Quality Improvement Science in the PICU

9

Matthew F. Niedner

Abstract

This chapter describes important aspects of health-care quality, what it is, and how to engage clinical providers in improvement efforts. The case for improving healthcare quality is outlined, with an emphasis on the special considerations within the pediatric intensive care unit setting. Common terminology to facilitate an understanding of quality is reviewed – including near misses, preventable and non-preventable adverse events. Models for understanding system and process performance are discussed – including the system of profound knowledge, aims oriented change, Plan-Do-Study-Act applications of the scientific method, and features of high reliability organizations. Provided is a basic tool kit for implementing change through the use of error proofing, process control, lean engineering, standardization, checklists, and protocols. Additionally, an introductory review of statistical process control and understanding variation through control charts is provided.

Keywords

Quality Improvement • Patient Safety • Medical Errors • Reliability • Translational Research

What Is Health-Care Quality?

In *To Err Is Human*, the Institute of Medicine (IOM) defined *health-care quality* as "the degree to which health services increase the likelihood of desired outcomes" and *patient safety* as "freedom from accidental injury because of medical care or medical errors" [1]. These two concepts are fundamentally linked, of course, and many notable voices have explicitly cited safety as the key dimension of quality, including the IOM, the Leapfrog Group, the Institute for Healthcare Improvement (IHI), the National Quality Forum (NQF), and even Hippocrates – *primum non nocere*. Other chapters of this textbook will focus more explicitly on patient safety, unintended harm, and the prevention of specific health-care associated complications. This chapter focuses on improvement models, general strategies, and practical concepts that can be applied in efforts to improve health-care quality.

The IOM outlined six dimensions of health-care quality [2]. The most fundamental attribute is that care should be safe. Care must also be effective and appropriately dispensed—that is, care must be provided to all who could benefit and not to those unlikely to benefit (avoiding underuse and overuse). Care should be patient-centered—respectful of and responsive to individual patient preferences, needs, and values, as opposed to provider or organizational motives. Quality care is timely—the right care to the right person at the right time—with waits and delays eliminated or minimized. Care should be efficient, actively seeking to identify and eliminate all forms of waste, be it time, equipment, supplies, or energy. The final dimension of care outlined by the IOM is equitability—the provision of care that does not vary in quality because of personal characteristics such as sex, ethnicity, geographic location, and socioeconomic status. It is a tall order, but ultimately attainable.

M.F. Niedner, MD
Pediatric Intensive Care Unit, Division of Critical Care Medicine, Department of Pediatrics, University of Michigan Medical Center, Mott Children's Hospital, F-6894 Mott #0243,
1500 East Medical Center Dr, Ann Arbor, MI 48109-0243, USA
e-mail: mniedner@med.umich.edu

D.S. Wheeler et al. (eds.), *Pediatric Critical Care Medicine*,
DOI 10.1007/978-1-4471-6362-6_9,

The terms quality assurance (QA), quality control (QC), and quality improvement (QI) are often used interchangeably in colloquial parlance, however, each has a slightly different meaning. The primary objective of QA is to demonstrate that a service or product meets a set of predefined expectations or requirements. This is achieved by comparing actual processes and/or outcomes to those specified criteria. QC involves the systematic use of performance monitoring and corrective methods to ensure that a service or product conforms to a desired standard. QI refers to the betterment or enhancement of a product or service compared to current, historical, or benchmark states. The term continuous quality improvement (CQI) is used to describe ongoing or iterative QI efforts.

Engaging Healthcare Providers in Quality Improvement

In spite of the known risks to patients from deficits in healthcare quality and safety countermeasures, most clinicians do not deliberately employ QI principles in their work, often believing that the responsibility for "systems issues" resides with the hospital administration [3–5]. Most clinicians focus on an understanding of pathophysiology and extrapolate treatment focusing on these principles. Figure 9.1 shows the translational research sequence of studies that transpire going from bench to bedside to practice guidelines. Research guided primarily by physiologic principles occurs in the early stages of translational research (e.g., T1 and T2). Late translational research (T3), such as implementation science, performance reliability, or improvement sustainability, is often confounded by local phenomena (e.g., staffing ratios, case mix), individualism (e.g., clinicians' experience base, leadership style), human factors (e.g., psychology, ergonomics), and nonmedical disciplines (e.g., business, economics, information technology, industrial engineering). These factors impact the generalizability of many quality and safety interventions. Tragically, the vast majority of discovery from T3 translational research is never published or disseminated—a lost opportunity in terms of the knowledge that *can* be extrapolated [3].

