

4.3 Description of data and feedback systems

4.3.1 Program Inception and Development

The data collected and reported to the anesthesia department at site I are referred to as the “Department of Anesthesia, Quality Indicators,” or “DA-QI.” The origins of the program have become somewhat obscure since its inception in the late 1990s. It appears that the hospital office of quality improvement and risk management was trying to expand a reporting program it had in place for obstetrics and gynecology to other departments and approached the chief of the department of anesthesiology.
“I quite naively just said, ‘of course, why wouldn’t we do this?’ I didn’t think there was any question or controversy. I told the department we would be getting these statistics now on our individual performance and it would be blinded and they would be coming in three months...so we just started.” (Former Chief of the department and principle architect of the DA-QI)

Interestingly, this individual did not see the collection and provision of individual quantitative practice feedback as exceptional or as an innovation at the time. This awareness came after the individual formed connections with other departments while fulfilling an administrative role. The choice of measures, their organization and formatting of the reports was made by this individual who decided to collect data on individual physicians in order to take advantage of their “natural competitiveness.” A contributing factor to this decision may have been the experience of having a colleague plead guilty to criminal negligence after a catastrophic outcome related to violations of accepted hospital and anesthetic procedure. There was suspicion that provider-level data might have provided objective warning though revealing a pattern of below-standard practice.

The department of anesthesia at Site II has been collecting quality data for over 10 years. Collection of these quality indicators grew out of a vision held by the board of trustees who wanted to improve the quality of care, not because of a perception of low quality, but as part of a strategy to grow the market share and reputation of the institution. The board recruited a CEO with experience in the manufacturing sector and a strong orientation towards measurement and quality improvement in order to carry out this vision. One of the physicians in the hospital had taken a sabbatical to work with the Institute for Healthcare Improvement and was successful in helping the institution secure a grant to support QI efforts. The institution was already a leader in receiving grant funding and its considerable experience with successful grant applications and existing research infrastructure for data collection and analysis was favorably recognized during the grant competition. The grant allowed the institution to support faculty interested in pursuing graduate level training in public health with a concentration in clinical effectiveness and a QI component. One of the first faculty members to complete such training was from the department of anesthesiology. For a few years, until internal training programs were developed, many faculty members from various departments and divisions were trained in this way.

During the same time, the department of anesthesiology appointed a new chief who also had a strong natural inclination to define quality and performance through quantitative means. This
constellation of factors - a chief who was very motivated to obtain quantitative quality data, a top management team and board of trustees committed to making quality a priority for the institution, an influential physician recently trained in quality improvement, resources from the grant and the ability to train faculty – lead to the department of anesthesiology becoming the pilot department for a new system of quantitative quality information gathering, reporting and use. Similar systems spread to other departments within the hospital. In general, implementation within new departments was found to be a five year process.
Table 4.2: Characteristics of the quantitative feedback collection and reporting system at each site

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Site I</th>
<th>Site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of program</td>
<td>Nearly 25 years</td>
<td>Just over 10 years</td>
</tr>
<tr>
<td>Stakeholders involved in the design</td>
<td>Chairperson of the anesthesiology department Medical Records</td>
<td>Chairperson of the anesthesiology department Subcommittee of anesthesia department members Hospital executives Hospital board of trustees</td>
</tr>
<tr>
<td></td>
<td>Subcommitte of anesthesia department members</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hospital executives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hospital board of trustees</td>
<td></td>
</tr>
<tr>
<td>Stakeholders involved in modifications</td>
<td>Chairperson of the anesthesiology department Quality improvement coordinator Subset of anesthesiology providers Medical Records</td>
<td>Chairperson of the anesthesiology department Subcommittee of anesthesia department members All department members to varying degrees</td>
</tr>
<tr>
<td></td>
<td>Medical Records</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subcommitte of anesthesia department members</td>
<td></td>
</tr>
<tr>
<td>Budget source</td>
<td>Institutional through medical records</td>
<td>Institutional through resources earmarked for data collection, analysis and QI projects</td>
</tr>
<tr>
<td></td>
<td>Institutional stipends for QI director and department chair</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>All charts from two hospital sites that contain an anesthetic record and a PACU flowsheet Exclusions: neonates, endoscopy, ambulatory ECT, incomplete records</td>
<td>A subset of surgical cases - data abstraction from patient charts A subset of patients and families - satisfaction survey A subset of anesthetic inductions - observed and rated All serious adverse events</td>
</tr>
<tr>
<td></td>
<td>A subset of surgical cases - data abstraction from patient charts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A subset of patients and families - satisfaction survey</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A subset of anesthetic inductions - observed and rated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All serious adverse events</td>
<td></td>
</tr>
<tr>
<td>Data abstraction</td>
<td>Health records analysts from the department of medical records</td>
<td>Trained Research and Quality coordinators</td>
</tr>
<tr>
<td>Data reporting frequency</td>
<td>Intended quarterly published reports</td>
<td>Weekly and monthly for data related to QI projects and initiatives</td>
</tr>
<tr>
<td></td>
<td>Reports delayed, sporadic and sometimes cancelled</td>
<td>Quarterly for dashboards</td>
</tr>
<tr>
<td>Report design</td>
<td>Visual scatterplot: Scores for each provider indexed to the mean for each indicator No longitudinal information</td>
<td>Run charts and control charts showing longitudinal variation compared to benchmarks or goals</td>
</tr>
<tr>
<td>Data communication</td>
<td>Paper copies of the report are left for anesthesiology provider in their mail slot</td>
<td>Multiple media are used: presentations at meetings, emails, newsletters, web site content</td>
</tr>
<tr>
<td>Actions arising from data</td>
<td>Individuals may reflect on their reports and take appropriate action</td>
<td>Group brainstorming and discussion of possible drivers</td>
</tr>
<tr>
<td></td>
<td>On occasion, report content has inspired group review of topics in rounds</td>
<td>Formal quality improvement projects aimed at improving particular measures</td>
</tr>
<tr>
<td></td>
<td>No expectations for use from departmental or institutional leadership</td>
<td>Reporting to hospital executive team and board of trustees</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reporting to insurers/ payers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Selected measures are available on institution’s website</td>
</tr>
</tbody>
</table>
4.3.2 Measure Choice, Collection and Communication

The measures chosen for collection at each site are summarized in table 4.

**Table 4.3: Quantitative performance feedback metrics chosen by each study site**

<table>
<thead>
<tr>
<th>Site I</th>
<th>Site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unplanned admission from day surgery</td>
<td>Serious adverse events, per 1000 cases</td>
</tr>
<tr>
<td>PACU arrival oxygen saturation &lt; 90%</td>
<td>Respiratory events in PACU</td>
</tr>
<tr>
<td>Anti-nauseant given in PACU</td>
<td>Anesthetic induction score</td>
</tr>
<tr>
<td>Initial PACU pain score &gt; 7/10</td>
<td>Comfort/ pain in PACU</td>
</tr>
<tr>
<td>Initial PACU temperature &lt; 36°C</td>
<td>Compliance with surgical site infection prevention bundle</td>
</tr>
<tr>
<td></td>
<td>Compliance with antibiotic administration</td>
</tr>
<tr>
<td></td>
<td>Patient and family satisfaction with care</td>
</tr>
<tr>
<td></td>
<td>Discharge time from PACU after representative day surgeries</td>
</tr>
<tr>
<td></td>
<td>On-time start in the OR</td>
</tr>
<tr>
<td></td>
<td>ASA units/ FTE/ Day</td>
</tr>
<tr>
<td></td>
<td>Indicators compared by race and insurance status (measure of equity)</td>
</tr>
</tbody>
</table>

ASA: American Society of Anesthesiologists physical status classification
FTE: Full time equivalent
PACU: post-anesthetic care unit

Measures for site I were initially chosen by the department chairperson at the time and were selected to reflect “hard” outcomes in order to maximize individual attribution and promote actionability. At the time there was little published information to guide choices in this area and the process was described as “intuitive.” Although documentation from this time is not available, personal recollections point towards a preference for outcome measures such as blood transfusions, cardiorespiratory arrests within 24 hours of transfer from the operating room, unplanned admissions and mortality. An emphasis was placed on measures that could be easily gleaned by minimally trained human data abstractors. At one point nine metrics were being collected but the anesthesia department was asked to decrease the number because of resource constraints. The current metrics were chosen by the anesthesia QI coordinator and the chief of the department with limited input from the larger group.

The DA-QI reports are generated from data taken from patient charts by employees of the health records department. These individuals abstract the information that is required for reporting to the Canadian Institute of Health Information (CIHI). Collection of data for the DA-QI is piggybacked on this process. The data abstractors do not have clinical backgrounds. Incomplete data is a rare finding but if it occurs the patient is omitted from the program as the data
abstractors will not look for it in alternative locations. The health records department does not keep a record of how many patients are excluded from the program based on missing data.

The data abstractors are using data definitions that are not regularly updated and this impacts data quality. For example, the definition of rescue anti-nauseant given in the PACU includes four commonly used drugs for this purpose, identified by their trade names. The data definition for this metric does not include generic names, nor does it include other anti-nauseant agents that may be used. The department currently hasn’t identified an individual responsible for maintaining the integrity of the data collection system. This lack of active leadership increases the risk that collected data will not accurately represent current care.

While the DA-QI reports are intended to be published quarterly, their publication has been frequently delayed and occasionally cancelled. The capacity of the records department to produce the reports has been reduced by a recent amalgamation with other hospital health records departments, increasing requirements for data abstraction to fulfill regional quality improvement mandates and human resource issues such as high staff turnover.

Hard copies of DA-QI reports are left for members of the anesthesia group in their departmental mailboxes. The reports provide the department with the number of surgeries performed that quarter, broken down by age group as well as a list of the 25 most common procedures and number of cases for each. The number of patients who experienced one or more of the indicator conditions, the total number of indicator events and events by indicator type are reported. Each anesthesiologist receives a report of the number of cases they performed at each of the two hospitals as well as statistics for ambulatory versus same day admission and inpatient surgery cases. A graph presents the total cases and number and percentage of all “indicator cases.” The number of these cases attributable to each anesthesiologist is presented in graphical form, compared to the mean (Figure 4). This is followed by graphs using the same format for each of the five individual indicators.

Each department member has been assigned an alphanumeric unique identifier (example, B1). It is not difficult for department members to identify the unique identifiers for the colleagues as they are assigned by working sequentially through the alphabet and numeric system as new members are added to the department. For example, physician A is a longstanding member of the department, while physician X is a newer addition. The identity of department members can
be revealed if they take a leave of absence as that results in a drop in case volume or a case volume of zero. Members with unique practice profiles, such as members nearing retirement that predominantly work in the ambulatory setting or who do fewer cases, can also find their anonymity compromised. The chief of the department has a key that links all department members to their unique identifiers. Other than receiving a hard copy of the DA-QI, no other methods of communicating the information to department members are used.

At site II the quality metrics were developed by a committee, presented to the department for feedback and further refined over a six to twelve month time period. Measurements were chosen for each of the IOM domains of quality (see Box 1) in an effort to reflect a balanced view. The department began by looking at safety measures in current use within quality assurance databases and efficiency and patient-centered measures described within the literature. They chose some metrics that were being used by pay-for-performance programs such as compliance with the surgical site infection prevention bundle. For several of the IOM dimensions of quality there were no existing or validated measures so the group created some. Metric design required balancing the anesthesiology department requirements with the resources required for collection and to act on the results. Each measure went through a pilot phase for testing and the results were presented to front line staff for feedback and refinement. The process of selecting or creating metrics, refining and testing them and designing dashboards and reports took approximately three years.

Some examples of metrics collected within the dimensions of effectiveness and patient-centeredness include patient and family satisfaction with the preoperative information given to them, pain scores in post-anesthetic recovery unit, respiratory complications in the PACU, and unanticipated admissions after ambulatory surgery. In the patient safety dimension, measures include the rate of serious adverse events and compliance with the surgical site infection prevention bundle. Within the work flow and efficiency dimension, there are three metrics: a measure of discharge time from PACU after selected day surgeries such as pediatric tonsillectomy, proportion of on-time operating room starts and a measure of work complexity and volume (ASA/FTE/day).
All of the measures capture the performance of the anesthesia providers as a group. Individual clinical providers who are out of compliance with some of the process measures will receive private reminders through email or conversations with quality leaders but reports on individual compliance are not kept by leaders or communicated to providers. Outside of the dashboard other metrics are added to answer particular questions or guide improvement projects. These are chosen by leadership in alignment with departmental and business priorities. Metrics are only chosen if there is a reasonable expectation that they can be improved by actions within the realm of influence of the department of anesthesia.

For metrics that can be gathered from the patient’s chart, such as respiratory complications and pain in the PACU, human abstractors are used to gather the information on every case in the first two weeks of each quarter. This amounts to approximately 10% of all cases. This sample size was determined after considering case variation and volume. Patients that go directly to intensive care units from the operating rooms are excluded. The research and quality improvement coordinators that abstract the data have been trained to collect information according to strict definitions. These individuals also collect data for research studies and quality improvement projects. They enter the information by hand into a database from which predefined and ad-hoc reports are produced.

Some of the measures require direct observation. The research and quality improvement coordinators observe 10% of all cases that occur during regular working hours using predefined rules that aim to ensure a diverse case mix. Afterhours and weekend cases are not sampled but emergent cases that occur during daytime hours are included. There is ongoing reconciliation and weekly auditing and multiple methods are used to capture some of the data of interest. For example, the charts of all patients who have a “code” called in the PACU are reviewed by the QI coordinator for the site to make sure that those events are accurately identified and captured within the metric for “serious adverse events.”

Site II went through a departmental goal setting process for each measure. Some goals were generated from benchmarks that were available in the anesthesiology literature but many of the metrics chosen were new and for these goals were developed by consensus:

**Box 1: IOM dimensions of quality**

<table>
<thead>
<tr>
<th>Safe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective</td>
</tr>
<tr>
<td>Patient/ Family-centered</td>
</tr>
<tr>
<td>Timely</td>
</tr>
<tr>
<td>Efficient</td>
</tr>
<tr>
<td>Equal</td>
</tr>
</tbody>
</table>
“Looking at the key drivers of the process we’d say perhaps we could reduce this by 50 or 60% based on real knowledge of the process and the system in which we were working. For some it was more consensus of talking to faculty…we said, ‘what do we want for ourselves…If I was a patient, what would I want that goal to be?’” (0295, site II)

“You know, most people, as professionals want to function at a very high level, over 95%. But in reality, as a first pass you can rarely do that. So you have to be talking about what you think it should be and what you think it can be within this time period.” (0618, site II)

The metrics are presented via a quarterly dashboard which relates the performance of the group to baseline data as well as goals or established benchmarks. A run charts or control chart is produced for each metric and provided with the dashboard (Figure 4). These reports are reviewed by the quality improvement coordinator, the division and section chiefs and the chair of the department. They are presented monthly to all anesthesia providers during department meetings. Quality metrics are not reported separately for physicians and CRNA provider groups. The quarterly dashboards are summarized on a yearly basis within the chief’s report on department activities and via periodic newsletters sent to the department. The dashboards and run charts are available to all employees through the intranet and are also presented twice a year to senior leadership and the board of trustees.

The quality reporting system at site II is also used to gather information related to QI projects. This information is presented in the same format as the quarterly quality dashboard. The current score or rate is indexed to both the historical mean and goal and presented in run or control chart format. The timing of publication and presentation of this data varies depending on the project’s nature. Examples were collected from this site where such data is presented daily, weekly, monthly, quarterly and yearly.

Neither site was leveraging the power of electronic health records to reduce the cost of data collection. The department of anesthesiology at site I was experiencing difficulties with obtaining IT resources and significant barriers to acquiring, customizing and supporting new applications, including an electronic anesthesia record. Site II has recently implemented a new hospital enterprise information system which has given them the opportunity to move towards automatically populating their database with electronic information. This will remove the need for human data abstractors and allow the group to increase their sample of cases. As they work through the data design and validation process it has become clear that the way the measures
have been collected in the past is not directly transferable to the new electronic charting fields. In some cases the anesthesiology group has been able to work with nursing management to change electronic charting to facilitate their data needs. In other cases the data definitions for some metrics will need to be changed to facilitate the use of electronic information. The group expects that they will have to wait until a new steady state is achieved in 6-12 months before making decisions based on the quality data collected automatically from the electronic chart.
4.3.3 Use of the performance feedback systems

Site I’s DA-QI program was designed and intended as “quiet personal leverage.” The leaders within the department expressed the opinion that the data provided was useful and that most department members would utilize it out of a professional obligation to work at improving their clinical care. There are no supports provided for mentoring the processes of reflection and practice change that leaders expect to occur around DA-QI reports. It was the impression of three of the department members interviewed that roughly half of their colleagues did not read their reports. Reasons that emerged from interview data for minimal report usefulness included problems with salience, relevance, interpretation and limited time and ability for practice change.

Within the first year after the program began it was noticed that individual performance variations were decreasing. The chief of the department at this time and the current chief of the department both attributed this reduction in variation to individuals modifying their performance to fit in with peers.

There are no formal programs set up to analyze and use the data at the departmental level. At times the QI coordinators, and other department members, have supplemented DA-QI report data with chart audits and reviews for the purposes of grand rounds presentations. This was more common during the early years of the program when specific measures were sometimes discussed at department meetings.

There have been occasions where the chair of the site I department, who has access to non-anonymized data, has used it as the basis to start a dialogue when gross deviations in individual performance are evident. One member, in response to a credentialing or licensing issue, wanted to use the data as evidence that he/she provided an acceptable level of care. The department, in that case, adopted the position that these reports should be used for learning and improvement and that the information contained within them be treated as confidential in the same manner as information shared within quality improvement and morbidity and mortality rounds.

Within the health region, the QI coordinator for each department is responsible for sending a quarterly report on quality improvement activities to the regional senior medical officer. The department of anesthesia includes some of their metrics as part of this report. This report does not result in significant dialogue between the department and institutional leadership on goals for
anesthetic care quality. The quality indicator reports are not shared with other departments within the institution or made public. There is very little awareness outside of the department of anesthesiology that a program providing individual practice feedback to anesthesiologists exists.

In contrast, the metrics collected by Site II are extensively used in learning and QI activities. The department discusses quality measures and information frequently within meetings and rounds. The dashboard is discussed each quarter and some of the metrics on it are discussed at monthly and weekly meetings along with other measures related to active quality improvement projects. Each metric is compared to an internal goal or external benchmark and when a deviation occurs, principles of continuous quality improvement (CQI) are used to try to effect change by identifying key drivers and planning interventions using the PDSA cycle. The anesthesiology department at site II currently has the resources required to perform four to five large QI projects a year. Most of these projects are expected to run for approximately three years which has been the length of time, in their experience, required to cement process changes into group practice.

Feedback to anesthesia practitioners is integral to these quality improvement projects and occurs at meetings where results are discussed and privately to individuals who are found to have deviated from expected processes. These individual conversations are initiated by individuals in leadership positions within the quality system. The feedback reminds practitioners of important initiatives, communicates the institution’s commitment and alerts the practitioner to episodes of non-compliance. These conversations are seen as opportunities to understand circumstances around failures and help clinical providers improve. Because the feedback flows both ways, when practitioners supply information about the circumstances surrounding the failure, system issues and barriers are identified so that they can be addressed. This individual feedback is private and does not influence yearly performance appraisals.

Information from the quality dashboard is used to maintain a quality and accountability dialogue between the department and upper management within the institution. Success stories from improvement projects are used to help market and promote the institution. Quality improvement has been recognized as an academic track within medicine by the organization. The information collected for the dashboard and various QI projects has resulted in publications and presentations
at conferences which further the academic careers of the involved individuals and add to the institution’s reputation.

Site II sees their system of measuring quality indicators as complementary to other methods of quality and safety improvement such as adverse event reporting. The department of anesthesiology has a quality improvement director for each of their two surgical facilities and these individuals review adverse events individually on a weekly basis and monthly with a committee. The department also holds the equivalent of morbidity and mortality rounds as part of the strategy to address these incidents.