In 2001, the IOM report, *Crossing the Quality Chasm*, elaborated on strategic solutions to improve safety and quality in health-care, emphasizing the importance of systems-based analysis, correction of latent defects in complex systems, transparency, multi-professional and multi-institutional collaboration [2]. Perhaps most importantly was the emphasis on overcoming the culture of individual blame and reliance on imperfect humans to simply try harder or do better. The IOM argued that imperfections in human performance should be expected, and that systems should be designed to anticipate and mitigate their impact on patient outcomes. With the sentinel IOM reports, important healthcare stakeholders have made significant commitments to address these issues [6]. In 1999, the Accreditation Council for Graduate Medical Education began requiring competence in systems-based practice for physicians in training. In 2005, the American Board of Medical Specialties outlined the kinds of patient-safety and quality-improvement content

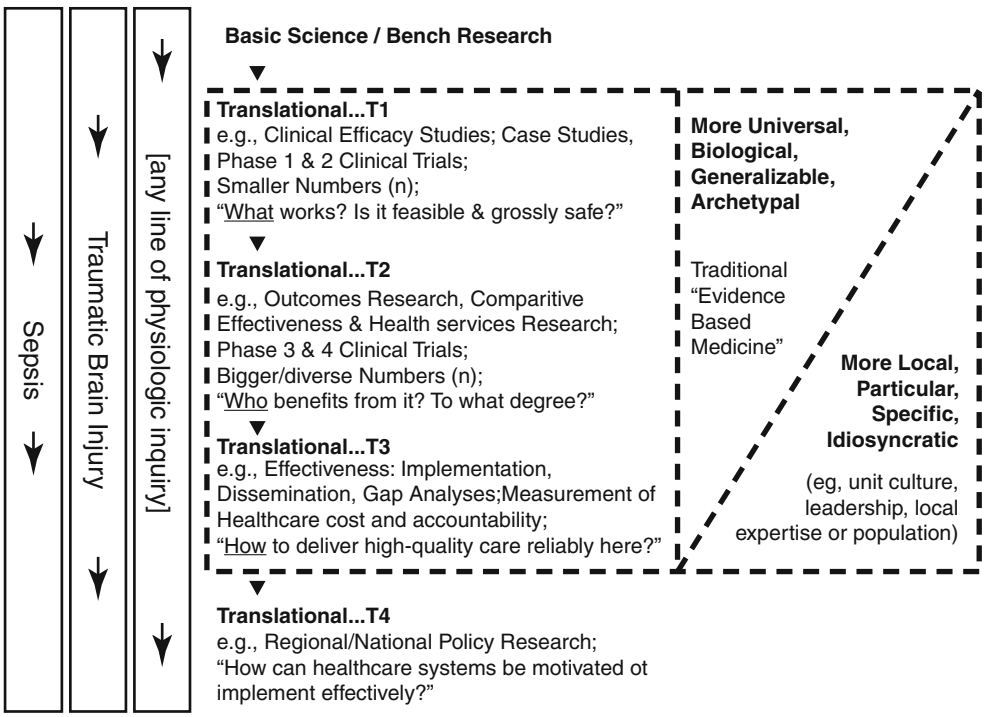

Fig. 9.1 A model of translational research stages

that should be included on board-examination questions in all medical specialties [7]. In 2007, on the heels of pay-for-performance approaches pioneered by some third-party payers, the Centers for Medicare & Medicaid Services declared that there would no longer be reimbursement for the additional costs associated with certain preventable medical complications – and many other insurers are now following suit [8, 9]. Most recently, in 2010, the Maintenance of Certification (MOC) program for physicians went beyond the traditional requirements—such as periodic examinations, continuing education, state credentialing—and required for the first time a practice performance assessment [10]. To maintain certification, physicians out of training must now demonstrate their ability to assess the quality of care they provide compared with peers or national benchmarks – and be able to apply the best evidence through follow-up assessments [11]. This latest element of MOC has only begun to play out, but it clearly represents a significant intent to encourage physicians to engage in clinical QI more substantively.

In the context of increasing public pressure for transparency, financial incentives for performance, as well as legal and regulatory drivers for improved patient safety and health-care quality, there is a rapidly flourishing academic dimension to QI. Some of the methodologies to analyze quality and safety in health-care with academic rigor are still in development, relatively young, and under adaptation from other industries and disciplines—and many are not familiar to practicing clinicians nor embedded in medical education [12]. Such analytic tools from human-factors engineering, psychology, industrial engineering and manufacturing are increasingly finding their way into the traditional stomping grounds of peer-reviewed medical literature [13]. Finally, calls for scholarly accounts of quality and safety endeavors—along with publication guidelines for proper peer review—have appeared in recent years [14, 15].

The Case for Health-Care Quality as a Priority for Pediatric Critical Care

The Pediatric Intensive Care Unit (PICU) is a place of converging threats to quality and safety—in essence, a canary in the coal mine. All the challenges that adult medicine has in defining, measuring, and improving health-care quality are amplified in pediatrics, where evidence for best practice is more limited, the data infrastructure less robust, and the potential loss of quality-adjusted life-years greater [16]. Furthermore, ICUs, emergency departments, and operating rooms are the locations where defects in care delivery are most prolific [1].

In the PICU, workflow is unpredictable, diverse, complex, technical, stressful, and invasive—often with little margin for error, as critically ill patients may succumb to insults that healthier patients might tolerate. Care must be multidisciplinary, with no one role—physician, nurse, respiratory therapist, pharmacist, or social worker—able to have mastery of all the information and skills necessary to treat the patient optimally. This makes the ICU critically dependent on teamwork and communication that must both function reliably in short time frames—and some of the most common and refractory sources of error are teamwork and communication failures [17]. There is a very high volume of "sharp end" activities—the point of care where all propagated defects emerging from the system actually reach the patient—a medicine injected, a catheter inserted, a chest shocked. Yet in spite of the pervasive risks in the PICU, adverse outcomes from errors are camouflaged. In general, more deaths occur in PICUs than in any other care units within a pediatric hospital, and adverse outcomes, although unwanted, are not unexpected. A death in a general-care unit stands out in bold relief and is scrutinized deeply for what went wrong—but how is it clearly discerned when morbidity or mortality in the PICU is from progressive refractory critical illness or from unrecognized (or unreported) medical errors slipping under the radar of awareness?

Pediatric critical-care medicine as a discipline has taken the initiative in asserting a voice at the national level regarding measures of performance quality and safety, instead of waiting for external entities to select or mandate such measures as targets for public reporting and pay-for-performance [18]. The density of error potential, the high level of vigilance, and the abundance of objective patient data make the ICU a locale that lends itself to measuring quality. Indeed, the NQF recently endorsed 48 quality measures, of which 23 are relevant to PICU care, and 7 are explicit measures of PICU performance. There are only 13 pediatric-specific measures, making PICU-specific measure representative of more than half of the NQF-endorsed measures for pediatric care (Fig. 9.2) [19]. Similarly, Medicare's recent announcement of numerous nonreimbursed "never events" included 8 metrics highly relevant to the PICU, although none that are PICU-specific (Fig. 9.2) [9].

The health-care industry has a fairly woeful track record of reliably delivering contemporary best practice. It is estimated that adults typically receive recommended, evidence-based care about 55 % of the time, with little variation among acute, chronic, and preventive care [20]. Although PICU-specific data on best-practice compliance are scarce, general pediatric data from a similar analysis suggest performance that is comparable to adult care on average, but a larger variability based on type of care [21]. Children receive an estimated 68 % of indicated care for acute medical problems, 53 % for chronic medical conditions, and 41 % for preventive care. What is more, data derived from such audits do not include many errors unrelated to widely accepted best