4.4 Summary

Both sites which were studied are tertiary health care institutions associated with university medical training programs. There were notable differences between them with respect to design and structure of the quantitative feedback programs. While site I designed a program to give individual providers quantitative feedback for personal use, site II collected group level performance data to be used by an organized quality improvement program. Site II was subject to greater external and internal pressures to collect performance data and this resulted in the involvement of leadership at multiple levels and formalized structures for accountability. Leadership at site II was able to provide institutional resources so that the department of anesthesiology could meet these expectations. A combination of heightened expectations and resources seems to have contributed to greater use of performance data for learning and improvement at site II compared to site I. Table 5 summarizes the key differences in the structure and context influencing the feedback programs at site I and II while figure 3 shows how the information was used schematically.
Table 4.4: Summary of key differences between sites

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Site I</th>
<th>Site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of measurement</td>
<td>Individual</td>
<td>Group</td>
</tr>
<tr>
<td>Feedback delivery</td>
<td>Written reports</td>
<td>Written reports, presentations, verbal</td>
</tr>
<tr>
<td>Feedback frequency</td>
<td>Quarterly (in practice less frequent with some reports delayed by 12 months or more)</td>
<td>Daily, weekly, monthly, and quarterly depending on specific feedback type.</td>
</tr>
<tr>
<td>Program design team</td>
<td>Select individuals within the department of anesthesia with very limited group input</td>
<td>Select individuals within the department of anesthesia with input from management/ administration and wider group of department members. Pilot phase for new measures.</td>
</tr>
<tr>
<td>Pressures/levers contributing to creation of feedback program</td>
<td>Internal: provider curiosity and motivation for improvement. Recent experience with care below standard by one department member with significant consequences</td>
<td>Institutional: seen as a way to position organization within the market External: data collection mandated by national and federal programs Internal: Provider curiosity and motivation for improvement</td>
</tr>
<tr>
<td>Sources of program leadership</td>
<td>Chief of department&lt;br&gt;Quality Coordinator</td>
<td>Chief of department&lt;br&gt;Quality coordinator(s)&lt;br&gt;Quality Improvement office&lt;br&gt;VP Quality&lt;br&gt;Organizational management team&lt;br&gt;Organizational board of directors</td>
</tr>
<tr>
<td>Resources</td>
<td>Limited&lt;br&gt;Funding for data collection and analysis not dedicated. Clinicians expected to learn and improve on their own time</td>
<td>Adequate.&lt;br&gt;Dedicated funding for data collection, analysis and to support clinicians in quality work</td>
</tr>
<tr>
<td>Accountability/expectations and their source</td>
<td>None formalized.&lt;br&gt;Chief of department assumes metrics will be used by individual physicians for learning. No consequences for information use or disuse</td>
<td>Formalized.&lt;br&gt;Department required to collect and use performance data by institution.&lt;br&gt;Achievement of group goals tied to variable compensation package for providers</td>
</tr>
<tr>
<td>Supporting organizational structures</td>
<td>Limited.&lt;br&gt;Some capability of medical records staff to assist in further information gathering Limited leadership and QI training available from health authority and province.</td>
<td>Exceptional.&lt;br&gt;Organizational quality institute with resources for data collection, analysis, QI science training, knowledge translation and change management.&lt;br&gt;Organizational training programs in leadership and QI.&lt;br&gt;Organizational recognition of QI as academic stream resulting in potential for career advancement.</td>
</tr>
</tbody>
</table>
Figure 3a: Mechanisms of learning and change stemming from quantitative performance feedback at Site 1

**Figure 3a: Schematic representations of quantitative feedback use for learning and improvement at site 1**

- **Effect monitoring**
- **Provider does not engage with report content**
  - Lack of time
  - Dismissal of content (perceived problems with validity, relevance, sampling etc.)
  - No formalized expectations, consequences or mechanisms of accountability associated with the program
- **No action**
  - Report contradicts other more trusted sources of feedback
  - Lack of time
  - Lack of support/maintenance for learning and practice change
  - Lack of improvement science and leadership training
  - Difficulties and uncertainties in goal setting and/or benchmarking
  - Current practice found to be adequate
  - Solution requires system change
  - Time and effort to change outweigh rewards
- **Reflection**
  - Information gathering:
    - Discussions with peers
    - Literature search
    - Drawing from current knowledge
    - Charter review/audit
- **Outlier assessment**
  - Leader initiates intervention for outliers (e.g., hallway discussions, formal meetings)
- **Communication of findings to the group** (e.g., department meeting, roundtable, etc.)
  - Personal behavior change
    - Natural (before and after) experiments
    - Lack of resources
    - Lack of improvement science and leadership training
    - Cultural barriers
  - No process or system change
  - Process or system change
  - Lack of resources
  - Lack of improvement science and leadership training
  - Cultural barriers
Figure 3b: Mechanisms of learning and change stemming from quantitative performance feedback at Site II

Group receives quantitative feedback reports on performance

- Evaluation of metrics
  - Retain, modify or retire metric?
  - Pre-determined goal achieved
  - Pre-determined goal not achieved

- Personal reflection (limited)

- Self-initiated practice change

- Key driver process adherence measured along with outcomes. Process becomes linked to outcome

- Group level metrics do not aid in personal reflection
- Cultural tendency to view quality as a system property not strongly influenced by individual providers
- Individual achievement not measured
- Providers expected to conform to standards in structural practice which may limit experimentation

- Feedback delivery and source influence uptake
- Intervention from leadership may be required when providers refuse to conform

- Lack of buy-in (perceived poor relevance, salience, evidence for process)
- Distraction
- Inability to make multiple behavioral changes at once
- Improvement project and/or data fatigue

- Measurement of process adherence (individual compliance)

- Provider out of compliance with process
  - Individual feedback
  - Feedback from practitioners on root causes of process violations reviewed to inform interventions and system changes

- Provider in compliance with process

- Group-initiated practice change—compliance expected by department chair and administration

- Education: rounds, meetings, memos, other communications
- Process changes
- System and policy changes; creation of PPOs and clinical pathways

Barriers include resources for data collection and analysis and non-clinical time
Site I

<table>
<thead>
<tr>
<th>Provider</th>
<th>Case total</th>
<th>Cases with PONV</th>
<th>PONV (% of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>324</td>
<td>48</td>
<td>14.8</td>
</tr>
<tr>
<td>A2</td>
<td>134</td>
<td>21</td>
<td>16.4</td>
</tr>
<tr>
<td>C4</td>
<td>53</td>
<td>3</td>
<td>5.67</td>
</tr>
<tr>
<td>C5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D2</td>
<td>98</td>
<td>16</td>
<td>16.3</td>
</tr>
<tr>
<td>D3</td>
<td>224</td>
<td>42</td>
<td>18.8</td>
</tr>
<tr>
<td>E3</td>
<td>125</td>
<td>27</td>
<td>21.6</td>
</tr>
<tr>
<td>F1</td>
<td>326</td>
<td>21</td>
<td>6.44</td>
</tr>
<tr>
<td>F4</td>
<td>411</td>
<td>41</td>
<td>9.98</td>
</tr>
<tr>
<td>F8</td>
<td>157</td>
<td>12</td>
<td>7.60</td>
</tr>
<tr>
<td>G1</td>
<td>202</td>
<td>25</td>
<td>12.4</td>
</tr>
<tr>
<td>G9</td>
<td>256</td>
<td>21</td>
<td>8.20</td>
</tr>
<tr>
<td>H4</td>
<td>177</td>
<td>18</td>
<td>10.2</td>
</tr>
<tr>
<td>H0</td>
<td>290</td>
<td>38</td>
<td>0.10</td>
</tr>
<tr>
<td>group</td>
<td>2762</td>
<td>313</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Site II

<table>
<thead>
<tr>
<th>Quarter</th>
<th>PACU</th>
<th>Cases</th>
<th>Respiratory Event Rate (%) in PACU</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 1</td>
<td>15</td>
<td>567</td>
<td>2.6</td>
</tr>
<tr>
<td>2013 2</td>
<td>12</td>
<td>620</td>
<td>1.9</td>
</tr>
<tr>
<td>2013 3</td>
<td>8</td>
<td>598</td>
<td>1.3</td>
</tr>
<tr>
<td>2013 4</td>
<td>9</td>
<td>614</td>
<td>1.5</td>
</tr>
<tr>
<td>2014 1</td>
<td>10</td>
<td>588</td>
<td>1.7</td>
</tr>
<tr>
<td>2014 2</td>
<td>5</td>
<td>650</td>
<td>0.77</td>
</tr>
<tr>
<td>2014 3</td>
<td>6</td>
<td>603</td>
<td>1.0</td>
</tr>
<tr>
<td>2014 4</td>
<td>7</td>
<td>498</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Figure 4: Sample quantitative performance reports using mock data.

It was decided not to include reproductions of actual quality reports as there is the potential for them to de-identify the study sites. Site I’s performance reports contain tables showing each anesthesiology providers case totals for the quarter broken down by the hospital where the care was provided and by ambulatory versus admitted patients. The report is multiple pages long with the first page showing totals for the quarter followed by one page per indicator. For each indicator, a table is produced which contains the denominator, the rate of “flagged” cases overall and broken down by hospital site and ambulatory vs admitted patients. A graph is produced which shows the relationship between the provider’s rate and the mean rate for the group. Site II’s performance reports consist of a quarterly dashboard. Each indicator on the dashboard is related to the historical average as well as the goal. A control chart is produced for each indicator.
5 Chapter 5 - Findings and Discussion

While both sites were collecting data, there were significant differences in terms of strategies, use, and perceived effectiveness. Performance data was also used at both sites for quality assurance purposes – i.e. to fill obligations for reporting to various governing bodies such as credentialing programs or a governmental health region. Those types of uses will not be considered here. The first research question was how each site uses their performance data to support learning and quality improvement activities and how much of this activity is related to the data that is collected. This learning could occur at the individual or group level and both motivation for learning and change and how this feedback is integrated with feedback from other sources for self-assessment purposes. Participants were asked what design features of their program contributed to usable data – such as metric selection, analysis and reporting format, feedback frequency and what improvements they would suggest based on their experiences. The third question was how contextual elements, such as organizational culture, presence and quality of leadership, institutional structures for communication and learning and available resources influenced both program design and effectiveness at each site. The contextual elements that heavily influenced the use of quantitative performance feedback at each site are summarized in table 5.

5.1 Learning and Practice Improvement generated by the Measurement of Quality Metrics

Participants at each site were asked how the information was used for personal learning and how it fit into other strategies for self-assessment and ongoing learning. They were also asked if the information was used to support learning within the anesthesia department and what they felt motivated individuals and to change their practice in response to quantitative performance feedback. The quality reports produced by site I were not considered a significant resource for learning by most of the participants. Participants reported being disengaged from the program and many did not review the DA-QI report or use it as a self-assessment or learning aid. The information was also not used in an organized fashion at the departmental level. In contrast, site II was actively using their data to guide improvement efforts and projects within the institution. Data was seen as essential to these activities and to the overall functioning and effectiveness of
the anesthesiology department. Site II did not produce individual level data and the group-level data was not seen as highly relevant for individual self-assessment and learning.

In general, the process of setting goals and using ongoing quantitative feedback to assess the impact of process and behavioral changes was seen as necessary to learning from the data.

5.1.1 Personal or Individual Learning

Learning in response to quantitative feedback was highly variable with some participants at each site reporting using the information to support self-assessment and learning activities and others reporting no use. The DA-QI program at site I was intended as an individual learning tool but its effectiveness was decreased by problems with design, leadership and resource availability. The program at site II was intended to help a group of anesthesiology providers learn by supporting their QI efforts but also resulted in some individual learning.

The majority of clinicians interviewed from site I expressed curiosity about their performance compared to peers and this was a commonly expressed reason for reviewing DA-QI reports. There was widespread uncertainty regarding how the data was collected and analyzed and how the reports should be interpreted. This uncertainty was expressed by clinicians that reviewed the reports as well as those that tended to disregard them. Interviewees that did not use quantitative feedback for personal learning gave various reasons including: poor salience of the chosen metrics to their practice, large delays in report publication and a lack of confidence in their ability to meaningfully interpret the DA-QI. These reasons echoed the defenses found by Bucher and Stelling during their study of how resident physicians reacted to expert feedback (Bucher & Stelling, 1977).

“They’re not important. I don’t think anyone looks at it more than a few seconds and just throws it away.” (0652, Site I)
Table 5.1: Contextual factors and barriers affecting usability

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Site I</th>
<th>Site II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning emphasis</td>
<td>Individual</td>
<td>Group</td>
</tr>
<tr>
<td>Motivation for practice change</td>
<td>Intrinsic: competitiveness, conformity, beneficence</td>
<td>Intrinsic: competitiveness, professionalism, beneficence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extrinsic: variable compensation, career advancement</td>
</tr>
<tr>
<td>Communication of quantitative</td>
<td>Hard copy report placed in departmental mailbox.</td>
<td>On demand – published on website</td>
</tr>
<tr>
<td>performance feedback</td>
<td>No discussion during meetings or appraisals</td>
<td>Weekly – department meetings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monthly - meetings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quarterly – dashboard publication, chief’s newsletter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yearly – departmental activity report including dashboard indicators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for all quarters</td>
</tr>
<tr>
<td>Feedback type</td>
<td>Individual metrics (intended quarterly but not available in last</td>
<td>Group metrics (monthly and quarterly)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>calendar year)</td>
<td>Personal individual feedback within the context</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of improvement projects results from process</td>
</tr>
<tr>
<td></td>
<td></td>
<td>adherence violations or failure to meet outcome goals (daily)</td>
</tr>
<tr>
<td>Department member orientation to</td>
<td>None</td>
<td>Every new member is oriented to reporting</td>
</tr>
<tr>
<td>program</td>
<td></td>
<td>system and receives basic training in QI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>methodology through rounds and meetings</td>
</tr>
<tr>
<td>Leadership</td>
<td>Passive and intradepartmental</td>
<td>Active and multilevel</td>
</tr>
<tr>
<td>Accountability</td>
<td>Department must demonstrate some QA/QI activity to hospital administration and regional health authority but this does not need to include quantitative performance information</td>
<td>Department expected to collect metrics and engage in downstream action by the board of directors and hospital management team</td>
</tr>
<tr>
<td>Performance measure choice</td>
<td>Metrics on individual performance</td>
<td>Metrics on departmental performance</td>
</tr>
<tr>
<td></td>
<td>Outcome measures</td>
<td>Process and outcome measures</td>
</tr>
<tr>
<td></td>
<td>High priority given to metrics that were thought to be</td>
<td>High priority given to metrics that were thought to be</td>
</tr>
<tr>
<td></td>
<td>be attributable to individual providers</td>
<td>actionable</td>
</tr>
<tr>
<td></td>
<td>Limited input from providers</td>
<td>Input from providers encouraged, all metrics</td>
</tr>
<tr>
<td></td>
<td>No goals or benchmarks</td>
<td>were piloted within the group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Department goal-setting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparisons to baseline and published benchmarks</td>
</tr>
<tr>
<td>Group norms and culture</td>
<td>Individualism</td>
<td>Team-based care emphasized</td>
</tr>
<tr>
<td></td>
<td>Anesthesia providers all autonomous</td>
<td>Anesthesia providers of different levels of autonomy and responsibility</td>
</tr>
<tr>
<td></td>
<td>Immature safety culture</td>
<td>Safety and learning culture</td>
</tr>
<tr>
<td>Barriers</td>
<td>Culture of individuality</td>
<td>Provider resistance to change</td>
</tr>
<tr>
<td></td>
<td>Passive leadership</td>
<td>Availability of resources</td>
</tr>
<tr>
<td></td>
<td>Availability of resources</td>
<td>Improvement project and data fatigue</td>
</tr>
<tr>
<td></td>
<td>Low quality improvement capacity</td>
<td>Duplication with other departments and services</td>
</tr>
<tr>
<td></td>
<td>Provider resistance to change</td>
<td>Stakeholder alignment</td>
</tr>
<tr>
<td></td>
<td>Stakeholder alignment</td>
<td>Timelines of change – projects require several years</td>
</tr>
<tr>
<td></td>
<td>Distrust of other clinical departments and institution’s management</td>
<td></td>
</tr>
</tbody>
</table>
The general strategy for learning from the DA-QI involved making an assessment of how far one’s practice and results deviated from the group and making practice adjustments to move one’s scores closer to the mean. Simple strategies, or level one problem solving, predominated during interview discussions. Examples included using prior knowledge of common causes of complications in the post-operative patient and general anesthesia knowledge to tweak intraoperative processes. Some anesthesiologists utilized medical records analysts and attempted to access more specific data such as cases flagged under a certain quality indicator, by type of surgery, or patient demographic but this was an uncommon strategy.

“At first I could not believe that my patients needed that much [anti-nauseant]. So I started going back to PACU and checking on the patients myself to see what was going on…I looked at what I do and then I started talking to some of my colleagues.” (0825, site I)

Receiving ongoing feedback allowed clinicians to engage in natural (before and after) experiments but delays in receiving information due to a lengthy reporting period was very limiting. The DA-QI program was receiving minimal support from the hospital and health authority and as a result there were significant delays in the production of reports as well as cancellation of reports. Limited resources impacted the type and amount of data that could be collected and distributed to the department. Even when the reports were available the resources and supports to help clinicians use the information for improvement were not.

“At getting information from a year ago, once every quarter…makes it difficult to know how changes to my care are impacting these indicators, and more importantly, impacting the patients.” (0777, site I)

Department members from site I relied largely on other methods of learning, such as observing their patient’s condition in the post-operative period, informal discussions with nurses and anesthesia colleagues, and formal CME activities over quantitative feedback. None of these preferred methods of learning incorporate the expert external feedback that has been suggested to be necessary for true professional growth (Eva & Regehr, 2005). There was no indication that informal peer support or mentoring for the purposes of reflection and practice change was occurring in the department and no formal mechanisms had been put in place.

Another way the reports were thought to function was through increasing mindfulness and awareness of quality and safety issues:
“We get this report every quarter and it reflected your care and your work, so without a doubt it raised the profile of quality of care and safety and made people think about it. They may not necessarily be thinking every day, but I think people think about it, probably in a fair bit of depth, quarterly when they receive their report and probably for a period of time after they receive it as well. For myself, every time I take a patient to the recovery room, I literally have these things in my mind. What’s my patient’s pain score? What’s my patient’s oxygen saturation? Are they going to need anti-nauseants?” (0777, site I)

Interviewees reported that learning and behavioral changes were difficult due to fixed practice patterns and a tendency for practitioners to focus on what they do well. Several interviewees suggested that mechanisms to support learning and practice change as well as clear expectations and accountability would be needed to increase the effectiveness of quantitative individual practice feedback:

“We have some individuals in the department...that are very resistant to changing anything about their practice. Some because that is their personality style...and some older people perhaps further away from their education that are hesitant to try new ways of doing things. So even given new information and feedback, their quality indicators, are not going to branch out and try something different.” (0437, site I)

“I think that people like to do continuing medical education to remind them that they are good, that they know a lot. It’s hard to hear about the stuff we don’t know about and I think we need to be forced in some way.” (0825, site I)

There was very little individual goal setting amongst interviewees from site I around DA-QI report results. Goal setting is an important step on the adult learning journey, (Brydges & Butler, 2012; Kaufman, 2003). Most of the participants were satisfied if their results were close to the mean. Participants that wished to make changes when their results were worse than the group average set vague goals such as, “try to do better,” as opposed to explicit goals associated with time frames.

Interviewees from site II overwhelmingly did not view their performance data as a tool for individual learning but did feel that it precipitated some self-assessment and could alter clinician self-perceptions. One individual spoke about how the quality and performance data had an effect on newer members of the department:

“I think, initially, the younger ones in the department have that sense of entitlement and ‘that’s never going to be a problem for me,’ and when it is a problem for them, all of a sudden it’s like, ‘oh, we need to do, x, y and z.’” (0754, site II)
One of the results of the system, where non-compliance with specific quality improvement initiatives leads to personal feedback either verbal or via email, and can impact variable compensation, was that individuals learned to identify their instances of noncompliance and communicate them to quality project leaders pre-emptively:

“I think it impacts some people’s practice because I will get emails from people saying, ‘hey, look, the reason we failed on this specific thing is because of x,’ so they want people to know where they are having problems.” (0754, site II)

Although the performance data was not seen to be a tool that could influence individual learning in site II, the use of individual feedback within the quality improvement projects was thought to precipitate learning. In many cases, strategies to support individual learning were utilized within quality improvement projects in an effort to improve the metrics. These strategies included mentoring of less successful practitioners by more successful practitioners and the creation of memory aids for process changes. One of the major differences between the two sites is that in site I, reflection and action were left within the domain of individual practitioners, whereas in site II, specific individuals and teams were allowed dedicated time for higher level problem solving. Once strategies were identified, they were communicated to the department and to working teams within the OR environment and individuals were encouraged and supported in making changes. The results of these case studies echo the results of studies on MSF – that reflection and high-level problem solving in response to feedback, whether qualitative or quantitative, are resource-intensive activities that require significant support (K. Overeem et al., 2009; K. Overeem et al., 2012; Sargeant et al., 2009).