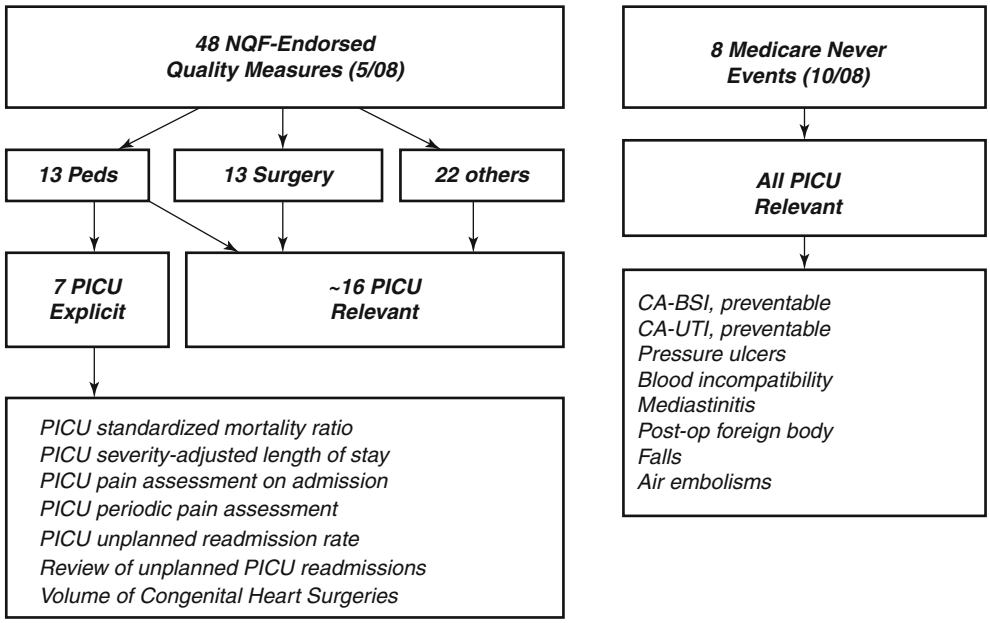

Fig. 9.2 Quality & safety measures endorsed by national entities. *CABSI* catheter-associated bloodstream infection, *CAUTI* catheter-associated urinary tract infection, *NQF* National Quality Forum

practices, nor those invisible to the audit methodology. In a survey of pediatric physicians and nurses, half filed incident reports on <50 % of their own errors, and a third did so <20 % of the time [22]. It is reasonable to conclude that most practitioners in the PICU are only aware of the tip of the iceberg when it comes to near-misses, preventable harm, and opportunities for improving healthcare quality.

Many quality and safety initiatives developed in the ICU find wide application in other medical environments and service lines. This not only makes the PICU fertile ground for the development of interventions relevant to pediatrics at large but also makes the PICU well positioned to further develop QI science itself [23]. If the PICU is to be the canary in the coal mine, then it can also be an incubator of improvement innovation.

Models for Understanding Quality

As experimental statistician George Box observed, "All models are wrong, but some are useful" [24]. Models, of course, provide an artificial structure for knowledge that reflects complex phenomena accurately enough to better enable understanding—ideally well enough to enable meaningful interpretation and constructive action. Clearly, healthcare quality is complex, but models can help us grasp what is essential. A few of the more common and useful models for improvement are touched on here.

A high-level perspective on managing quality is encapsulated by Deming's system of profound knowledge [25]. This model is comprised of these four domains:

1. *Appreciation of a system*: understanding the overall processes involving suppliers, producers, and customers (or recipients) of goods and services; insight into the interdependence and dynamism of complex institutions; recognition of how the whole is greater than the sum of parts.
2. *Knowledge of variation*: understanding the range and causes of variation in quality (e.g., common versus special cause); ability or access to statistical sampling in measurements.
3. *Theory of knowledge* (*or epistemology*): understanding the merits and limitations of what is knowable, how we come to understand through experimentation, and the roles of modeling, prediction, and justification.
4. *Knowledge of psychology*: understanding concepts of human nature, including inter-individual variation, communication styles, beliefs, assumptions, intrinsic/extrinsic motivation, and the will to change; insight into the collective impact of individual psychology toward the behavior of groups, morale, and teamwork.

Perhaps one of the simplest ways to think of QI is as a strategy to close the gap between actual practice and best known practice. The estimated time lag for scientific knowledge generated in randomized clinical trials to be routinely accepted into medical practice is 17 years, a rather shocking testimony to the size and persistence of the gap, and a provocative invitation to close it [2]. Figure 9.3 is one depiction of how QI fits in the context of actual, best, and idealized performance. If the graph is taken as a survival or time-to-adverse-event curve from some identified measure of quality or safety, one can assume that the ideal outcome is 100 % perfect over time. Much conventional research is focused on closing the gap between current best practice and such an idealized practice—that is, taking the best known mousetrap and incrementally making it better. In contrast, much QI is focused on closing the gap between actual practice and best

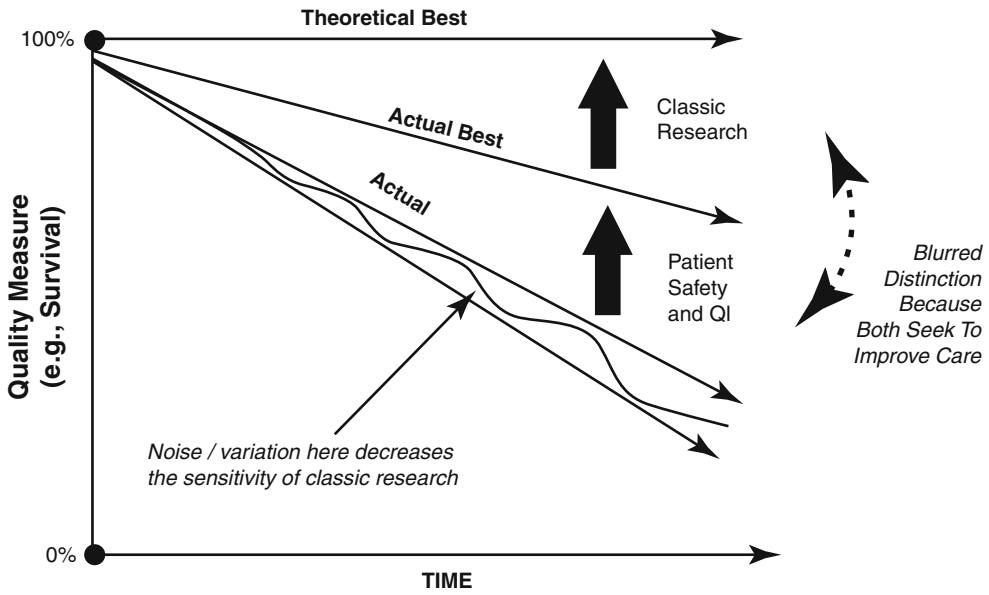

Fig. 9.3 A comparison of classic research versus quality improvement. Classic prospective interventional research and quality improvement (QI) both seek to move clinical care closer to an idealized theoretical best outcome. Classic research often focuses on novel means – such as more sensitive diagnostics or superior therapeutics. QI often focuses on closing the gap between common practice and established best practice—such as thru standardization, decision-support, and automation. Reducing errors and performance variation in operations through quality control can reduce the noise introduced into measures of research interest, thereby improving the sensitivity of research efforts. Thus, clinical research and clinical QI are synergistic

known practice—that is, taking the best designed mousetrap already known and ensuring that it deploys flawlessly every time it is indicated (and not when it isn't) in the specific local context of deployment. Complex systems and multi-step successes are often a basis for the gap between actual and best practice. If one is to start with valid, evidence-based health-care guidelines, many steps must be executed correctly for the best care to reach a patient. Staff must be aware of the evidence, accept it sufficiently, know how to apply it, work in a care environment that makes it feasible, remember to do it real-time, have agreement from the care team and consumer, and ensure that the care is actually performed in the right time, place, and manner. If each of these steps were to occur with 95 % fidelity, best care would be delivered only about 70 % of the time.