The quality data system at site II heavily promoted individual learning in the areas of quality improvement theory and methodology. All members of the department are expected to be familiar with standard QI tools such as run and control charts and key driver diagrams and the PDSA cycle. The institution has committed resources so that new hires receive orientation and training in basic QI methodology. Individuals with an interest in this area are encouraged and supported in obtaining extra training.

Interviewees in site II felt that individual learning, on its own, was not an effective way to improve care. They preferred attempts to change standard workflow and processes.
“Seeing the data repeatedly almost makes people jumpy about it. But it wasn’t until they implemented the [improvement] bundle and we had specific feedback and we talked about it at meeting after meeting after meeting…that we finally got the outcomes we were looking for.” (0508, site II)

5.1.2 Learning within the anesthesia department

Site I’s DA-QI reports had been discussed in the past at department rounds – especially when wide variation in scores or outliers were observed. The QI coordinator is expected to undertake one major project a year and the choice of topics has occasionally been influenced by results in the DA-QI reports. The resulting audits and chart reviews have been presented to the department along with relevant recent literature and practice guidelines. The onus seems to be very much on individual education and practice change as a method of improvement as opposed to system change. For example, the metric for unplanned admission from day surgery was felt by many of those interviewed to reflect surgical care and patient selection rather than the care of an individual anesthesiologist. However, the anesthesiology group could reduce unplanned admissions by reviewing and adjusting guidelines for appropriate selection of ambulatory surgical patients or requirements for patients with certain risk factors to be seen in the anesthesia consultation clinic prior to booking.

Anesthesia providers at site I are not able to easily access quality improvement and leadership training and the absence of this training during medical education was seen as a significant barrier to system change.

“It’s not just enough that you go to some conference and hear about the evidence on the latest drug or new technology. You have to be learning about system change and improvement…trying a new drug, trying a piece of equipment, that’s the easy part. The hard part is how do you change the whole system and the process?” (Site I, 0465)

Site I interviewees felt that their clinical responsibilities left little time for pursing improvement work. Funding for non-clinical time was only available for a select minority of department members with formal administrative roles. Members wishing to take on QI projects would do so at significant personal cost by giving up either leisure time, or income-generating clinical time.

At Site II, leadership emphasizes the importance of collective and team learning over individual learning when using the quality data. The process of learning from the data is a departmental
exercise. The department meets to brainstorm and discuss the suspected key drivers leading to desired and undesired outcomes and to propose modifications to current processes:

“When we have decided to make a change, that change has not been made by me, it’s been a collaborative effort. I may be driving the conversation but it’s something that we all agree upon: ‘hey, let’s try this, this might work, it might be a good idea.’” (Site II, 0047)

After process changes are enacted, data is monitored for both compliance and changes in patient outcome. Through an iterative process, the most influential key drivers are discovered as attempts to link process changes to changes in quality metrics are made. Quality leaders and their teams are very active in steering and focusing efforts and in working to ensure compliance. The use of ongoing monitoring of their efforts through data collection and presentation allows the QI teams and the anesthesia department to objectively assess their efforts.

Site II had provided several individuals from the department of anesthesiology with compensated non-clinical time. Even so, respondents reported still doing a great deal of work on their own time and during small breaks in clinical duties during the day. Interviewees from site II struggled with balancing clinical and nonclinical duties especially as surgical case volumes fluctuate significantly depending on the time of year. During months with higher case volumes, non-clinical time is often not granted as clinical activities take precedence.

Another way that the quality metrics supported group learning was to provide objective comparisons of care at the main hospital and the mostly ambulatory surgical center. The ambulatory center had high scores on operating room efficiency, controlling post-operative pain, patient satisfaction and decreasing adverse respiratory events. By analyzing the underlying systems and processes, those that made the ambulatory site so successful were identified and extended or adapted to the main hospital site despite a different patient population and level of acuity. Since the same anesthesiologists provide care at both sites, having metrics reveal a high level of success at the ambulatory site increased group self-efficacy and supported translation of knowledge to the larger site despite a higher complexity surgical population.

Several interviewees from site II stated that questioning and pushback was an important part of the learning process and facilitates open discussion and communication. The chair of the anesthesia department at site II spoke at length about how ongoing discussions aid learning and
help the group to move through Berwick’s four stages of “facing reality:” realizing there is a problem, realizing the problem requires action, taking ownership of the problem and finally acting to solve the problem (Berwick, 2003):

“They’ll say, ‘this is not a problem.’ So then you measure it and show that it is a problem…Level two is, ‘it’s a problem but it’s not that bad.’ So knowing that you are going to have level two from the beginning, before you even start, you need to say to them, ‘if this was you, how would you define this as a problem? Give me a number.’ So when they say this problem is not that bad, you can say to them, ‘but you told me that it was a problem if it was more than 2%! ’ Level three is, ‘it’s significant problem that we need to do something about, but it’s not our problem, it’s somebody else’s problem.’ So take, for instance, pain in the PACU, they’ll say, ‘we’re doing everything we should do so it’s actually a nursing problem. You should ask nursing to fix it.’ So you have to have an argument for that well laid out ahead of time as well. Saying in the beginning, ‘well, when we defined quality for the anesthesia service, you said that pain in the PACU was part of your professional activities, now you are saying that it isn’t part of your professional activity?’ And they’ll say, ‘well, okay…’ It’s almost like you have to have a tape recording that you can play back to them.”

Feedback helped the group at site II to learn how to improve processes and outcomes and increased appreciation of group efficacy. Even if individuals couldn’t see the value in a certain project or requested process change, feedback showed them linkages between process modifications and outcome improvements. This was the case with a project aimed at reducing costs through decreasing fresh gas flow rates during anesthesia:

“That project…irritated people to begin with, especially when you have someone coming around to the operating room and saying, ‘why are your flows so high,’ and it’s because [you’re] waking the patient up. Those kinds of things piss people off but, you know, when they send us back information and say, ‘we bought 100 less bottles of anesthetic gas and we saved this in oxygen costs,’ then that makes a difference.” (0754, site II)

5.1.3 Motivation

Interview subjects at both sites spoke about the strong effect of intrinsic motivators on personal and group learning and their desire to provide high quality anesthetic care. Site II also made use of the extrinsic motivator of variable compensation tied to group achievement on quality metrics. The proportion of variable compensation associated with each metric was set by leadership within the anesthesia department and this was made possible by a salary model of remuneration for both nurse and physician anesthesia providers.
Interviewees from site I expressed strong motivation to provide a safe and effective anesthesia service and doing so was wrapped up in the identity of anesthesiologists. Intrinsic, personal motivations were important for learning and practice change. The most commonly mentioned motivation for reading the DA-QI reports was curiosity and the most commonly mentioned motivations for practice change based on the results was competition:

“Physicians…all believe they’re the best, they love being shown that they are better than their peers and will do things on their own to improve their own practice so they can look as good, or better than, their peers.” (0465, site I)

Other intrinsic motivators mentioned included benefits for patients and team functioning:

“For me personally, [changes] do have to have a positive impact on patients. I’ll put in a little more work…if I can see a benefit to the people I’m going to look after and then, you know, if it might have a positive impact on the team.” (0437, site I)

Some interviewees at site I ascribed negative personal characteristics to individuals that would be primarily influenced by external motivators, such as extra compensation. The use of extrinsic motivators was thought to increase the risk of metric manipulation.

“The metrics we choose for pay for performance are critical and it’s very easy to get a perverse outcome if you choose the wrong metric. As soon as you use that [money] as your reward, to me that sends the wrong message…I believe you get far better results if you focus on providing exceptional care for patients and being able to pat yourself on the back, saying, ‘I’m part of the team that does that.’” (0465, site I)

The importance of intrinsic motivation was also stressed by interviewees from site II. Anesthesia providers are viewed by the chair of the department as professionals who are motivated by altruism and desire a high level of achievement and are able to incorporate data into these intrinsic drives:

“They all want the best for their patients so they have to believe that this project or this thing will make a difference to their patients. So if you can make a good case for that, and that involves measurement, and they actually believe the measurement, then that’s the best motivator.” (0618, site II)

The data collected by site II is seen as a way to illustrate quality problems in concrete terms and increase provider motivation for behavioral change. The ability of quantitative feedback to increase motivation for change has been shown in a number of studies (Amin et al., 2007; Donaldson et al., 2005; Mangram et al., 1999; Poniatowski et al., 2005). Another way that the
use of data has been motivational is by giving the group a sense of efficacy and success. As change and improvement occurred, practitioners were seen to have enhanced acceptance and engagement with the measurements and behavioral changes.

“The group realized that we can actually do this collectively, it’s not completely out of the realm to be able to get to 95% success with this. So I think that’s really changed. It’s taken about 10 years and they’ve seen all sorts of projects come and go.” (0295, site II)

“It’s strange, but on a daily basis it feels like nothing ever changes. However, when you step back and look at the changes that occur over a two year period, for example, it’s significant.” (0508, site II)

Competition between physicians was seen as motivating by interviewees from site I but those interviewed from site II stressed competition with other institutions. Providers at site II were proud of their institution, its reputation as a leader in health care quality improvement, and the special programs it was known for. While some of the metrics on the quality dashboard were reportable by other institutions, for example the metrics used for P4P recovery from CMS, many of the metrics lacked benchmarks. Competition with other institutions seemed to be reputation based rather than data based with the data program being seen as enhancing the institution’s reputation rather than allowing direct quantitative comparison. Motivation to participate in QI activities and to modify behavior in order to support QI projects was rooted in this sense of professional pride.

“We all know the programs that the hospital is known for. And I think that people who work within those programs have pride. I routinely hear the pride for these programs from our nurses when they talk to our patients and families.” (0508, site II)

“I think that physicians are very competitive in their nature, I just do. So when I see that we are scoring below other hospitals or scoring below where we should be or that we aren’t significantly higher…if you told me that [name removed] hospital scores 90% on this, I can guarantee you that most of our department would be gunning for 90% or higher…probably higher.” (0679, site II)

Site II makes use of extrinsic motivators though the use of variable compensation schemes which are tied to group achievement of goals associated with specific quality improvement projects. The amount of variable compensation awarded in this way is a small percentage of total compensation for anesthesia providers but was deemed to be effective because providers are seen to want to succeed:
“I think it’s the high achiever personality type that is in this field. When the bonus doesn’t come through, and there have been years, it’s to demonstrate that we are serious about this expected behavior. But it’s really the principle of the thing that makes everyone behave that way.” (0508, site II).

Tying the variable compensation to group achievement was seen by one interviewee as a way to increase cooperation:

“Everyone’s bonus was based on the group’s performance. To be aware that we all need to improve as a group means everyone is motivated to work together.” (0508, site II).

Trisolini suggests that extrinsic motivators work very well for rote tasks, (Trisolini, 2011) and the success seen with their use by Site II may be due to the presence of a well-resourced QI program that could simplify the complex process of improvement into a series of compliance steps or tasks for clinicians.

Another source of motivation at site II is provided by the institution’s decision to recognize quality improvement as an academic stream within medicine. Participation in quality work is associated with opportunities for career advancement within the health care center and the affiliated university and accords academic status within the group. Some authors have classified elements such as respect, pride, academic and clinical reputation as extrinsic motivators, (Trisolini, 2011), while others have classified anything other than monetary compensation or fines as intrinsic motivators, (Pink, 2011). Both of these authors agree that these elements can be powerful motivators for physicians engaging in complex problem solving.

5.1.4 Strategies for Self-Assessment and Sources of Feedback

One of the first steps of practice change is determining that current behavior does not lead to the desired outcome. Performance data can provide a practitioner with objective evidence of their practice outcomes and this can lead to self-, and/ or group-assessment.

The individual-level feedback provided at site I prompted some self-assessment activities but uptake of the reports and their use for personal learning was low amongst those interviewed. Physicians who thought the DA-QI reports were useful spoke about integrating them into larger strategies for self-assessment, but physicians who did not find the reports useful placed greater emphasis on other forms of feedback. Both groups used information from patients and nursing staff, particularly those in the PACU and OR, as their most salient sources of feedback:
“What I rely more upon, is what my patients are telling me, and what I hear from the recovery room nurses…because they work with me.” (0652, site I)

Additional sources of feedback that supplemented or supplanted the information contained in the DA-QI reports included interacting with colleagues and discussion of anesthetic care and near misses or poor outcomes during in the department’s quality improvement round.

“I know how my colleagues operate. We all see how we deliver patients to the recovery room and we frequently deal with each other’s patients in the recovery room when they are not here, late in the afternoon.” (0825, site I)

“The department felt they [cases] spoke more to our circumstances here. It seemed more urgent…I think it [case-based QI rounds] prompted everybody to look at their own practice because they could…see that this could be me.” (0825, site I)

Self-assessment activities were not strongly inspired by the group level feedback provided by site II. After being presented with the data, providers at site II reported transiently emphasizing process adherence in their daily practice but no one reported being inspired to undertake personal learning projects or more involved practice assessment. There was a suggestion that physicians need to be shown data and given feedback in order to facilitate accurate self-assessment but this did not emerge as a major theme in the set of interviews from site II.

Site II has an extensive infrastructure to support quality improvement projects and it could be that the emphasis on using the data to inform system changes reduces the perceived need by clinicians to use the data for spontaneous individual practice assessment.

At both sites, participants felt that the quantitative feedback they received was not representative of their skills as practitioners or the overall quality of their practice. Some of this sentiment was due to the impression that metrics in use inadequately conveyed meaningful patient outcomes or experiences, some with the limited metrics in use and some with concerns about sampling.

5.2 Performance measure choice

The content of metrics chosen at each site influenced the ability of clinical providers to use them for practice and system change. The attributes of useful metrics, as described by those interviewed from both sites included representativeness, salience to anesthetic practice and clinical problems and actionability. These attributes are also favored by the American College of Cardiology Foundation and American Heart Association who have released advice on the
methodology of performance metrics based on literature review as well as experience within their own field (Spertus et al., 2010). Clinicians preferred metrics where there was less perceived ambiguity of interpretation. Metrics that were designed to be actionable were more likely to lead to practice changes and downstream effects compared to those that were chosen for other reasons.

The experience of site II suggests that intelligent linkage between different types of measures is more useful than trying to find the perfect mixture of measures from the domains of structure, process and outcome. For example, site II has been able, by measuring both an outcome and its purported key driver processes, to show temporal linkages between outcome improvements and increased adherence to process. These linkages have proved to be very motivational for providers working within this system. For example, the measure for the surgical site infection prevention bundle was not initially accepted. It was introduced at a time before there was sufficient published evidence for its components. Seeing that compliance with the bundle was linked to a lower incidence of SSI increased both acceptance of this measure and adherence. Provider buy-in was seen to be influenced by the length of collection as well as the amount of effort taken to ensure robust, high quality data and transparency in data collection methodology and analysis:

“We have a long track record of [data collection] so they understand that when we present data, it is reliable data…that it is at least adequately sampled, that there are objective measures for it and a timely reporting system.” (0047, site II)

Using different types of measures in association could also reveal linkages between structure and process adherence and structure and outcome that could be used to guide organizational planning and policy implementation. This may require multi-organizational data.

Site I only measured individuals and site II only measured the group. Neither site has addressed the complex issue of which measurements may be best suited to individual measurement and which would be more suited to group measurement.

5.2.1 Useful metrics are representative

The major limitations that clinicians from site I identified within DA-QI reports was that the current metrics were too limited and did not provide a balanced view of anesthetic care quality.
None of those who were interviewed felt that the metrics they received adequately or accurately reflected the care that they provided to patients. Metric choice in this setting was constrained by resource availability.

“I’ve wracked my brain for a couple of years now over how we could measure quality indicators that really indicated the quality of anesthetic care. I don’t think that these do a good job of that. They are things that are easy for clerks to find in a chart and to record, but I’m not convinced that they truly indicate quality care.” (0777, site I)

Preconceived notions of quality were used to generate measures and over time it was the impression of those interviewed that the chosen measures reinforced how quality was conceptualized and defined attribution for different aspects of patient outcome and experience. Several participants wrestled with the inability to define where anesthetic responsibility for patient outcomes ended.

“I don’t know if [unplanned admission after day surgery] is a surgical issue or an anesthetic issues, but it is an anesthetic issue as far as this audit is concerned.” (0652, site I)

Interviewees from both sites understood that metrics could be used for purposes outside of quality improvement and learning such as performance management, funding, privileging and credentialing. Representativeness and balance in metric choice can obviate some of the concerns that clinical providers have whenever their performance is measured and reported to outside groups.

Interviewees from site II thought that the metrics that made up the anesthesia quality dashboard were largely representative of the care provided. These metrics captured more domains and aspects of care compared to those used by site I. Site II was able to access greater resources and this allowed them to spend a considerable amount of time designing and piloting their metrics and consulting with stakeholders.

The use of less than universal sampling was felt by many members of the group to decrease data quality and introduce error. One respondent reported that the sample size for certain metrics that required direct observation by the data abstractors could vary based on workload and they felt this variance could make the data more difficult to interpret and use. A larger sample would also allow the group to develop and apply risk adjustment models.
Some of the interviewees from site II expressed frustration with the measure of “professionalism” that forms part of their yearly evaluation and which may affect their variable compensation. Although this is not a measure present on the quality dashboard, dissatisfaction with this measure illustrates how non-representative measures are likely to be received and supports Hall’s suggestion that clinicians would approach measures that don’t accurately reflecting performance with anger (K. W. Hall, 2010).

“[The professionalism measure] has bred some ill will as well as some grumbling. I’m struggling a little bit because I’m not sure it’s changed behavior actually, and it’s caused some frustration. I think that’s because it’s fairly arbitrary. What it’s made me realize is that there is no room for arbitrary. It’s specific people high up in the department that are doing the rating - arbitrary feels negative and creates quite a bit of frustration and it feels disrespectful.” (0679, site II)

Although those at site II felt that aggregate measurement is preferred from both philosophical and utilitarian viewpoints, they are subject to external programs and pressures that require individual measurement. This necessitates a need to rationalize the two approaches. A reduction in the costs of data collection would exist if the same measures could be used for both internal QI work and meeting external program requirements but these would not necessarily be the most useful metrics for either purpose. The group reports difficulty in identifying measures that would be both attributable to individuals and representative of the care provided.

5.2.2 Useful metrics are salient

Interviewees were in agreement that measuring the quality of anesthetic care was a difficult prospect and recognized that different viewpoints would exist both between, and within, groups such as clinicians, patients, and payers. Site I’s program operates in a silo and the needs of patients, the hospital or the insurer were not considered when choosing metrics. Metric choices at site II reflected some outside influences including the existence of a P4P program and a business need to measure and improve efficiency.

The anesthesiologists interviewed at site I wanted measures that were important to them personally for the purposes of learning but examples of these varied widely from procedural success rates, measures of the efficacy of post-operative pain control, chronic pain treatment, measures of efficiency of operating room function and tracking common complications from anesthetic procedures. A high value was placed on measures that would reflect high quality
clinical evidence, best practice or that captured a highly efficacious process. The need for metrics to evolve with changing technology or after maximal improvements had been made was appreciated. Metrics with minimal inter-clinician variability did not hold a lot of salience for clinicians from site I. Measures for pain scores and PONV rescue therapy in the PACU were seen as patient-centered, and since a high level of importance was put upon patient experience, these metrics largely accepted although it was felt that there was room for improvement with respect to their design. One participant expressed frustration with the lack of more specific data that would help with result interpretation:

“They’ve got pain score in there but length of time that they stay with a pain score that’s above 7…Is it just once that it happened or was it really hard to control their pain?”
(0437, site I)

Delayed feedback on aspects of anesthetic care which occur in front of the anesthesia provider, such as initial oxygen saturation in the PACU, were thought to be of limited value because corrective action could be immediately applied and the consequences to the patient were minimal. Quantitatively capturing events that occurred outside of their immediate care where feedback is generally lacking, such as measures of the efficacy and safety of the post-operative pain service, was felt to be desirable. Interviewees from both sites placed a high value on measures that would capture patient experience and satisfaction:

“Satisfaction is maybe too easy, but some sort of patient perspective on the care they receive. You know, were they treated respectfully and politely and did they feel they were in safe hands and were they given the information they needed and had an opportunity to make decisions for themselves and have autonomy.”
(0777, site I)

Measures of “soft” skills such as communication, teamwork, and aspects of professionalism including the quality of interpersonal interactions with patients, were thought to be useful. The distinction between performing a process in order to achieve compliance and performing the process in a way that promotes mindfulness was expressed by the desire for metrics that would capture some of these soft skills.

“The degree to which people truly embrace the [WHO Surgical Safety] checklist as opposed to tolerating it…because, you know, there are some people who will go through the motions because that’s what they are supposed to do. I would love to be able to measure who really truly is thoughtful and makes good use of the safety checklist.”
(0777, site I)
The approach at site II with respect to salience was different. The quality dashboard at this institution is reported to management and parts of it are made public with accompanying expectations for ongoing high performance and improvement. Metrics that facilitated this type of communication were considered salient. A second type of salient metric was one that allowed a particular problem to be captured and monitored while improvement efforts are underway:

“There are certain things that are fundamental to your business that you are going to measure always at a certain frequency but there are certain things that you know will be more of an immediate problem or there’s been a change in the environment that’s caused this to be an immediate problem. So then you’ll have to add measures. Some will fall off once you’ve achieved what you wanted to achieve with it and others are so fundamental to your business that you always measure them.” (0618, site II)

Site II ensures that its active measures are salient by using feedback from group members to create new metrics to deal with ongoing problems and frustrations with practice. Leadership is involved in selecting new measures so that they are aligned with departmental and business priorities:

“Our CRNAs complained to our chairman about dissatisfaction in our process for doing inpatient preoperative assessment[s]. And he heard it so repeatedly that he actually…well he knew it was a problem and he was planning on delaying a year before tackling it, but because he heard so much complaining about it they reprioritized and made it a project for that year.” (0508, site II)

The anesthesia providers interviewed from site II accepted the measures of the quality dashboard as being useful and pertinent to their practice but there were difficulties with acceptance when the program began. Acceptance improved over the years through two mechanisms. The first was alteration of some measures based on feedback from the anesthesia providers, such as when the patient satisfaction measure was changed after complaints that it was too subjective. The second mechanism was when department members saw correlations in the data develop between adherence to process and improvements in outcome. This implies that perceived salience of a measure can be increased through positive experiences related to its use.

Site II is currently undertaking a large project to design measurements for patient experience. Partly this has to do with a P4P program related to experience and satisfaction scores, and partly to do with a more holistic way of approaching quality in health care:
“What we are coming to realize as we start to think more and more about experience, is that safety and experience and productivity and flow and all these other initiatives really go hand in hand. A patient and family’s experience is about their experience in total. They don’t compartmentalize the way we do. They don’t say, ‘okay, that was a safe experience but the flow was not good…’ So we’re trying very hard to cut across these silos and think more holistically about the way that patients and families experience care.” (0679, site II)

Designing measures through the lens of patient experience as Site II is doing may be one way to address the need for metrics that capture patient transitions, shared accountability and cut across diverse care areas within an institution (Fisher et al., 2007; Myers et al., 2008; Spertus et al., 2010).

5.2.3 Useful metrics are designed to be actionable

Both sites chose metrics that were presumed to be actionable. Site I chose to measure aspects of care that could be affected by individual anesthesiologists by focusing on the patient’s immediate condition in the PACU. This strategy has resulted in DA-QI reports that many anesthesia providers do not consider to be useful because they mirror feedback that is received real-time when providing patient care.

Site II spent more time and effort than site I when designing metrics and focused on measuring outcomes that have key drivers under the influence of the department of anesthesiology. Unless the department feels it can influence the outcome, it will not be included in the data collection program. There are strong expectations from the organization’s leadership that the data will be used to alter systems of care.

“I think [department members] have a fair amount of say, usually through their division chiefs, [in] what needs to be measured. I think the leadership has to push back a little bit. We can only measure so much. Everybody would like to know this, this and this but it costs a lot to do that. And so the leadership, their job is to say, ‘that’s a good idea, let’s say we find this, what are you going to do?’ There has to be an action associated with it.” (0618, site II)

5.2.4 The number of metrics should not be overwhelming

A barrier that was mentioned many times by interviewees from site II was that a large volume of quality improvement projects caused data fatigue, burn-out and difficulty knowing which projects deserved more provider attention. Front line clinical providers at site II often have their daily workflow and processes affected by multiple improvement efforts. These projects may
originates within or outside of the worker’s own department or division. Feedback pushback (Lubarsky et al., 1997) in this system is likely occurring.

“Some of the burnout comes from…project fatigue really. There are a lot of projects going on hospital-wide.” (0508, site II)

“I would personally like to see them do fewer measurements at a time…if they could tone down the number of run charts they are doing on a daily, or a monthly basis it would make things a lot easier. I think there are just too many projects going on at any one particular point in time and it’s making it difficult for people to concentrate.” (0754, site II)

Leadership recognizes that each additional project requires front line staff to change their behavior and shift their attention and that this is a threat to productivity and meaningful improvement. They are actively trying to sharpen the focus of improvement projects:

“…you have to be careful that you don’t pick too many things, because then they’ll neglect other things. All this fits into how many projects you can do at once and what are the key messages. This is our biggest problem, how many projects can you do? What are the unintended consequences…if you do these what isn’t going to be done?”

(Department Chair, site II)

5.3 Communication and Collaboration

Information from quantitative performance feedback must be effectively communicated to providers in order to result in learning and practice change. The information needs to be presented in a way that minimizes difficulties with interpretation and analysis. Providers also need mechanisms to share their interpretations and solutions so that peer-to-peer learning can occur. Cross-divisional communication mechanisms will allow learning to occur amongst the different disciplines of the perioperative care team including anesthesiologists, surgeons and nursing staff. Communication between the clinical division using quantitative feedback and organizational management with respect to program impacts and successes helps to secure ongoing support including funding.

5.3.1 Indicator and Report Interpretation

There were significant differences between site I and site II with respect to how quality reports were organized and the resources available to help members of the anesthesiology departments with interpretation. At site I, the DA-QI reports present individual scores for each member related to a horizontal line that indicates the group mean. The report is a static representation of
a particular quarter which makes it difficult to determine how results are changing over time. Because no confidence intervals were included, providers had difficulty determining if deviations from the mean were significant.

At site II, the quality dashboard shows that quarter’s results for the group compared to baseline data and the goal for each indicator. Run or control charts for each item show how the group’s score varies over time. There appear to be advantages to using well known formats from QI science such as the ability to share information with QI professionals and other institutions already familiar with these formats and the ability to leverage existing educational materials when educating employees and providers.

At site I the reports are considered an individual learning tool and are rarely discussed in the group setting. New members are not formally oriented to the program or given assistance in interpreting the reports. Several uncertainties and frustrations with respect to report interpretation were voiced by interviewees from site I. Compared to other feedback that clinicians might receive, these reports were not considered timely. Reports have been delayed, sometimes by over 12 months, or not produced at all. Even when reports were received regularly, members had concerns about the remoteness of feedback from the clinical encounters and tended to view the information as abstract and non-urgent. Immediate feedback, such as from patients, nurses or monitoring devices was felt to be more actionable. Interviewees from site I thought that de-identifying the reports made them less threatening but also less salient and more difficult to learn from.

“They are essentially anonymous so no one individual has to admit that they are the outlier. And they don’t typically know who the other outliers are. Whereas, when you’re in a quality improvement round, and you’re sitting in a room and there’s 20 or 25 of you, and you are discussing a case, it usually becomes apparent who the anesthesiologist was that provided that care and…you’re in much more of a situation of taking responsibility for your actions, learning from your colleagues and showing them the grace that you want to be shown when it’s your turn…so it’s much more on a human level. Whereas, with these quality indicators, you can separate yourself from them.” (0777, site I)

Interviewees spoke of a tension between the need for anonymity and transparency in reporting. Anonymity is felt to potentially decrease feelings and shame and blaming behaviors in response to feedback data (Werner & Asch, 2005) and this sentiment was shared by interviewees from site I. Transparency in reporting may increase accountability and the potential for learning from
peers. Several of the participants reporting knowing the codes for a fair number of their colleagues but no one thought that incomplete or imperfect anonymity was a significant problem. The potential for shame from non-anonymous reports was felt to be low because of the low variation in DA-QI results between individuals.

“I don’t think it matters because I have an idea what kind of anesthetics my colleagues give anyway, based on water cooler talk and nurses’ talk…and it’s not like there are a lot of outliers if you look at them.” (0437, site I)

“Personally I would have preferred that my name is on here and not just a number…if I knew that you were Dr. X and I see your pain scores are beautiful, I could go to you…and I could just say, ‘oh, Dr X, what do you do for pain?’” (0825, site I)

Some clinicians wanted reports that were summative or provided a judgment and found that they currently received too much information. The reports show an individual’s score in relation to the mean for the group and the individual scores of colleagues. When asked how far above or below the mean one would have to fall before triggering some kind of response, most participants were unsure and felt that inclusion of standard deviations, confidence intervals or more sophisticated statistical analysis would be useful. Most members expressed satisfaction if their scores fell close to the mean while group leaders expressed satisfaction if variation was low.

“I could be well within the standard deviation and still trying to change my practice…there’s not enough statistical support around it to inform our efforts.” (0777, site I)

The mean was seen as a surrogate for appropriate treatment and deviations above and below were often viewed as under- or over-treatment. Without using a benchmarking method, such assumptions could be very misleading for clinicians (Cook et al., 2012). Comparisons to other groups, benchmarks or trends in performance over time were suggested methods for increasing usefulness. To illustrate, one participant, in regards to their pain scores in PACU, said:

“I like to be here [at the mean]. I don’t like to be down here or anywhere here [above the mean]. Because these guys are living dangerously, in my opinion. These ones [patients with low pain scores who have received more narcotic] stop breathing, get lots of headaches, and these ones [patients with high pain scores] are shouting and screaming…” (0580, site I)

This individual has assumed that the group mean represents optimal pain management – where the tradeoffs of having low pain scores are balanced by guarding against adverse events
associated with narcosis. However, this may not be the case. Benchmarking or adding in patient satisfaction with pain management and the addition of measurements for adverse events associated with narcotic-based pain management such as nausea, sedation, delirium and respiratory depression, would allow more nuanced interpretation and goal setting.

Most of the interviewees from site I were aware of the effect of different practice mixes on the data, both dramatic and subtle ones. The time spent at each hospital, proportion of ambulatory surgical procedures, amount of after-hours work, poor quality charting, and differences in PACU nurse training and documentation norms between sites were identified as potentially affecting the results. Some clinicians would make internal adjustments for this when interpreting the reports while others would use the case summary section to find colleagues with comparable practices so that their scores could be used as a benchmark. No one mentioned that they would like to see formal risk adjustment techniques applied to the data but separating out physicians with similar practice mixes in the reports was seen as useful.

The idea that colleagues could “fudge their numbers” in order to falsely improve their results was prevalent amongst those interviewed. There was also concern that poor charting could affect rankings. Other investigations have shown that extrinsic rewards and punishments are associated with increased attempts to manipulate quality metrics (Lilford et al., 2004; Spertus et al., 2005). This was seen in the data from site I. Individuals in the group had personal experiences with colleagues attempting to manipulate the metrics and these manipulations were felt to be in response to fear that negative personal or professional consequences would result from the reports.

“People might be embarrassed and they might be anxious about this. And they might go to severe lengths to make sure their numbers look great. I’ve seen that. For instance, the desaturation on arrival. ‘On arrival’ becomes a very vague term.” (0825, site I)

“In the early days when [colleagues] were concerned about their scores they sometimes were pressuring the recovery room nurses to be more favorable in what they wrote down.” (0777, site I)

The difficulties with case volume and how very low case volumes may affect results was not mentioned. The number of cases collected for each report was felt to be adequate and most felt that their practice mix and volume remained stable over time. Site I had not gone through a process to determine if the sample size per report was large enough to show significant
differences between practitioners, to determine the stability of the practice and case mix over
time, or what sample size would be required for risk adjustment. There was a considerable
amount of practice heterogeneity between site I anesthesiologists as their work was divided
unequally between two facilities and unequally between surgical sub-specialties.

The reporting system at site II uses formats drawn from quality improvement science -
dashboards, run charts and control charts. Care is taken to keep the format for data presentation
consistent throughout the institution. When new anesthesiology providers join the department
they are formally oriented to the quality system, the quality dashboard and the use of run and
control charts. The significance of each measure, the way it is collected and the implications for
compliance are laid out. In addition, the institution has structures in place so that all clinical
providers receive grounding in improvement science. Consistency in reporting format and
formal training in report interpretation facilitate communication of quantitative performance
information throughout the institution.

"Having [leaders] driving this work, training and teaching providers from the front line,
all the way up the ranks, how to do improvement work, how to look at the data, how to
learn from the data, how to develop improvement work based on the data, has really shot
our institution to a higher level" (0679, site II)

“They use that standard format we have – the key driver diagram [that shows] your goal,
your key drivers and your interventions and suggested PDSA changes. We’ve gotten so
used to seeing that. Run charts with baseline data…the format for each of these [quality
initiatives] is the same.” (0295, site II)

The department discusses quality measures and information frequently. The dashboard is
discussed in detail quarterly and some of the measures on it are discussed at monthly and weekly
meetings as well.

Several members interviewed from site II believed that larger samples would improve the
usefulness and ability to interpret quality information by decreasing alpha error and would also
allow for individualized feedback. Those who were heavily involved in quality improvement
work reported hearing frequent complaints about sampling from colleagues. This occurred even
though sample size had been predetermined and the rationale for each sampling strategy had
been explained to providers.
“Even though the sample size has been validated to get the response we need [I hear], ‘I don’t believe that because the sample size is so small.’” (0679, site II)

Site II is undergoing a process to allow quality metrics to be gathered automatically from the electronic health record. This will increase the sample size and give the group an opportunity to apply risk-adjustment models.

Risk adjustment was seen as more critical for individual measures than for the group measures. Interviewees expressed a view that although group case mix was reasonably stable over time, individual anesthesia providers may have widely differing practices with respect to case mix complexity and acuity.

**5.3.2 Communication around program results and implications**

The program at site I resulted in some additional episodes of peer and group communication around anesthetic practice. Interviewees reported using peers as a learning resource when reflecting on, and reacting to, their feedback. The reports inspired some departmental rounds focusing on quality improvement. No new communication channels seemed to be opened by the presence of the program. While the DA-QI reports themselves were not universally seen as a useful learning tool, the conversations and discussions that occur around them were favorably regarded as such. Improved communication to clinicians about the content and purpose of the reports as well as increasing report-related discussions within the anesthesiology group would be a useful strategy towards increasing their effectiveness. Both Edmondson and Westrum have stressed the importance of communication in fostering and sustaining a culture of learning and safety (Edmondson, 2011; D. Parker et al., 2006; Westrum, 2004).

The department previously discussed the indicator program on a regular basis during meetings but hasn’t done so for four or five years. Several newer members expressed confusion about the intent and structure of the program and reported receiving no information on it when they joined the group. These newer group members have been left to “fill in the blanks” which may decrease the likelihood that they will expend the effort to reflect upon and engage with their data. This participant has wrongly assumed that the DA-QI is a program run externally by administration for a purpose other than learning and quality:

“I can’t see that it has to do with improving our anesthetics. Because this would not be how you would go about improving the department of anesthesia’s anesthetics. So there
must be some other reason that the health region has for having initiated it in the first place. I would be curious about what that was. You’d think they would come and tell us about that…but they don’t.” (0437, site I)

Increased communication about the program from the department’s leadership might increase its perceived importance by department members and increase use. These discussions are likely to be important in the context of data that is imperfect and difficult for members to interpret. The absence of formal conversations about the program has left many group members questioning its purpose, usefulness and relevance:

“I cannot recall the last conversation I had about this, with anybody.” (0825, site I)

“The last time we got this, the last time one came out, we did have a chat about it...in the OR lounge. It was a bit of a complaining session and I don’t recall the exact nature of what it was about. Kind of like, why do we still get this? What does it mean? What are we going to do about it? What should we do about this when we get this information? What are we meant to do about it? And nobody really had an answer.” (0437, site I)

The ability to share information and collaborate with outside parties was seen as very desirable by site I interviewees when attempting to interpret and react to the metrics. Interviewees wished to be exposed to similar data from other anesthesiology departments for the purposes of benchmarking and learning about data element selection, collection, presentation and use. They also felt that sharing the structure and outcome of QI projects with their colleagues in other departments would be a worthwhile learning opportunity. Pooling data and joining registry programs was seen as beneficial so that benchmarks and comparisons could be established. The group had found little support from the health authority in the form of the resources required to join those types of programs.

“I wanted to make us part of a European post-operative pain group. It’s massive and they were looking for North American health regions to join. I wanted this because that would be such a resource, because you would have access to what happens in other places in their database and solutions to their problems...but the health region declined.” (0825, site I)

Interviewees at site II emphasized the importance of communication around quality metrics. The group had multiple structures to facilitate communication of quality data and quality improvement activities. Data is presented regularly at meetings and discussion is encouraged. The data is also available on the hospital intranet and summaries are periodically sent to providers through email and newsletters. Communication from leadership to front line staff
occurs around methods of data collection and analysis, choice of metrics, results, and any resulting action plans or projects.

“They just hear it over and over and over again. Every month they are seeing a report back on maybe 4-5 projects that are going on and then they see the dashboard and see what our measures are for quality of care.” (0295, site II)

“We have a monthly clinical steering committee meeting and really what it is, is the entire department sits together and really for an hour and fifteen minutes we just have presentations on the various projects in the department and how they are coming along...[the] surgical site infection bundle for example, has been on the radar for five years now. For five years, every month, we’ve been presented with the data on how we are doing...I think it’s important that it’s talked about frequently.” (0508, site II)

New staff members meet with division leads who orient them to the system and explain expectations and any associated variable compensation rewards. The quality director and chair of the department both placed great importance on maintaining approachability and facilitative ongoing discussion about quality metrics and improvement efforts. Interviewees at site II stressed the need for clear communication of expectations, especially around behaviors and processes that are targeted for quality metrics:

“We’ve agreed that we have to make sure we’ve made our expectations clear before we hold anyone accountable to certain behaviors or certain standards.” (0679 site II)

Individual feedback has been a component of many of the QI projects within the department. Providers are used to getting emails, phone calls or visits when they deviate from a process or fail to meet a target and will attempt to pre-empt these conversations. This provides a source of new ideas that can be fed back into the key driver framework and PDSA cycle for the project.

“So if we have somebody deviating out of our parameters I kind of talk to them to help them through the process to help them improve.” (0095, site II).

“I will get emails from people saying, ‘hey, look the reason that we failed on this specific thing is because of x,’ so they want people to know where they are having problems.” (0754, site II)

When a group of clinical providers is motivated and communication lines are open, ideas from front line providers can flow through to project leaders and be incorporated into solutions (Lee, 2010):
“I’m very blessed with the people I work with because you can throw several things at them and not only is their response that they take what you throw at them, but they come back and say, ‘you know, we were thinking and there’s about five things we need to do in order to improve this,’ and they usually can make a lot of things better.” (0508, site II)

Discussion around push back from providers was used as a learning tool for the group and the dialogue was seen as a necessary part of obtaining buy-in. Communication and discussion was often mentioned as a key strategy to overcome barriers that were associated with front-line anesthesia providers and other members of the perioperative care team. Issues with program acceptance have been overcome through discussions about the sources of the data, analysis methods and by demonstrating data usefulness.

“I do a lot of talking, let me tell you. A lot of talking, counselling - a lot of listening. And I don’t argue. I just try to give them my impression of it and get them to buy in. But we still have some very resistant outliers. It’s just managing them and knowing how to manage them.” (0295, site II)

When the program started at site II, some of the pushback from providers was in the form of personal attacks on the individuals involved in the design and implementation of the measurement system. Communication was important both to obtain the cooperation of these individuals and to keep such behavior from poisoning the working environment:

“People…actively engaging in a campaign to discredit certain things…either the data or the people. It tended to be a lot more passive aggressive – private conversations, gossiping, that kind of stuff. It was pretty ugly. I think that you need individual conversations with people…privately. You talk about professional behavior and respect for people. Why would this person who’s your colleague have an interest other than the genuine? It’s very difficult because it’s not like you know necessarily who and where this is coming from. It’s very insidious and very difficult.” (0618, site II)

Communication of the data and effect of improvement efforts on the metrics to other departments was seen as a way to increase collaboration, especially with the department of surgery.

“I think as other departments see that we are committed to providing safe, top quality care, that they are more likely to listen to what we have to say when we have data and they are willing to come along side of us in that effort.” (0047, site II)

Quality improvement is a key part of site II’s business model and every department is engaged in quality work. But this quality work can become redundant and conflict without collaboration between departments and this has been identified as a significant issue:
“I think about a year ago they realized that there is so much improvement work going on that it may be sometimes counterproductive. People don’t realize in two different departments that they are doing the same thing or that they may have done it two years ago. They needed more coordinated activity so they did [seminars] so that they were going from kind of micro systems to macro systems. We had…many divisions represented there and we each presented what we were doing and found that there was…stuff connected. So that’s when they actually connected departments, divisions and kind of made it a more systems approach versus kind of individual projects going on.” (0295, site II)

5.4 The Role of Leadership

In both departments, leadership was seen as important for initiating the program, setting the tone of its use and enforcing behavioral and attitudinal norms towards QI and patient safety. The characteristics of an organization’s leadership are important for determining culture and the ability of individuals and teams within the organization to learn (Edmondson, 1996; Edmondson, 2003; Schein, 2004; Westrum, 1991; Westrum, 2004). The program at site I had been led by the individual who designed the DA-QI but when this person took on another role, a leadership vacuum resulted. Leadership for the program at site II came not only from the chairman of the department, the QI director and others involved in quality work, but from the highest levels of management in the institution including high level executives and the board of trustees. This leadership was active and ongoing.

5.4.1 Site I – Limited leadership

The chair of the anesthesiology department at site I was approached by the medical records department with a proposal for a program to gather and distribute quality metrics in the mid 1990s. While this individual is reluctant to take credit for the program, their initial support for the idea appears to have been somewhat unique. The medical records department had approached other departments within the institution but only the department of anesthesiology chose to participate. This individual valued making an effort to measure quality even if the measurements were imperfect and the implications were not known. The program was launched without a consultation or study period and was imposed upon the department with little fanfare or discussion. Members were simply informed that blinded performance metrics, which had been pre-chosen, would be distributed to them on a quarterly basis for their private use. This may have reduced push back from department members but the lack of any expectations from leadership has resulted in a program with poor uptake and limited effectiveness.
The leaders of this group feel it is important to maintain a strong divide between quality improvement efforts and administrative concerns such as performance management, credentialing and privileging. One of the strategies for maintaining a barrier between quality improvement and performance management was through separating different kinds of leadership and responsibilities:

“The chair of the department is seen as…the boss and you don’t want this seen as performance management. You want this seen as being led by someone from the group, separate from the performance, as learning and improvement.” (0465, site I)

No leaders or champions for this program have recently emerged from the ranks of the department and the program appears to be orphaned. Members that have joined the department within the last five years are unaware of the purpose of the program and have created their own narratives in regards to its intent. This study participant, a newer department member, emphasized quality assurance over individual physician learning and improvement:

“I would hope it’s…ultimately to try to achieve the standard of care for these metrics for our patients. I think that’s probably the end goal. I think there are probably secondary goals…feedback and self-assessment. I guess the QI director is supposed to look at the overall rates of most of these things and compare them to the literature.” (0555, site I)

“You need some explanation regarding it. I mean the how and the why and the background of it. So that first day I got one as a staff person, here I was [saying], ‘What? What’s this about?’” (0437, site I)

Leadership is important for advocating and obtaining resources through advocacy and working with administration. A lack of support and investment from higher levels of leadership hampers the usability of the DA-QI program. Resources are required to collect, analyze, reflect and act on the data. The reports are often delayed or cancelled because of a lack of resources in medical records, statistician support for data analysis that would aid in interpretation is absent, and non-clinical time for individuals to react to the information and engage in QI work is very limited.

Clinicians from site I struggled with insufficient support to help them develop skills in leadership, data analysis and quality improvement science. There was frustration expressed by one individual that had pursued additional training in QI on their own time because the health authority had subsequently shifted to LEAN methodology and viewed that training as obsolete.
There were several suggestions that leadership training should be part of the undergraduate and graduate medical curriculums. Mentoring and coaching of physicians to help them develop leadership skills was seen as an overlooked but important part of the natural evolution of a clinician into an “expert.”

“You would not expect a physician to look after a patient without having some book knowledge and then supervised practical experience. You would not expect anyone to be a leader...though we do expect people to be leaders with no didactic training, no supervised coaching experience...any high performance athlete would of course have a coach, why shouldn’t each individual physician who wants to improve and excel?” (0465, site I)

5.4.2 Site II – Supportive leadership from multiple institutional levels

At site II, the top executives and board of directors consider QI an institutional priority and have invested in the infrastructure necessary to support it, including training opportunities, as well as statistical and IT support. They also make it possible for faculty members to access non-clinical time. Apart from providing the necessary infrastructure, the highest levels of management are also involved in communicating the vision and importance of the quality work to staff.

“They talk about how we have completely saturated the market locally and that the only way we can have success and grow as an institution is to attract national and international patients. And then you start talking about why would somebody get in an airplane and fly over several hospitals to get to your hospital...and one of the reasons is that we want to provide the best care.” (0508, site II)

“What I think is different about this institution is that everyone is expected to [collect and use quality data] and they say, ‘here are your resources...’ It’s an incredible infrastructure that I think is because of leadership. They are just 100% committed.” (0295, site II)

“Here I think if we don’t [collect and use quality data] we can come under fire for not doing it...and that’s what’s different, I think, about this institution. Everyone is expected to do it and they give you the resources to do it.” (0295, site II)

“What I’m finding is that you need key executive leadership support for any kind of initiative. Whether it’s service excellence and improving customer service, whether it’s improving outcomes, whether it’s creating a safety culture, anything you want to try to do, you have to have that very passionate leadership from the top...meaning that they believe very strongly in this kind of work and are talking about it all the time, championing the successes or talking about where we need to improve and they’re willing to put the resources behind making this happen...for the most part this work is non-revenue generating. When you take me out of the operating room two days a week you could say that’s a money loss because I’m not generating revenue. And yet we have
leaders at this institution that see the bigger picture and realize that improvement work is equally as important.” (0679, site II)

The anesthesiology department was one of the first areas of the institution to establish an organized system of quality information collection and analysis. The chairman of the site II department had a background in business and empirical research and a strong natural tendency to rely on empiric information in decision making.

“Quality is a number. It’s not a thing. It’s a number or a series of numbers…I was a coach of a hockey team. The first question you ask is, ‘how good is your team?’ You know, what’s the win-loss record, what are the statistics of each of the players? So when I came into the department, the business of delivering anesthesia, it’s like the business of hockey is winning games. You need to know how many games you win, how many you lose, the ratio of goals and goals against and all that. So it’s a natural question to ask.” (Department Chair, site II)

This individual became the chair of the department at a time when the board of directors in the hospital had recently articulated a vision of using quality improvement methods as a way to increase the prestige, reputation and market share of the institution. They recruited a CEO with a background in manufacturing, leveraged an individual who had recently returned from an academic sabbatical pursuing training in this area, applied for and received a large grant to support QI work, and began identifying potential leaders and sending them for training. The anesthesiology department became a pilot area for a new way of using data-driven strategies to improve clinical quality.

The chair at site II sees an ongoing need for leadership in quality work and is very involved, along with the quality director and quality team members, in leading and shaping the program. This individual articulated a philosophy that if anesthesia providers cannot deliver high quality care, the fault lies with the inability of leadership to provide the necessary resources:

“The theory is that we are all good and if we don’t achieve goodness it’s because we are missing things…so it’s not a character flaw, we aren’t saying [that] people have character flaws, I’m saying that everyone has strong character, they are trying hard, they are professionals…but we aren’t achieving the level we want to achieve, then the system, meaning the leadership, has not delivered the resources, the process and the information [for the people] to do what they are capable of…do you want to, as a leader, lead by implying that your people, the people who you are responsible for leading, are flawed? Or do I take the approach that you guys are all really good and the problem is in me helping you?” (0618, site II)
One of the key functions of leadership is to communicate the intent, nature and use of performance measurement (Edmondson, 2011). The chair provides a large amount of highly visual support for the individuals doing the quality work and also spends considerable time communicating with members about expectations for professionalism and behavior. One of these expected behaviors from all anesthesia providers is compliance with programs aimed at improving quality through process changes.

“They try to make the goals achievable but very black and white. The feeling is that they want you to be successful but it is expected that you perform.” (0508, site II)

“It often takes the chairman of your department saying that this is an important initiative, this is what we are doing now. There are no ifs ands or buts about this. No question. You cannot do something like this just on people’s good will alone. It has to be made very clear that this is a priority.” (0679, site II).

The larger quality projects are determined by leadership so that they are rational, fit the overall needs of the department, and support the strategic direction set by the institution’s board and management team. The top-down approach to choosing quality improvement projects can sometimes be a source of tension and dissatisfaction for members of the department who feel that solutions and insights “from the trenches” are not being leveraged fully.

“When we take [the six month quality improvement] course, members of our department are assigned what project they are going to do. So as much as it makes sense to say, ‘you know in my daily work I’ve noticed that this is a problem and this is what I’m going to do to improve it,’ it’s not really what happens. What happens is that you are assigned something that [leadership] felt was important for the department. You may or may not agree.” (0508, site II)

Multilevel leadership at site II is present. The board and senior executives set the strategic direction and lend importance to the work while the department leaders maintain focus and triage projects. Interviewees stressed the value of communication about priorities and quality projects from those in high leadership positions such as the CEO of the hospital.

“…the senior leadership of the institution as a whole have really made [quality and safety] a priority. Whether it’s the CEO or the chief of medical staff, the chief of surgery or the board of directors, the commitment there to quality and safety is really unbelievable. [It’s communicated] by personal appearances by important people. It’s part of every anesthesia meeting, it’s something that we hear about from our chair of anesthesia. It’s something we hear about from the chair of surgery. It’s something that we hear personally from the chief executive officer of the hospital. It’s something that we see on
our website. It’s something we read about in our emails. It’s something that comes in a video clip sent in an email to everybody…” (0047, site II)

Each project has leaders and champions that provide support and motivation for change efforts and to lend weight whenever individual feedback is used. Distributed leadership with a consistent message, and harnessing leaders that already exist within teams has been instrumental in allowing change to happen:

“The leader of the nurse practitioners is an incredibly great people person and is just really good at working with the group…really good at saying, ‘you know, I know this is a pain, but this is important for the department and we’re going to achieve it. We’re going to get together as a group and I know we can do it.” (0508, site II)

“We had a CRNA…for a while and I don’t think [their] fellow CRNAs liked emails from [this individual] saying, ‘you didn’t do this.’ We had to change it so it came from [the quality lead] or it would come from me. So I think, some of the measurements, when we don’t meet them, it’s better when it comes from the leader of that measurement, not someone else that’s helping with the project that is a peer.” (0095, site II)

Anesthesiologists from site II are fortunate to have access to a quality and safety center that employs individuals with expertise in behavioral change, adult education and knowledge translation. The quality and safety center provide accessible leadership and QI science training to front line clinical staff. Approximately 15 members of the department have received support to complete an in-house quality improvement course which includes a practicum. In addition, the department of anesthesia has a tradition of providing mentorship for junior faculty who are interested in adding non-clinical components to their practice in safety, quality and outcomes, bench research and anesthesia education.

The literature supports that team and organizational learning requires organization from leadership such as the availability of training and other supportive structures (Adler et al., 2003; Dukerich et al., 2002; Edmondson, 2002), and the effectiveness of the program at site II was attributed by interviewees to the existence of these opportunities.

5.5 Group norms and culture

There were significant differences between the approaches to quantitative feedback and quality improvement and culture between the two sites. Within the interview data obtained from site I, a strikingly strong theme of individualism and personal responsibility was found and this was highly influential when it came to the DA-QI’s design and functioning. The ability of site I to
fully leverage the DA-QI as a learning and QI tool suffered because of an immature safety culture within the surgical services. In contrast, site II emphasized a shared group responsibility for care and many interviewees felt that individuals did not have the capacity to bring about more than cursory improvement. They emphasized the multidisciplinary nature of modern surgical and anesthetic care and the principle of “multiple touches” and felt that group level metrics and organized projects within and across departments were the most worthwhile avenues for quality improvement work.

At each site there appeared to be reciprocal relationships between culture and the measurement program. The prevailing culture was reflected in the design and functioning of the program at each site and the programs in turn influenced the culture. The use of individual measures by site I reinforced a culture of individualism. The measurement program at site I did not contribute to the creation of a stronger safety culture, likely because of a lack of active leadership. Direction from leadership has been recognized as one of the most important features shaping organizational culture (Westrum, 2004). The use of group measures and goals by site II was used as part of a transformation to create and then strengthen a culture that places a high value on teamwork and continued improvement. This culture was the result by the focused and consistent efforts of many levels of leadership within the institution.

5.5.1 Site I - Individualism and personal responsibility

Members from the anesthesiology department as site I had a strong shared tendency to emphasize individual responsibility for anesthetic care quality and expectations for individual providers were high. Site I anesthesiologists generally practice independently and rarely take over cases from each other or provide break relief. The fact that anesthetic practitioners supply care in solo fashion was mentioned many times. Learning and practice improvement in anesthesia were seen as individual activities. The belief that each anesthesiologist has the power and self-efficacy needed to affect patient experience and outcome was common amongst interview participants. This belief was shared by both junior and senior physicians within the group. The feedback in the DA-QI reports is organized by individual provider and during interpretation the amount of variation between individuals is seen as important. When approaching possible solutions to this variation, individual practices were seen as the main target
for investigation and change. Relying on individuals to self-identify areas for learning and practice improvement was seen as a way to decrease the potential for blaming and shaming.

This emphasis on individuality may contribute to a preoccupation with measures that are directly attributable and dismissal of measures that could capture shared accountabilities or team functioning. There was a suggestion that an anesthesiologist’s ability to affect patient outcome ended with care transfer from the post-anesthetic recovery areas to the ward and very little traction for the idea of collecting measures with longer time lines. Measures that were not felt to be directly attributable to individual anesthesiologists were dismissed.

Strategies for practice change were also largely individualistic and often focused on first order problem solving and informal natural (before and after) experiments:

“I look to see how I compare to the rest of the group, and when I find an outlier that stays that way for more than one quarter, I say, what can I do with my practice to improve that? In my mind, it is an experiment” (0777, site I)

“For me, if I scored poorly in something, I would go look up the most relevant, recent, management guidelines for whatever I’m doing terrible at…and try to do better.” (0555, site I)

“There’s no need for audits or discussions because the information is micro. You know exactly what you need to do to improve.” (0626, site I)

The way feedback is presented to the group emphasizes individual results and individual solutions. The lack of benchmarking to other groups or published event rates deemphasizes any need for group improvement. Individual practice change was seen as a powerful way to change patient outcomes, and the loss of opportunity for group learning and system change was at times underappreciated. In the following example, an anesthesiologist struggled with logistical challenges during a case that required multiple blood products. This individual is exploring a potential solution to this problem that relies on a behavioral change. But neither the problem nor the potential solution was shared with colleagues and no system changes were contemplated.

“I know the small things I can personally do to make my next massive transfusion case better. I know what I need to do next time and if somebody gave me feedback and said, you know what, you gave too little FFP (fresh frozen plasma) too late or whatever the issues was, I can say, I know that and the reason for that is that I could only get one unit of FFP at a time…I don’t believe I can say, ‘but it’s the blood bank that didn’t give me the FFP.’ Next time I have a massive transfusion at that hospital I will make sure that I
have a unit clerk that stands outside my door to organize the porter to go to the blood bank to get the blood and the FFP and whatever I need right away…I will do the organization differently and I know that I will be able to effect that change.” (0825, site I)

When asked if the program was useful on a department level, one of the responses illustrates just how deep the sense that anesthesiology practice is an individual responsibility runs. Here, department level learning was interpreted as collective individual learning, rather than group learning:

“I start with the assumption that I’m an average person and if I look at it and reflect on the numbers in a way that impacts my practice, then I assume that other people do as well. So in the sense that I believe it probably impacts the majority of members in the department than, yeah, I think it’s good on a departmental level.” (0777, site I)

The ability to personally make decisions and practice changes to benefit patients was seen as highly personally rewarding:

“I will feel empowered if I can make those little changes for the safety of my patient. I think our systems today tend to disempower you. I feel that, I see that frustration on a daily basis in the OR. Physicians feel like they are not part of the decision making process anymore. But getting practice feedback and…correcting those mistakes or deficiencies yourself will you give that feeling…I am in control and I know what is good for this patient and I am making decisions that I need to, [in order] to save this patient.” (0825, site I).

The group does not share DA-QI information with non-anesthesia colleagues, such as OR nursing staff or surgeons, and the program was not known outside of the department. The group expressed several concerns about sharing information with other departments and with administration. There appeared to be ongoing distrust and fears of punitive action or consequences for poor performance which implies the absence of a safety culture within the organization (Edmondson, 1996; Edmondson, 2011; Westrum, 2004). Sharing and reporting of non-anonymous performance data, implying some element of public shaming, within the hospital was seen as an attempt to spur rapid change within “problem” departments:

“You will see [QCDSM metrics] posted on walls as you walk around. And in fact, for the orthopedic surgery [department], because it’s been a crisis, the data…is not even blinded. It’s up on the wall for the public to see – [WHO surgical safety] checklist performance by Dr. R., Dr. F., Dr. Jones…names and results posted on the wall.” (0465, site I)
Interviewees spoke of a complicated and not entirely workable relationship with the department of surgery which has discouraged efforts at cross-departmental QI efforts. Joint morbidity and mortality rounds were suspended several years ago because the climate did not support psychological safety and attendance was poor. These experiences made the anesthesiologists reluctant to pursue joint QI projects or to focus on patient processes and outcomes that had shared attribution with surgery.

“The surgeons were the larger number of individuals in the room. They were certainly the more outspoken individuals in the room. They were not shy to attack one another, seldom learned from the care that was provided or the problems encountered in the care and certainly the anesthesiologists didn’t have an opportunity for learning or improvement in that environment… I think there is the potential for learning to do something with surgery and anesthesia, but that experience with M&M rounds makes me really hesitant to go there.” (0777, site I)

“The moment we put the two or three departments together they became so defensive that it was just a mess.” (0825, site I)

“I distrust the head of surgery. Our department is a very cohesive department, it’s welded together, and the department of surgery is incredibly incoherent and fractured. It would be incredibly difficult to effect change in that department.” (0825, site I)

In addition some of the anesthesiologists had experiences with 360-degree feedback assessments in the health region being used by administration for human resources management. These individuals were very reluctant to have performance data published or shared outside of the anesthesia department. Prior experiences with administration resulted in distrust of the morbidity and mortality round process:

“[the anesthesia group] did not trust the senior administration that the information would be protected… they thought they would use it for disciplinary action. It is only in the last 3 or 4 years that I believe we have totally overcome that.” (0825, site I)

5.5.2 Site II – Team effort and group responsibility

In contrast to site I, where quality is seen as originating from a well-trained and conscientious anesthesia provider, quality at site II is seen as a system-based property. This may be because of exposure to different academic traditions. Allopathic medical education has long emphasized the conscientious provider and is geared towards ensuring that physicians have solid groundings in knowledge and skills in order to provide appropriate care. It is only relatively recently that medical education has begun to include elements from improvement science and systems
thinking and care has begun to shift towards team-based models. Physicians at site II have had greater exposure to improvement science and business management compared to those at site I and as a result approach quality as a system property:

“If you read the work by Deming, who is the father of quality improvement, his belief is that quality is a system-based property and not an individual-based property. While individuals do clearly influence the product, talking about the system which they are a part of and working as part of a team…that is how we affect change and quality. So we decided to measure as a department and that this is what we would stand by and who we are…not what you do and what you do and you’re better…and all that.” (0618, site II)

The structure by which anesthesia providers deliver care is also different in Site II. At this institution, the faculty anesthesiologists are salaried and supervise CRNAs, residents and fellows. Patients receiving anesthetics in this type of system may have multiple providers through hand-offs and break relief as well as through supervisory relationships. Having multiple members of one department interact with the patient in this way may lead to greater awareness that outcomes are difficult to assign to any one individual.

“There was no way we could measure individual performance because patient care is influenced by multiple touches.” (0618, Site II)

Individual performance measurement was seen by the chair of the anesthesiology department in site II as a form of human resources management that relies on identification and eradication or remediation of poor performers rather than as a tool that could help individuals improve:

“Do you want, as a leader, to lead by implying that your people, the people you are responsible for leading are flawed? And I’m going to go and weed out the flaws?...The problem with [bad apple] theory, which is widely used in the legal and human resources domains, is that it’s never been shown to be true. But unfortunately, a certain amount of leadership, if there is a problem, [finds it] easy to say, ‘I’m going to remove the bad apples, fire a few people, go pointing fingers at a few people…’ but that doesn’t lead to a strong organization or one that can change.” (Department Chair, site II)

Seeing quality as a system property and attempting to measure it required that members of the department set aside their individualism and embrace a team and system-based approach. When quality measurement was initiated, some members of the department chose to leave because they found it difficult to conform to new expectations for behavior. Since the department started talking about measuring quality, 35 faculty members have been added. This significant expansion allowed the department to guide culture change towards increased cooperation and
teamwork through recruitment of like-minded individuals. The use of group level measures was felt by many of the subjects interviewed to contribute to a sense that all members of the anesthesia department were working together towards common goals. Researchers have suggested that team based measurement may avoid the team dysfunction and anxiety that is observed with some individual performance measurement programs (Bowman, 1994; Deming, 1986; Prince, 1994; Çiçek et al., 2005). Although the members of the anesthesia department at site II are not working as a team in the sense that they are working together deliver care to any one patient, they are a group with similar interests and goals and this relationship was strengthened by group level measurement. The leadership at site II put considerable effort into supporting a culture of learning and safety at their institution and performance measurement was used as a tool to support those efforts.

“I would say that’s been the big [change] – more team-based care and respect for your teammates. So if your teammate needs [something] from you, you do it to improve the outcome as opposed to getting in an argument. You can’t just do whatever you want to do. You have to follow processes and pathways that we’ve agreed to and there will be consequences if you don’t. [The chairman] isn’t going to support you if you don’t follow those things whereas in the old regime you could do whatever you wanted to do and leadership always backed you up because you were the ‘doctor.’” (0618, site II)

“I think we have a better sense of team [since measuring quality], in that this is all our responsibly and that we really need to work together and watch each other’s back, if you will, and really bring things to other people’s attention…” (0047, site II)

“There is considerable pride in who and what we are and in what we have achieved. The fact that we are a very busy hospital and we work very hard, but we work to be the best. It gives us a goal and it gives us a common purpose and we really celebrate the achievements as they come.” (0679 site II)

This culture change was the result of the expectations and priorities of leadership being communicated clearly and repeatedly during the time of change. The culture is sustained by communication of these expectations to prospective new hires during a comprehensive orientation phase and ongoing communications to the group.

“The new expectation is that you have to work together to achieve this, and we’re all agreeing to this and moving forward. Not everybody gets to do whatever they want to as long as the results are okay. The ends [no longer] justify the means. Here, it’s the means that will get us to the ends.” (0618, site II)
By providing some training in QI science to all staff, the institution has created and fostered the shared norms and language required for a strong safety and improvement culture. This culture cuts across departments and divisions. For example, the department of anesthesiology, which has a strong QI and team culture, does not find itself having to contend with defending that culture against opposing or blocking attitudes.

“Here we say that everybody has to be inoculated…everybody speaks the same language, everybody understands run charts, everybody understands control charts, they know what PDSA is, it’s just an incredible kind of movement that’s occurring. I think that’s the reason why we are where we are…if you took this department and transplanted it somewhere else I don’t think we would have had the same outcome.” (0295, site II)

In summary, the measurement program at site II has functioned in a way that provides a common set of goals for members of the anesthesia department, a way to share information and knowledge and allows group members to feel a sense of ownership – all hallmarks of a more mature safety culture (Westrum, 2004).

An idea that emerged from interviews at site II was that the quality improvement work had created an environment where the emphasis on processes and tasks was replacing humanistic aspects of practicing medicine or nursing. A focus on tasks takes a clinical provider’s time and attention away from relationship building with patients and their families.

“Everybody wants to have that connection, it’s just making sure that they have time to do it and they have the capability to do it. What we struggle with, is we’ve taken away a lot of the time in people’s days and we’ve become very task oriented. As we try to make things as safe as we can and we try to institute all these protocols and make sure everything is done in a uniform way, we’ve taken away some of the time to…just be with patients.” (0679, site II)

“There are so many different initiatives going on hospital-wide that it has become very run chart oriented and those of us that were there in the olden days when we didn’t have run charts and things like that...sometimes we all wish it was as simple as just taking care of the patient.” (0754, site II)
5.5.3 Performance measurement can emphasize differences between types of providers

In general, at both sites, providers that had more autonomy or who were more secure in their positions within the organization were less likely to report discomfort with having their care measured or receiving quantitative feedback.

When Site I began the DA-QI program it created anxiety within department members who had completed their medical training outside of Canada and so were practicing under provisional or temporary licensure. This anxiety was not shared by Canadian-trained and foreign-trained physicians who had already completed all domestic credentialing requirements. This divide between physicians of different backgrounds resolved over time as foreign trained physicians within the group obtained domestic credentials in anesthesiology or went on to other types of practice and members saw that the DA-QI information was not used for staffing decisions or performance reviews. However, had it not resolved, this could have had a significant negative impact on department functioning and morale.

Site II uses a mixture of different types of anesthesia providers: faculty anesthesiologists, resident and fellow physicians and registered nurse anesthesiologists. The quality metrics that are collected and published quarterly apply to all of these members and the results are not stratified by provider type. The results are presented to all providers at a quarterly meeting. A large number of quality projects occurring at site II use individual feedback as a strategy to improve compliance with process measures and to identify opportunities to change workflow to help providers reach compliance. Providers of different backgrounds appeared to have different levels of reception when it came to individual feedback.

While none of the faculty anesthesiologists reported any negative associations with giving or receiving individual feedback, this was not true for the CRNAs who were interviewed. It may be that the CRNAs receive more individual feedback than the faculty anesthesiologists because they provide most of the immediate anesthetic care or it may be that differences in the provider “hierarchy” change the amount and type of feedback that is received. For example, one CRNA described an environment where failure to comply with a quality initiative resulted in the filing of incident reports by PACU nurses.
Other than spending the majority of their time providing one-on-one care for surgical patients, CRNAs may be subject to more quality improvement efforts and more individual feedback than faculty anesthesiologists because of a physician-nursing divide. Although CNRAs provide anesthetic, not nursing care, other nurses may view them as peers and thus expect a greater level of compliance with nursing-initiated projects and be more willing to provide feedback. In the example above, the CRNA felt that the feedback was not appropriate as there was a limited amount of evidence for the intervention in anesthetic care while the perspective of the PACU nurses might have been different because the type and intensity of care is different.

Although the quality measurements do not differentiate between providers at site II, the CRNAs that were interviewed perceived more opportunities for physician anesthesiologists to participate in quality improvement efforts and projects. They reported not having the same opportunities with respect to providing input into problems and potential solutions and felt they had insights “from the trenches” that were not being fully leveraged.

“I would like to see more CRNA involvement. We are staffed with an anesthesiologist that may watch two or three rooms, I mean, they are doing patient care, but we are doing the cases. I think it would be nicer if we had more involvement in the whole process and I think that if people understood more about the process they would come up with new ideas and better ideas about how to measure.” (CRNA, site II)

Quality improvement is such a vital part of this institution’s strategy and vision, and has been recognized as an academic track within the hospital and the associated university. CRNAs reported not having the same access to nonclinical time or training and this disparity means that CRNAs cannot access as many opportunities for academic and research oriented career advancement as faculty anesthesiologists.

“I’ve done a lot of quality improvement stuff [with the department of anesthesiology] but I’ve never been primary investigator because I’m a CRNA, not a physician.” (CRNA, site II)

The CRNAs and faculty anesthesiologists that were interviewed were specifically asked if quality projects had changed relationships in the department or lead to any differences in the way that faculty anesthesiologists and CRNAs were treated. They did not report that this was the case but one CRNA described choosing to work on quality projects as a way of increasing the status of nurse anesthesiologists:
“I was very interested in having them understand that CRNAs did prove to be an important part of what they did…so by getting CRNAs involved in hospital wide committees and groups they knew that we existed and the types of things that we did.” (0754, site II)

Unfortunately residents and fellows were not accessible for interviews at the time of the study. They may report feeling different pressures from the quality projects that incorporate individual feedback compared to faculty or CRNAs and this would be an interesting avenue to explore in the future, especially as data and feedback use in clinical anesthesiology is likely to increase.

5.5.4 Anesthesiologist Identity

The anesthesiologists that were interviewed from Site I spoke of the safety of modern anesthesia and improvements in safety over the years with great pride and this formed part of their sense of identity. While the DA-QI program was not felt by those interviewed to have lead in any meaningful way to this group-view, this group norm may have contributed to acceptance of the program.

In contrast, providers at site II identified with the institution and its priority of quality improvement rather than with the achievements of their specialty in patient safety. They expressed pride about being early adopters within the institution and felt that the program had raised the department’s image:

“We felt as a department that this was a significant part of what we should be doing…that we should be aware what our outcomes are and that we should be improving our outcomes…over the years it’s become part of who we are as a department.” (0047, site II)

“The department has exploded and the academic work, the reputation of the department and all this quality [work] has all been a part of this metamorphosis that has occurred.” (0508, site II)

5.6 Summary

Both of the study sites should be commended for their attempts to use performance feedback to aid anesthesiologist learning and improve patient care. It is clear that this is an undertaking that requires significant commitment and resources on the part of the individuals and organizations involved. Both sites reported struggles with using the information effectively and their difficulties and real and potential solutions are highly instructive (Tables 6 and 7). The types of
barriers that emerged from interview data corresponded to elements that have the capacity to increase Performance Improvement Capability (PIC) as described by Alder (Adler et al., 2003). Barriers that were common to both sites included a finite amount of resources such as funding for data collection and analysis and access to non-clinical time. Site specific barriers were passive leadership, limited opportunities for leadership and QI science training, and poor safety climate in site I and too much improvement work causing provider fatigue as well as dilution and replication of efforts in site II. Alder described how changes in structure, strategy and culture have the largest effect on PIC and the struggles of site I in these areas appears to have limited the effectiveness of their measurement program.

Unfortunately neither program seemed to result in significant amounts of individual learning. For site I this was most likely due to a combination of low provider buy-in, program design choices, a low level of active leadership, and lack of resources including mentorship for providers. The act of reflection when receiving feedback seems to require considerable effort on the part of the practitioner (Brydges & Butler, 2012; Eva & Regehr, 2005; Sargeant et al., 2009). The recognition of this effort, along with the consistent expressions of interest in mentorship and support from interviewees in this study as well as previous studies (Sargeant et al., 2009), suggests that resources need to be made available to individuals who are expected to learn and improve through the process of reflection.

Group level data, such as that collected by site II, did not prompt self-assessment and individual learning with respect to skills as individual anesthesia providers. The use of group level data in a program of CQI did result in learning in other areas. Providers became comfortable with basic QI science and learned to incorporate process changes into their practices in order to comply with institutional demands. By focusing on group-level metrics, site II may be ignoring opportunities to use their methods and experience with data collection and analysis to assist providers in self-assessment and skill improvement.

Group learning occurred secondary to the feedback system at site II and was made possible by strong leadership, the availability of resources and structures to hold the group accountable. Group learning at site I was minimal. Resources at site I were constrained and a strong culture of individuality may have discouraged group learning activities.
Many design choices influenced the ability of the sites to use their performance feedback. Choice of quality indicators, their presentation, the level of measurement and how they were analyzed and presented to members affected provider commitment to change and the ease in which they could understand and plan based on the data. Translating intention into action required a supportive culture and leadership and the availability of resources for learning and change.

Interviewees from both sites recognized that multiple stakeholders can be a barrier. When metrics are going to be shared with different parties, getting agreement on their purpose and content can be difficult. While anesthesiologists want practice feedback to help guide learning and practice improvement, hospital administration, insurers and taxpayers want measures that speak to efficiency and accountability. Patients may want measures of safety, quality, efficiency and accountability. Without support and resources from those in administration, most anesthesia departments would be unable to find the resources required for ongoing data collection, analysis and reporting. Compromises might have to be made by the anesthesiologists in the form of including metrics that focus on accountability and efficiency in order to secure resources from administration. The presence of these metrics may decrease the effectiveness of the program for learning and may impact the culture and morale of the anesthesia group as well as provider buy-in.

“There has to be a fire wall between saying, ‘we’re doing this to learn and improve’ versus, ‘we’re doing this to punish you if you are a bad doctor,’” (0465, site I)

Both sites had struggled with provider buy-in when the feedback programs were being developed and introduced. Site I has not been able to overcome these, but because there is a lack of expectations from leadership, the push-back takes a non-disruptive form - providers discard the information. Providers at site II did push-back against the program during its early days and this was reduced by dialogue from leaders about the importance of the program. Even after the department was able to show improvements in care, ongoing struggles with obtaining the cooperation of the front line remain.

“There will always be people that will find some way to have the data not work for them.” (0679, site II)
A lot of the cynicism and skepticism was reported to decrease over time as people realized that quality and performance data was not going to be used to single out or punish providers and that expectations from leadership were not unreasonable:

“They’ve realized that we’ve really looked to see what’s going on. And sometimes it’s a provider issue but most of the time there are systems issues that you can work to improve.” (0295, site II)

Program acceptance was increased amongst the providers by sharing results of efforts and especially having data that allowed linkages between process and system changes and improvements in patient outcomes.

A further difficulty that other centers may encounter when beginning a program of quantitative feedback for the purposes of learning and improvement is the slow rate of change. Improvements on measures and changes to the culture of site II took close to ten years. They also report that quality projects need to run for an average of three years before new processes and group learning become cemented into daily operations so that the change becomes sustainable in the absence of ongoing feedback and efforts. Change on this time scale requires stable and committed leadership and funding.

“…the transformation has been amazing. This takes time, it really does. We didn’t get to where we are in a few short years. This has been an ongoing process with lots of fits and starts and grumbling and complaining and, ‘how are we going to afford this,’ and, ‘how are we going to take time out of our busy clinical practice to measure all this stuff.’ You know, lots of questions, concerns and objections. But to see where we are today is really neat.” (0047, site II)

At the time of the study, neither site was using data abstracted from an electronic health record, including an AIMS. Site I had not implemented an AIMS and site II was in the process of undergoing data mapping and evaluation to transition from human data abstraction to electronic capture. Organizations that have an AIMS and enterprise EHR in place will need strong data governance policies and ongoing audits to ensure that high quality data is being collected. These systems allow near universal sampling of cases and the increased sample size may allow comparisons between providers as well as risk adjustment. Unfortunately neither case presented here provides direction on how to harness this potential. Given that some providers from site II reported feeling overwhelmed and fatigued by the amount of data being collected, reported and
acted on at any one time, such as system is likely to need strong leadership to set priorities and avoid exceeding the capacity of providers to change.
Table 5.2: Site I – Barriers to using quantitative feedback for learning and quality improvement and potential corrective strategies

<table>
<thead>
<tr>
<th>Barrier</th>
<th>Consequences</th>
<th>Potential Corrective Strategies or contributing factors</th>
</tr>
</thead>
</table>
| **Lack of Resources** | Limited ability to collect data  
Limited data analysis  
Delayed and cancelled reports  
Limited QI and leadership training for providers  
Limited ability to conduct QI projects | Resources could be improved by raising the visibility of the program, obtaining supportive leadership from higher levels, increasing alignment with mandated quality programs from governing and funding bodies and increasing alignment with the needs and wants of other stakeholders including patients.  
The cost of data collection could be reduced by the implementation of an EHR including AIMS |
| **Lack of non-clinical time for providers** | Limited self-assessment and reflection.  
Limited learning and communication of efforts | Tradition in medicine of providers completing CME on their own time. The expectation that physicians will engage in improvement and learning activities on their own time may be unrealistic when those activities require changes to care processes within an institution.  
Having leadership for program emerge from higher levels could result in funds for non-clinical time |
| **Low motivation for change and low provider buy-in** | Many providers ignore the quality reports or read them but do not reflect on their practice or change their behavior. | Expectations from leadership would help create a sense of urgency. Accountability structures would improve compliance. Rewards for improvement may increase motivation as well as program visibility. Ensure the consistent availability of reports based on higher quality measures. Chose measures and set goals based on group consensus and available benchmarks. |
| **Safety culture** | Immature safety culture and poor relationship with both the department of surgery and administration means that efforts are not shared and program visibility is poor. Improvements and projects are constrained to individual anesthesiology providers. | A concerted effort from strong motivated leaders to change culture over time would be required. |
| **Level of measurement** | Individual measurement can engender a sense of personal pride and accomplishment but undervalues the effect of team based care. A myopic view on individualism decreases the likelihood that system changes will be undertaken and may lead the group to focus only on aspects of care that can be modified through individual behavioral change. | Addition of group measures and cross-department measures and concerted efforts to raise group achievement through group practice or system change. |
### Table 5.3: Site II - Barriers to using quantitative feedback for learning and quality improvement and potential corrective strategies

<table>
<thead>
<tr>
<th>Barrier</th>
<th>Consequences</th>
<th>Possible corrective strategies or contributing factors</th>
</tr>
</thead>
</table>
| **Lack of resources**           | Limited ability to collect data, especially the amount of data needed for risk-adjustment  
Limited data analysis  
Limited ability to conduct QI projects | The cost of data collection could be reduced by the implementation of an EHR including AIMS  
Forming partnerships with other institutions and with registry programs to pool data could result in the ability to risk adjust even if a single institution’s data is sampled and could result in cost-sharing and leveraging economies of scale. |
| **Lack of non-clinical time for providers** | Projects may take longer to plan and complete | Balancing clinical and non-clinical time for providers is difficult for all institutions but could be improved by accurate staffing projections, stable fellowship programs and setting realistic expectations and goals based on the need to maintain clinical capacity. |
| **Level of measurement**        | Group-based measurement ignores the contributions and variances between individuals. Some metrics may be necessary to ensure all providers are meeting an appropriate standard. Scoring well on individual metrics provides a positive sense of reinforcement. | A balance of individual and group metrics to guide both system changes and personal practice changes may be ideal. The theory and practice of this is complex and currently poorly understood. |
| **Data fatigue/ QI project fatigue** | Providers have limited capacity for change. Being subject to many requests for process changes distracts and overwhelms providers and may cause disengagement from the QI program. | Increase the emphasis on provider-passive system changes when planning quality projects. Appoint team leaders to evaluate project burden and determine how to prioritize/ emphasis different changes and to cap projects affecting individual workers. Increased collaboration between units and divisions to reduce redundancy and repetition |
6 Conclusions

Theses case studies contribute to our understanding of performance measurement and feedback as a part of anesthetic practice by providing examples of how feedback may be successfully or unsuccessfully translated into individual and group learning. Some of the design and contextual elements that influenced the ability of each site to learn from the data collected have been identified and explored. Providers at each site shared their experiences with, and perceptions of, the system to which they had been exposed and their insights are valuable for those contemplating the use of quantitative performance feedback in their own setting.

6.1 Different performance measurement and performance improvement models were selected by Site I and Site II

Both of the study sites are using data collection as part of a strategy towards increasing the quality of their anesthetic care but their programs are based on different models of performance measurement and performance improvement. Site I embraces the notion of “motivated and conscientious medical professionals,” that are willing and able to autonomously incorporate individual quantitative feedback into ongoing practice evaluation and improvement. Site II built its program on the assumption that quality in clinical care is a system property and chose to focus on group level quantitative feedback. This data is used to guide and monitor organized quality improvement efforts. These models each have benefits and tradeoffs which should be considered by organizations considering the implementation of a quantitative feedback program for providers.

The program designed and implemented by site I assumes that anesthesiology providers will use their performance data to identify gaps in their care and that they will be able to strategize and act in order to address those gaps. One benefit of this approach is the potential to identify providers who are struggling with certain tasks or aspects of care so that supports can be put into place. One of the major drawbacks, as illustrated by site I, is how easy it is to underestimate the amount of time, effort and resources required by providers to understand, reflect and act on their performance data. The realities of time-consuming clinical work, difficulties in choosing salient and actionable metrics, uncertainties associated with interpreting the data and a need for resources and direction from leadership accounted for much of site I’s inability to translate quantitative feedback into tangible learning and improvement. A second drawback was that the
focus on individuals did not help foster a common quality agenda that leadership could use to motivate efforts. Most of the learning that occurred was individual learning. Mechanisms were not in place to encourage providers to share this experiential information.

The anesthesiology group at site II, in contrast, did develop a common agenda for quality improvement based on group-level quantitative feedback. This program provides an example of the amount of resources and active leadership that is required to move from a simple feedback system to one that is effective at changing practice. Group level feedback was used to design and monitor the effect of quality improvement projects and to provide linkages between process adherence rates and improvements in clinical outcomes. Much of the learning that resulted was group level learning that was crystallized into best-practice processes. The program at site II improved measures of patient care but there were tradeoffs. Elements of care that are under the direct influence of individuals may be missed and alternative mechanisms to identify opportunities and support individual practitioner learning need to be in place. Site II’s emphasis on standardization of care via protocol and process adherence led to the perception on the part of some providers that their skills were not valued and that their interactions with patients were less fulfilling. Providers that did not accept a model of performance improvement based on adherence processes left the organization.

While site II practitioners spoke of quality being a largely system-based property, they were bound by pressures through the credentialing bodies to address individual performance during yearly employee reviews. Individual performance was also thought to be important in areas such as interactions with patients and their families and gathering individual feedback was thought to be an important step in helping practitioners assess their communication styles.

The ideal quantitative performance feedback system may be a hybrid between these two models where individual and group metrics are used rationally as tools to achieve different goals. Even if site I’s program had more active leadership and greater resources, the impact may still have been less than the impact of quality improvement projects at the group level. In a complex and multidisciplinary environment, group-level projects seem to have greater likelihood of influencing the myriad of providers and processes that contribute to a patient’s final outcome than individual-level changes. That said, medical care is accomplished by individual actions within a complex system and the use of individual feedback to help practitioners reflect on their
knowledge and skills and identify areas for improvement should not be neglected. The case studies presented here suggest that both supportive leadership and quality improvement training are needed in order to equip practitioners to learn from their individual data.

There were commonalities when looking at individual or group level data selection and presentation that affected usability. These included measure selection, strategies for data collection, and data analysis and presentation. Participants shared their insights and experiences and suggestions for program improvement.

6.1.1 Design features of feedback systems that support learning and QI activities

Design features that influenced usability included metric choice, the methods used to analyze the data and how the data was presented. The methods used to communicate data to providers, leadership and management and the frequency of these communications also affected usability.

Anesthesiology providers supported measures that were perceived to address important gaps or problems in current care, that were collected and analyzed with valid and transparent methods, and that were congruent with the performance measurement and improvement models endorsed at their particular site.

Differing ideas of how “quality” is created in medical care resulted in different purposes of measurement and measurement choice at each site. At site I, interviewees spoke about wanting metrics that would provide quality assurance. They wanted to “define quality” and show that they were providing quality anesthesia. They struggled quite a bit with how to choose measures that would reflect this at the individual provider level. Their concerns were that metrics be salient, attributable and easy to interpret. Site II uses measurement within a program of continuous quality improvement. They wanted their measurements to be actionable. They accomplished this by choosing measures where the key drivers affecting the process or outcome were perceived to be under the control or influence of anesthesia providers. The salience of the metrics in the eyes of individual practitioners from site II was increased by the perception that they were being used to guide system improvement efforts.

Anesthesia providers from site I did not endorse the view that the metrics used by the DA-QI addressed important gaps or problems with current care. In some cases there was redundancy
between the metrics and other more immediate sources of feedback such as direct patient observation. In other cases, practitioners felt that the metrics did not capture conditions that were associated with significant morbidity or poor patient experience. Individual variations in the metrics that had been chosen were thought to be low which resulted in the perception that there was not much room for improvement.

The metrics chosen by site II were selected to provide a balanced view by representing the different domains of quality suggested by the IOM. The design of metrics for the department’s quality dashboard was a long process that involved practitioner input and feedback. Each metric was piloted and refined before being put into use. This resulted in the group of anesthesiology providers gaining a sense of ownership of the system and a higher degree of provider acceptance. The frustrations and problems reported by practitioners from site II are translated into new quality projects and metrics by leadership and this increases the perceived salience of the quantitative performance feedback.

Measure choice influenced not only provider acceptance and buy-in of the program at each site but also the availability of resources. Site I failed to include metrics in the DA-QI that would be important to stakeholders with the power to provide resources or a political voice that may result in increased resources. For example, metrics that would capture patient perceptions or quality of experience or those that would measure efficiency or timeliness that would be of interest to administrators within the health care organization. The connection of site II’s program to a wider variety of stakeholder groups was certainly a factor in terms of its success and relevance to the institution’s efforts. These stakeholders included members of the department of anesthesiology, the chief of the department and other anesthesiologists in leadership positions, hospital administration, the board of directors and executive team, and patients and their families. While the goals of these groups are different, the measurement program provided a common mechanism by which they could be pursued. The information that the anesthesiology providers wanted to monitor and improve their practices and support learning and academic goals was imbedded in a program that maximized the institution’s recovery of funds from CMS through improved adherence to P4P metrics and created a story that could be used to try to increase market share.
Quantitative performance feedback was analyzed and presented differently at the two sites. Site I chose to present individual scores referenced to the mean while site II chose to present the current score referenced to historical scores as well as the group-defined goal.

The reports generated by site I created a great deal of uncertainty and this was a significant barrier to their use. The group mean did not provide an adequate benchmark or standard for performance and the omission of confidence intervals did not allow providers to determine if deviations from the mean were significant. Because data was only presented for the last quarter, there was no way to easily determine how individual or group performance was varying over time. This made it difficult to connect changes in behavior with changes in performance. Most members of the department had received no explanation or instruction in the data format and it was unclear where such questions could be directed. Long time gaps between when care was provided and when reports on that care were generated reduced a perceived need for change. Without consistent reports providers couldn’t gauge the effect of any changes they did make in their practices.

Site II presents data using a quarterly dashboard which shows group performance in relation to baseline performance and a pre-determined goal. A run or control-chart is produced for each metric that shows variations over time and the current and historical difference from target performance. All providers receive an orientation to these formats when they join the department and ongoing guidance in data interpretation when it is presented at department meetings.

A presentation format that was understandable and non-ambiguous was desired by all providers. This goal can be met by providing information in relation to a clear target, including measures of error or random variance to assist in interpreting deviations from the target, and showing variations over time to allow connections to be made between performance data and behavioral and process changes. The use of formats that are accepted and widely used within quality improvement science is desirable and standardization of format within and between institutions will allow information to be shared and understood without repeated orientation sessions or the need for specialized education. Educational sessions for members of the organization can be drawn from existing material rather than being created for every unique reporting format.
Quantitative performance information must be communicated to providers and leaders in order to allow its use. Attributes of successful communication included active modalities that forced engagement, repetition, use of leaders to convey information and organized activities to encourage discussion. There was a significant difference between site I and site II in the use of effective communication of performance data. At site I, communication was passive and the lack of discussion around the results made it easy for providers to ignore the information. At site II communication used a variety of mediums – both active and passive. The results are presented for group discussion each quarter. The group forum encourages providers to actively engage with the data through interpretation, speculation and hypothesis generation.

In summary, quantitative performance feedback appears to result in more use when the chosen metrics are perceived to be relevant and actionable by practitioners and a performance gap or significant variations in performance are evident. Presenting the information so that variations over time are emphasized, and the distance to a goal or benchmark, makes it easier to interpret. When confidence intervals are included, it helps providers determine if the data represents real differences in performance or random variations.

6.1.2 Use of electronic health information systems to support data collection and analysis

At the time information was gathered for these case studies, neither site had electronic health record systems capable of capturing information for their quantitative performance feedback reports. Participants from both centers recognized the benefits that such a system could bring. Site I has no plans to implement an electronic health record including or AIMS and continues to use paper anesthetic and nursing records, with reliance on human data abstractors. Site II has just implemented a new enterprise system although they have had an AIMS for several years. Both centers reported that further resources in the form of IT support and electronic data capture would support their efforts to learn from performance data.

Site II is decreasing its reliance on human data abstractors by mapping the data from the enterprise system to their current measures in order to increase automation. The proposed benefits of this move include the ability to increase the sample size toward capturing all cases. A larger sample size will allow development of a risk adjustment model. Risk adjustment was proposed to provide two benefits – one, it would account for some of the variance over time and
between sites allowing more accurate benchmarking and data interpretation and, two, it was suggested that it would allow comparisons to be made between practitioners.

The implementation of site II’s enterprise system did not include work to define and configure the data elements needed to seamlessly transition the feedback system to electronic data capture at the time of the go-live. While the vendors do provide some, “out of the box” functionality for generating reports, it is largely geared towards providing administrators with operational data rather than providers with clinical data. In order to support clinical data needs, the group has needed to make changes to the templates for structured clinical documentation and in some cases has chosen to redefine outcomes using data elements that are easily obtainable.

The use of electronic health record information to provide feedback to providers is an emerging area and one to which greater attention should be paid during the design and implementation phases. An explosion of work in this area is to be expected as an increasing number of sites come on-line.

6.1.3 Supporting structures and resources

There was a large difference in the resources available at each site for providers wanting to learn and modify practice in response to quantitative feedback and the amount and type of supportive structures that had been placed around the program.

Site I did not surround its quantitative feedback program with supportive structures. Departmental rounds provided a forum during which quality and clinical problems could be discussed but quantitative feedback had not been a recent or recurring agenda item at these meetings. A forum that would facilitate sharing ideas and learning from peers was lacking. Providers from site I also had difficulties accessing training in the areas of leadership, systems change, and quality improvement science. Other resources for learning, including non-clinical time, and mentorship were scarce. Participants from site I spoke of feeling overwhelmed by the demands of daily clinical work which was an impediment to learning.

The quantitative feedback program at site II was supported by a host of organizational structures including an institute that provided QI training through several different formats and intensity levels and access to expert consultants in the areas of statistics, change management, knowledge translation and adult learning. Other organizational structures that were supportive included
forums and dedicated time for discussing the data and sharing ideas and experiences both within the anesthesiology department and across different clinical departments. Providers at site II were also able to access non-clinical time in order to contribute to group learning and improvement projects and mentorship was available to assist junior faculty in developing these skills.

The presence of supporting structures and availability of resources were key features that allowed site II to generate learning and improvement activity from their quantitative performance data.

6.2 Integration of quantitative feedback into self-assessment and learning strategies

The process of integrating quantitative performance feedback into self-assessment for the purpose of learning is complex and requires significant effort. The differences between the amount of learning and change occurring at the two sites in response to quantitative feedback are partly due to resource availability. The steps needed to learn from the DA-QI reports at site I are time intensive and require special skills but providers have not been given the resources necessary to accomplish these tasks. In addition, the strength of the motivators built into the system at site I do not outweigh the amount of effort required for learning and improvement. Site II has provided the resources required to build quality improvement capacity and support learning projects. The motivators employed by this institution balance the efforts required of the providers to generate learning and improvement.

The effort needed to generate self-assessment and learning from individual quantitative feedback was underestimated when the program at site I was designed. Providers at site I did not have access to supported non-clinical time to engage in the resource and effort intensive process of reflection, learning and practice change. When feedback from the metrics was incongruent with other sources of feedback, such as that received from nurses or patients, it was perceived to be the least useful by clinicians at site I.

Individuals at site I need to add several layers of interpretation on the DA-QI report before it provides information that can be integrated into self-assessment. First, the mean group performance needs to be compared to a benchmark or goal and a measure of error needs to be available to determine if the group score represents a significant deviation from this standard.
The use of goals or benchmarks is important because there is no guarantee that the group mean represents optimal care. Individual scores could then be compared to both the group score and the goal or benchmark to determine if significant deviations are present. Individual scores are not risk adjusted so that individuals may need to identify a comparison group of peers with similar practice patterns.

If a performance gap is identified, the provider then needs to analyze the actions and processes that affect that particular outcome and plan changes. This is a skill set that is underrepresented in medical training and this skill gap is likely one reason why clinicians at Site I reported either ignoring the DA-QI reports, or attempting small natural experiments based on first order solutions. Practice changes resulting from DA-QI feedback would need to be within the realm of control of the individual provider and are not likely to be sustained if they are not congruent with the current systems and culture of the organization.

Both difficulties with interpretation and the temporally removed nature of the feedback contributed to the tendency for clinicians to discount it. Publication of the DA-QI reports has been delayed so that the time gap between the most recent reports has been over 12 months. They currently do not provide the kind of temporally close and ongoing feedback that could help guide practice change.

Quantitative group-level feedback at Site II was used to help design and monitor quality improvement projects. These projects often used individual feedback via conversations or emails to highlight occasions where processes were not adhered to. Respondents from site II did not feel that this individual feedback informed their self-assessments but it was effective at changing behavior. It also served as a way to begin a discussion about features of the system which contributed to process failures. Frequent prompt feedback caused individuals at this site to learn to pre-empt these discussions by offering explanations and reporting system problems even prior to being contacted. This resulted in a feedback loop so that information and solutions from front line workers could inform QI projects and initiatives.

The program at site II emphasized collective learning and there were many structures in place to allow the communication of performance information as well as the strategies being used to tackle problems and encourage collaboration between individuals. A key decision that helped in supporting learning was to collect data that allowed testing of linkages between processes and
outcomes. For example, site II measures both adherence to a surgical site infection prevention bundle and the incidence of surgical site infection. Several of the respondents from site II spoke of the motivating effect on the group when infections decreased during the same time frame that use of the bundle increased. Illustrating these linkages with in-house data was highly motivational for site II clinicians and resulted in greater support for the data gathering program and buy-in for other quality efforts.

6.3 Perceptions of quantitative feedback systems

None of the providers that were interviewed perceived quantitative performance feedback or measurement as an inappropriate or ineffective learning tool. There was sentiment from both sites that quantitative feedback should not primarily be used for staffing, privileging or disciplinary purposes.

Participants from site I felt that the DA-QI could be effective if different metrics were chosen, the information was presented differently, and more resources were available to support efforts. Members of the anesthesiology department at this site expressed regret at the opportunities for improvement and learning that were missed by the current system.

Participants from site II viewed their system of quantitative performance feedback as a strength of their organization and department. They described the measurement program as one tool had helped their organizational culture place a greater emphasis on safety, collaboration and professionalism. Respondents from site II connected the concept of professionalism with continuous improvement and being held accountable for results.

At both sites a tension between the conditions required for system change and the view of healthcare workers as autonomous experts existed. Site I’s culture of individualism and emphasis on individual measurement and individual solutions overshadowed and reduced opportunities for group learning. The environment of Site II, which emphasizes uniform practice and group learning, led some providers to feel that their expertise was undervalued. There was frustration with QI projects that required providers to enact individual behavioral changes (i.e. “do better”) in order to conform. This frustration was higher when good quality evidence for the process was lacking and when the institution’s own data did not yet show that adherence to the
new process was associated with outcome improvements. Attempts to embed processes into workflow before the evidence is available that the process is of benefit should be questioned.

Individual providers felt that being asked to perform new processes perfectly was burdensome and receiving feedback only about their process failures gave the impression that their clinical skills were being ignored. Respondents from site II mentioned feeling distracted and fatigued by a large number of reports and projects. There is a limited capacity for change and members reported feeling overwhelmed by the number of improvement projects occurring around them and amount of behavioral change they were being asked to accomplish. This sentiment was more common amongst the CRNAs interviewed compared with physician anesthesiology providers.

There may be a difference in how performance measurement and feedback is perceived by anesthesia providers of different types and autonomy levels. This difference could be the result of historical factors, different socialization during training, different perceptions of skill and worth with respect to provider type, or a difference in willingness to reveal these types of perceptions.

Neither site had experienced a significant amount of non-desirable learning on the part of clinicians such as learning to manipulate metrics. Neither of the sites studied provided optimal conditions for exploring this. Site I’s program had no built-in element of accountability or consequences stemming from it so that there are no perceived advantages to metric manipulation. Site II’s program measured group achievement and dilution within the group would decrease any incentive for individuals to manipulate their personal results.

6.4 Contextual elements that support or impede learning and change efforts in response to quantitative feedback

The single most important contextual element that contributed to the amount of quantitative performance feedback use at each site was the quality of leadership provided. The activities of leaders, and the involvement of different levels of leadership, influenced all of the other contextual elements including resource availability, support, extrinsic and intrinsic motivations, organizational culture and the ability of the site to overcome barriers.

Leadership within the anesthesiology department at site I was not active in managing the measurement program; there was no identified individual or team charged with maintaining it
and no strategy at the group level for data use. Leadership from site I expected individual providers to analyze their data and use it to help guide practice changes but this expectation was not made explicit to members on an ongoing basis and it was not tied to any mechanisms that would have increased accountability. Site I’s program did not receive any meaningful leadership support from outside of the department.

In contrast, multi-level leadership from site II actively promoted and maintained that site’s quantitative performance feedback program. This leadership was provided by the board of trustees and executive management, leaders within the anesthesia department including the chair, QI director, and from champions for each metric and project being undertaken. One of the functions of leadership was to raise program visibility and relevance in the minds of anesthesia providers. At site II the highest levels of leadership within the organization frequently communicate their expectations for data-driven improvement efforts, teamwork and compliance to front line staff through personal appearances, newsletters and emails. These expectations are also communicated by the chair of the department and other leaders within the group. Leadership within the department provided a strong sense of direction through prioritizing projects and targets for the change process and by ensuring that every new department member is educated about the system, metric interpretation and use, and their responsibilities.

Leadership has a profound effect in shaping institutional and group culture. The feedback system at site I was designed by leadership in order to not interfere with a culture that emphasized the individualism of clinicians. The program was not actively managed and this “hands-off” style was congruent with, and further entrenched individualism within the group. Seeing each anesthesiologist as an “island” of knowledge and skills did not encourage the cross-colleague communications that are required to propagate knowledge throughout a group or allow multiple individuals to collaborate on projects at the system level.

The feedback system at site II was used by leadership there as a tool to shift institutional and group culture. At the institutional level, mandating a continuous improvement program for the department, facilitating academic promotion for QI work, and tying some of the compensation package for providers to achievement on quality metrics was a way for upper management to emphasize the importance of these programs for the institution’s ability to be competitive. Leadership at the group level entrenched the importance placed on quality improvement by
upper management by echoing a similar message, recruiting like-minded clinicians to the department and by forcing engagement with the quality information and feedback system. The result is a group that is very aware of, and places a high importance on CQI and by and large are willing to conform and standardize their anesthetic practice to comply with institutional policy and procedure.

Participants from site II shared some of the difficulties they were facing in their daily anesthetic practices as a result of this culture. These included the impression that their skills and judgment were undervalued, and the reality of quality improvement project demands distracting them from immediate patient care activities. Examples of the latter included: internal visitors to the operating room commenting on compliance violations during critical times such as induction or emergence from anesthesia and receiving compliance violations for failure to complete a small task in the context of a complex, critical or unstable case scenario. For example, one practitioner recalled that during a value project aimed at reducing fresh gas flow rates and thereby anesthetic gas usage, they were questioned about their appropriate use of a high fresh gas flow rate during anesthetic emergence and volatile agent washout. In another incident, a practitioner was made the subject of an incident report for failing to put a new positive pressure cap on an IV access port. These experiences created some job dissatisfaction and were seen as a job stress and may not be entirely consistent with the concept of a just culture. Some clinicians believed that leadership at site II could have provided more support and insulation of their staff from these experiences by prioritizing quality projects and placing them in context and providing praise and recognition for staff involved in complex case scenarios.

One of the ways that leadership influenced usability was by generating and harnessing motivators for engaging with the system and changing practice. Intrinsic motivators were assumed to be present at both sites but only site II leadership actively appealed to these when addressing front line providers. Site II also used extrinsic motivators such as variable compensation and the potential for academic promotion. This difference was due in part to organizational human resource differences between the two sites and in part to philosophical differences. While interviewees from site I had philosophical arguments against using extrinsic motivators such as monetary rewards, their program wasn’t receiving the same amount of attention and buy-in from practitioners as the program at site II.
Site I anesthesiology providers are remunerated via a fee-for-service model by a universal insurance plan. Leadership within the anesthesiology department at site I do not have the power to provide any monetary motivation to anesthesia providers in order to encourage certain behaviors. There are areas in Canada that are considering pay-for-performance programs and if there is alignment between the interests of clinical groups generating quantitative performance information, the health authorities and the insurer, then rewards may become available in these areas. The providers at site II were salaried and leadership at that institution was able to tie variable compensation to group performance.

Site II clinicians have been exposed to pay-for-performance through CMS (Medicare and Medicaid) while site I clinicians have never been exposed to a similar scheme. Familiarity and experience with tying compensation to performance may play a role in the differing views encountered. Site II had great success improving their results when yearly bonuses or variable compensation was tied to achievement of group goals and their impression was that extrinsic motivators of this kind are very effective at getting the attention of clinical providers. The features that allowed this element to contribute to program success seemed to be clear communication of expectations and support from leadership in conjunction with attainable goals set with the input of group members. The effect of extrinsic rewards in organizations with less evolved learning and safety cultures may be different. In other studies, there is a suggestion that the use of extrinsic motivators associated with performance measurement may decrease the functioning of teams (Çiçek et al., 2005).

Instead of propagating arguments against extrinsic rewards in health care, policy makers and program developers should look to evidence of their effect in motivating clinicians to achieve group and individual goals. Intrinsic motivators may not have the power to result in lasting change especially if it is felt that current performance is adequate. Extrinsic motivators send a message to providers that the current level of performance is expected to improve and provide a level of accountability. This strategy worked well for site II when external motivation was provided through variable or bonus compensation and when no negative motivators were used.

Both sites thought that appealing to the professionalism, compassion and empathy of clinical providers was a powerful motivator with respect to learning and quality improvement. Harnessing competitiveness was seen as a way to motivate clinicians by interviewees from both
sites. The relative effect size of extrinsic versus intrinsic motivators was not elucidated in this study and further studies that would look at this issue with clinical providers would be helpful for guiding the design of quantitative performance feedback programs. What does seem to be important is that the strength of motivators on which the program rests be sufficient to counteract the amount of effort required to learn from quantitative performance feedback.

Leadership was also necessary at both sites to obtain resources to support change and improvement efforts. The availability of resources was a barrier common to both sites. After inception, the program at site I had limited impact because of an institutional culture not supportive of change, an administration not willing to commit resources, and leadership that was unable to combat these forces. Within the anesthesiology department, leaders could improve the choice and design of quantitative feedback and advocate strongly at higher levels within the institution for the resources required to produce timely reports and provide the time for department members to discuss them and generate action plans based on them. Unfortunately, the feedback is seen as a private matter and the program begins and ends within the anesthesiology department – a view that certainly doesn’t assist in obtaining resources which might improve its impact.

Research into 360 degree assessments indicated that supporting structures are necessary in order for clinicians to learn from their feedback (K. Overeem et al., 2009; K. Overeem et al., 2012). Learning from the measurement and feedback system at Site I was hampered by the lack of supporting infrastructure including poor or no training with respect to report interpretation, no statistical support, and little to no support in the realms of root cause analysis, change management or mentorship for individual clinicians. Interviewees from site I frequently reported using first order problem solving and work-arounds when discussing both problems they encountered in their daily practice and their performance metrics. Without tools and mentorship, several of the interviewees were not confident in their ability to choose behaviors or process modifications that would improve their results. Supporting structures such as training in QI methods, supported non-clinical time, IT support and mechanisms to facilitate group learning would give busy clinicians the resources needed to consider higher levels of problem solving.

Site II’s program had the support of the institution’s highest levels of management, a dedicated budget and access to supporting structures. Anesthesiology departments that are looking to
emulate the successes of site II will likely need to obtain similar levels of support including a sustainable funding plan.

A barrier that was common to both sites was initial resistance to measurement and using measurement to change processes and behaviors. This barrier was rooted in organizational culture at the time of program implementation.

The strong culture of individualism at site I, where anesthesiology care quality is seen to be determined by the effectiveness, attention and diligence of individual anesthesiology providers, informed the choice to measure on the individual level. Introduction of measurement and feedback did not change this culture and may have reinforced it. Because the measures are for individual consideration and do not feed into any group activities, the program created by site I could become eroded or ended without causing any external pressures on the anesthesiologists.

In contrast, leaders at site II viewed quality as a property of the system that could be influenced by group effort. The use of group level measures to provide feedback, group goal setting and projects aimed at improving group metrics, have contributed to a feeling of cohesiveness and teamwork at this site. The department of anesthesiology, along with the other clinical departments at this site, is expected by the executive team and board of directors to gather data and show action resulting from it. The external pressures felt by the anesthesiology department at site II help to keep their measurement program relevant and necessary to the continued functioning of the department. The concerted work of leaders at site II to emphasize teamwork and continued improvement has shifted organizational culture towards the generative ideal of Westrum().

6.5 Guidance to other centers and areas for future research

Analysis of these two cases may assist other groups who are contemplating the use of performance measurement so that they may design programs that will support their clinical work, institutional goals and their vision of an ideal organizational culture. The contribution of active leadership from multiple levels within an organization to the success of such an endeavor cannot be over-emphasized.
In order to support clinical work, measures should capture aspects of anesthetic practice that contribute to positive patient outcomes and be actionable. A group of clinicians looking to add performance measurement to their quality improvement strategy should attempt to choose measures that solve their real and immediate problems rather than only selecting those that have been published or recommended as quality indicators in the literature.

Elements of program design that led to successful use included: choosing metrics that were supported by clinical providers as well as administrators and patients, including clinical providers in the process of choosing metrics, setting goals to create a sense of ownership and ongoing dialogue, presenting metrics in a way that decreased ambiguity of interpretation, and creating channels for consistent communication of results throughout different levels of the institution. Group learning appears to be reinforced by evidence of efficacy such as seeing improvement and meeting pre-determined goals and benchmarks. It is highly desirable to measure both outcomes of interest and key driver processes. Linking process changes to improvements in outcome measures was highly motivating and gratifying for the group at site II and provides evidence that can be used to advocate for, and obtain, program resources from the institution. Comparative outcomes research via large database projects such as the National Anesthesia Clinical Outcomes Registry (NACOR)\(^2\) and the Multicenter Perioperative Outcomes Group (MPOG)\(^3\) can help identify the possible key drivers for a range of anesthetic and perioperative outcomes. Performance reporting that includes both outcomes and process adherence related to key drivers can help provide causal linkages that increase clinician buy-in and motivation. Similarly, data that has the potential to link structural elements to process adherence and structural elements to outcome would be valuable but, due to the difficulties inherent in changing structural elements, will most likely require multi-center data.

Although site II has achieved significant success through their measurement program, improvements in their methods are still possible with respect to capturing the effects of multidisciplinary care on patient outcomes. The measurements used by the organization are largely organized by division or department and duplication occurs. This represents a barrier to

\(^2\) [https://www.aqihq.org/introduction-to-nacor.aspx](https://www.aqihq.org/introduction-to-nacor.aspx)

\(^3\) [https://www.mpogresearch.org/](https://www.mpogresearch.org/)
improvement and change. An ideal future state would be to link key drivers from multiple specialties including anesthesia, surgery, nursing, physiotherapy, nutrition and other ancillary health services to patient outcomes in a more holistic manner. This will require institutional structures and support to allow multidisciplinary collaboration and improved communication.

Despite the success at site II in using measures to improve group performance, neither of these two sites is successful in the use of individual feedback to drive anesthesiologist learning and quality improvement. However, they do provide clues as to the barriers that would have to be removed for a best case scenario. Site I provided quantitative individual feedback that was largely ignored because of delays in communication, a lack of leadership and lack of clinician support. Site II provided qualitative individual feedback in response to process failures which was effective because it was nearly immediate, personalized, and because there was a strong signal from leadership that certain behaviors were desired and would be rewarded. Individual feedback systems that work will likely result from designing measures that fill the requirements of the American College of Cardiology Foundation/ American Heart Association (ACCF/AHA), “valid, reliable, actionable, and measureable,” (Spertus et al., 2005) coupled with expectations and support from leadership including explicit goals (Ende, 1983; van de Ridder et al., 2008) and the resources required to attain them. There were differences between physician and nurse anesthesia providers at site II with respect to attitudes towards receiving feedback and the optimal strategies to help providers accurately assess their skills and improve may be different for different types of providers or providers at different stages of their careers and/or learning trajectories.

Before offering individual feedback it would be prudent to consider the investment required by clinicians in terms of time in order to reflect, learn, and change. Care providers are likely to need help developing and maintaining self-assessment skills. Leaders should find ways to provide them with a sense of self-efficacy to support not only individual learning but also collaboration and system change. The optimal amount of, and ongoing nature of quantitative clinical feedback has not been well established.

Even with this best case scenario, an individual-based feedback system may not be as powerful for improving patient outcomes as a system based on group achievement. The complexity and team-based nature of modern surgical care is not reflected by individual level metrics. The
sample size of cases generated by any individual anesthesiology provider may not be sufficient for comparisons to be made between providers or for one provider over time. This lack of power is due to a complex practice mix with a lot of variation between individuals and over time.

Choosing multiple levels of measurement is likely to enhance system efficacy. Individual-level measures at site I did not result in much system change or group learning and the addition of group measures could potentially contribute in that regard. There could be opportunities for improvement under the control of individual providers at site II that are not captured by their data system. Traditional methods of ensuring the competence and safety of individual anesthesia providers such as random chart audits, allowing reporting of concerns by all members of the clinical staff and the systems for filing complaints and civil suits could be supplemented by individual level data. Performance measurement at the group level does not help individual clinicians accurately self-assess or choose areas for targeted improvement. The structure and content of performance measures that would best support individual clinician reflection and learning are not fully known and could be a focus of further research. When clinicians are exposed to group or individual performance measures, along with tracking measure improvement, attention should be paid to further investigating how the data is integrated into self-assessment activities and used for learning so that future feedback system design can benefit.

In the realm of patient safety, the concept of a “just” culture speaks to a balance between the responsibilities of individuals and the responsibilities of the hospital or institution in contributing to patient safety. It is possible that viewing quality in the same way, as resulting from a balance of individual acts and system functioning, and using that viewpoint as the basis for creating a program of performance measurement would result in a feedback system which is more powerful than either site that was studied.

In the analysis of safety breaches, Reason suggested a model he termed the “culpability tree” which consisted of a series of tests that could be conducted to analyze the contribution of individual acts and system functioning to the incident (J. Reason, 2003). This framework has been adopted by health care systems in North America (Etchells, Lester, Morgan, & Johnson, 2005) and within the NHS (Meadows, Baker, & Butler, 2005). When analyzing the performance of a practitioner that differs significantly from peers, adaptations to this framework could be helpful. An approach similar to the culpability tree could also help to decrease practitioner
anxiety around performance measurement as the tree emphasizes the potential contributions of 
the system to the outcome such as poor communication of expectations and requirements.

Monitoring the changes in individual scores, as well as the group score, when a process change is 
implemented, could identify practitioners that are slow adopters. The inability to change could 
be associated with poor buy-in, not understanding or agreeing with the rationale behind the 
process change, or could represent a practitioner that is struggling. It could also represent a 
practitioner thoughtfully customizing care for particular patients. Without targeted performance 
measures at the group and individual levels, the important interplay between standardization and 
customization in clinical medicine will go unmeasured, and perhaps underappreciated. The 
measurement programs at Intermountain Healthcare (Berman, 2007) and the Boston’s Children’s 
Hospital SCAMPS program (Darst et al., 2010) both recognize the value in recording and 
analyzing the reasons why clinicians deviate from suggested process and use this information to 
refine and improve processes. These programs effectively place clinician judgment above 
bureaucracy while still leading to improvements in patient outcomes overall and increased levels 
of standardization.

Practice change activities must be perceived as attractive enough to draw the attention of busy 
clinicians from time consuming and complex care activities. The success of site II with practice 
change at the group level was made possible by the support of leadership and provision of 
adequate resources. These resources included human data abstractors, IT and statistical support, 
protected and compensated non-clinical time and access to experts and training in QI, change 
management and adult learning. This level of support requires that the measurement program be 
important to the continued success of the institution. For site II, this pressure came from a desire 
to increase their market share of national and international referrals and from P4P programs and 
from a conviction that measurement is a key player in helping groups improve their performance. 
Institutions in other health systems may not face the same pressures and it is unlikely that they 
will perceive performance measurement and feedback as a priority investment without 
environmental factors that mandate data collection and reporting or that link performance to 
funding or organizational success.

Most institutions already have structures in place that support change, quality improvement and 
clinician learning. The practical challenge for most anesthesia departments wanting to use data
to inform their improvement and learning efforts will be to leverage this existing infrastructure. As these resources are finite, proof of concept projects and the ability to construct a business case may help these departments receive greater consideration. Toward this end, including measures of cost-effectiveness and developing the skills to estimate cost savings will be important.

It is interesting that both of the performance measurement systems studied developed independently of other information systems in the health care setting. As more centers are implementing electronic health records and AIMS we are likely to see an increase in innovative ways of using data to improve clinical care. One of the challenges for the future is to integrate data from various health care information systems in a rational way to allow monitoring of the many facets of patient care, regardless of provider discipline, and their effect on patient outcomes. Such a system would likely have a profound social environmental effect on health care including changes in traditional patterns of communication and in the provider hierarchies typically found in health care settings. For example, a health care setting where a patient’s outcome is attributed to the principal surgeon is likely to have a different implicit or explicit social hierarchical structure compared to one where that outcome is shared between providers from surgery, anesthesia, internal medicine, nursing, physio- and occupational therapy and others.

These social and environmental changes and their effect on clinical practitioners when they are exposed to feedback on their practices from information systems is an area that could benefit from further study. Providers may be subjected to excess stress when feedback is provided in the absence of a just culture that could lead to poor provider retention, increased rates of burn-out, dissatisfaction and other negative effects including health effects on the providers themselves. Providers that worked within site II that did not adjust to the new expectations were able to move to other institutions. If performance feedback systems become widespread within medicine, such providers will be forced to integrate and conform and the best ways of helping them do so will require attention. Training programs within undergraduate and graduate medical education may need to be adjusted to provide a stronger grounding in CQI. There is a fundamental tension between the ideas of CQI, which value conformity and standardization, and the traditional view of the physician as an autonomous and accomplished expert. This tension will require negotiation as more feedback is provided and higher levels of conformity to policy and
procedure are expected as these are likely to be seen by clinicians as erosions of their autonomy. The impression that the information system is monitoring care may allow busy clinicians to reduce mindfulness during their daily work. While heavy involvement of clinicians in improvement work and allowing their knowledge and experiences to shape improvement projects and priorities is one way to mitigate this tension, the time demands of most clinical practices and the high value placed on clinical work over other kinds of work within medicine make changes difficult.

These are difficult and complex questions that are not easy to study with quantitative data, the approach that is more familiar to most physicians and stressed during medical training. Further quantitative evaluation of practice feedback systems can evaluate their treatment effect and those studies should be accompanied by for further qualitative study of the cultural and social effects as well as the mechanisms of learning that support them.

The experiences of these two sites can be added to information gleaned from the experiences of health systems that have implemented pay for performance programs and other disciplines within medicine that have used quantitative feedback as a learning tool. It is hoped that the experiences of these two centers may be helpful to other anesthesiology departments as they formulate data use strategies. Such centers should be encouraged to be mindful of the effects of these systems on provider functioning and the social context and culture in which medical care is delivered as this area warrants further study.
7 Appendix A - Descriptive and contextual data collection form

Date:

Investigator:

Site Description

Nature of organization (academic/teaching vs community, profit business model and ownership)

Number of ORs:

Surgical volume:

Number of surgical beds

Number of surgical or general ICU beds

Specialized surgical programs or designations (i.e., trauma center, transplant, cardiac, etc.)

Number of clinical providers in the anesthesiology department

Quantity of different types of anesthesia providers (Fellowship, board cert, CRNA, GPA, Resident, Students, other?)

Anesthesia services provided (in and out of OR locations, clinic specialty services, telehealth, etc.)

Site characteristics:

Location

Referral base

Size of urban setting and demographic makeup

Number of nearby competing inst/organizations
Description of quality system

Data collection

Data sources

Data transformation

Data cleansing

Data validation

Data security

Data governance

Data Analysis

Risk adjustment

Predefined vs ad hoc query capabilities

Special queries

Users

Data presentation method

Data Reporting

How is the data organized in the report/ report content

Who receives the reports

Frequency of reports

Obtain copies of reports
Data use

What happens to the report data? How do the recipients use it?

Uses other than performance reports (e.g., comparative outcomes research)

Quality Improvement

Who is responsible?

Training?

Support?

QI Strategy:

Education

Projects

Special committees/working groups

Order sets

Clinical pathways

Other
8 Appendix B – Interview guide

The purpose of this interview today is to gain an understanding of your impressions of your group’s system for gathering, analyzing and using performance data to improve anesthetic practice. This interview should take approximately 60 minutes and will be confidential. There will be an opportunity at the end to ask questions. You may withdraw from the study at any time, and your responses will not be used.

Introduction

Obtain informed consent

Check and start audio or visual recording equipment

Questions:

Can you describe your roles within the department of anesthesia and within the quality system?

Prompts: Time spent and responsibilities, Support (non-clinical time, sessional, salaried), activities

What has the impact of the system been?

Has the system changed relationships amongst the staff?

Has the system impacted your clinical practice? (examples, changed workflow, created or changed relationships, had an effect on reflection or discussions with colleagues)

Do you think that your culture has been influenced by this system? In what way? How did those changes come about?

Do you think the system has improved quality? How?

Do you think the system is useful? Why or why not.

What elements are critical for its functioning?

What strategies for using the data have been most useful? Why?
Why was the system developed?

How has the system evolved or changed since it began? What caused those changes?

What roadblocks were hit during the development and implementation of the system? How were these bypassed or resolved?

Are there currently any barriers to the use of the system? How could these be bypassed or resolved?

What has the role of leadership been in the functioning of the system? Where have the leaders come from?

What has the role of management been?

Is there anything else that you think has been important to the functioning or use of the system that you would like to share with me?

Is there any functionality to the system that it currently doesn’t have but that you would like to see?

**Background information:**

Date of interview:

Interviewer’s name:

Participant’s name:

Preferred reach by email or phone for follow-up and contact information?

Training and years in practice

Subspecialty area or subspecialty training

Years with the current organization

**Closing**
Thank you for your time.

Do you have any questions for me?

Remind the participant that contact information is on the information letter they received at the beginning of the interview.
9  Appendix C – List of abbreviations

ACC: American College of Cardiology

ACGME: Accreditation Council for Graduate Medical Education

ACS: American College of Surgeons

ACSTQIP: American College of Surgeons Trauma Quality Improvement Project

AHA: American Heart Association

AIMS: Anesthesia Information Management System; also Anaesthetic Incident Monitoring Study

AQI: Anesthesia Quality Institute

ASA: American Society of Anesthesiology physical status classification. “ASA I a health normal patient, ASA II a patient with mild systemic disease, ASA III a patient with severe systemic disease, ASA IV a patient with severe systemic disease that is a constant threat to life, ASA V a moribund patient who is not expected to survive without the operation, ASA VI a declared brain-dead patient whose organs are being removed for donor purposes.” (www.asahq.org/resources/clinical-information/asa-physical-status-classification-system)

CABG: Coronary Artery Bypass Grafting

CDS: Clinical Decision Support

CEO: Chief Executive Officer

CIHI: Canadian Institute for Health Information

CLABSI: Central Line Associated Blood Stream Infection

CME: Continuing Medical Education

CMS: Center for Medicare and Medicaid Services

CRNA: Certified Registered Nurse Anesthetist
CQI: Continuous Quality Improvement

CVC: Central Venous Catheter

DA-QI: Department of Anesthesia Quality Indicators

EHR: Electronic Health Record

ERAS: Enhanced Recovery After Surgery

FPPE: Focused Professional Practice Evaluation

FTE: Full-Time Equivalent

HITECH: Health Information Technology for Economic and Clinical Health

IOM: Institute of Medicine

JCAHO: Joint Council on the Accreditation of Health Organizations

MHAUS: Malignant Hyperthermia Association of the United States

MPOG: Multicenter Perioperative Outcomes Group

MSF: Multi-source Feedback

NACOR: National Anesthesia Comparative Outcomes Registry

NSQIP: National Surgical Quality Improvement Project

OPPE: Ongoing Professional Practice Evaluation

OR: Operating Room

P4P: Pay for Performance

PACU: Post-Anesthetic Care Unit

PBLI: Problem Based Learning and Improvement
PDSA: Plan-Do-Study-Act

PIC: Performance Improvement Capacity

PONV: Post-Operative Nausea and Vomiting

PRAN: Pediatric Regional Anesthesia Network

QCDSM: Quality-Cost-Delivery-Safety-Morale

QI: Quality Improvement

REB: Research Ethics Board

SCAMP: Standardized Clinical Assessment and Management Plan

SCIP: Surgical Care Improvement Project

STS: Society of Thoracic Surgeons

VA: Veterans Administration

WHO: World Health Organization
10 References


Collins, K. M. T. (2010). Advanced sampling designs in mixed research. current practices and emerging trends in th social and behavioral science. In A. Tashakkhori, & C. Teddlie (Eds.),


Kievit, J., Kurkerink, M., & Marang-van de Mheen, P. J. (2010). Surgical adverse outcome reporting as part of routine clinical care. *Qual Saf Health Care, 19*(6), 1-5.


culture on multiple outcomes of care for coronary artery bypass graft surgery patients. [ ]

*Med Care, 38*(2), 207-217.