It is not uncommon for classic research and QI interventions to be confused, as they both are seeking to improve patient outcomes in data-driven ways, although typically through different mechanisms. Classic research and QI are often touted as serial, but in truth they are concurrent and synergistic. Classic research in early translational stages (T1/T2) certainly defines many of the best known practices that QI agents strive to implement in later translational work. Yet improvements in the consistency and reliability of baseline clinical performance can reduce noise, improving the sensitivity of analyses aimed at detecting small incremental improvements in best practice (Fig. 9.3). For example, if children are not dying or having prolonged hospitalizations due to nosocomial infections, medication errors, and surgical complications, then sensitivity will be increased when trying to detect the impact of a novel intervention on survivorship and length of stay.

Hierarchy of Defects	Methods to Remediate
System Vulnerabilities *(Risk of Failure)*	*[MORE PROACTIVE]*
	Failure Mode & Effects Analysis
Near Misses *(Failure, No Harm)*	Reduce Steps (simplify / "lean")
	Reliable Steps (process control)
	Root Cause Analysis (RCA)
Adverse Events *(Harm)*	*[MORE REACTIVE]*

Fig. 9.4 Hierarchy of defects and methods to remediate

Another useful QI model relates to the hierarchy of defects in a complex system. *Latent system risks*, *near-misses*, and *actual harm* are points along a continuum. Figure 9.4 demonstrates how this continuum matches up to methods commonly employed to remediate such defects. There is a long-standing debate between whether it is more advantageous to measure risk, errors, or harm—but in truth, each has advantages and disadvantages, and all are widely used. A brief consideration of each measurement follows.

Table 9.1 Types of waste in the lean model

Type of waste	Definition	Heath care examples
Correction	Rework because of defects, poor quality, mistakes	Revising incomplete or illegible forms Order entry error
Overproduction	Producing the wrong things or producing more/sooner/faster than required by downstream processes	Unused or too-frequent laboratory testing Too frequent clinic appointments
Motion/ movement	Unnecessary physical activity (motion) by people or relocation (movement) of people/materials	Walking to office supplies or exam room Searching for misplaced equipment or chart Multiple patient room transfers
Waiting	People, machine, and information idle time	Patient waiting in waiting room Providers waiting for lab results
Inventory	Information, material, or consumers in queue or stock	Stacks of medical notes to be dictated Excess stored supplies in stock room
Processing	Unnecessary or redundant handling or processing	Reentry of patient demographics Repeat collection of data
Underutilization	Tasking staff below their capacity or abilities	Nurse tending phones to refill prescriptions Surgeon operating on one patient per day

Detection and elimination of latent defects in a complex system provides the ideal solution to improving quality and safety, as it is the furthest point upstream from harming a patient. Failure Modes and Effects Analysis (FMEA) is one powerful strategy to identify ways in which a complex system can fail on the basis of known historical performance of constituent parts of a device or process. This allows potential defects to be designed out of the system (or planned countermeasures to be devised) before a design actually culminates in an actual product or active process. FMEA is widely used in manufacturing and engineering industries where device performance is fairly predictable (as with an intravenous pump or telemetry unit), but it is increasingly applied in the service industry—even though human factors and dynamic phenomena are more difficult to model [17]. Limitations of analyzing latent system vulnerabilities include: lack of good historical performance data upon which to base the model, the risk of unforeseen perturbations in complex and interdependent systems, unpredictable and dynamic changes in the system, lack of intuitive guidance to the sources of risk, and the theoretical nature of some assumptions and conclusions in the absence of measurable errors or harm.

Another top-tier tool in system improvement is process streamlining through the elimination of waste—be it time, energy, materials, or process complexity. As Albert Einstein said, "Make everything as simple as possible, but not simpler" [26]. If a desired outcome can be achieved in fewer steps without loss of fidelity or performance, it will likely be safer, because eliminating unnecessary steps removes some opportunities for errors to creep into a system and simultaneously reduces the number of variables when trying to understand ongoing failures. Furthermore, elimination of waste improves value from a cost–benefit perspective. This kind of streamlining to optimize value-added output is the basis for lean design, originally applied to production lines but increasingly applied to service lines. The contemporary paradigm of lean production and management is based on the Toyota Production System, and lean strategies have been successfully adapted to health care [27–29]. Specific examples of the kinds of waste identified and eliminated in the health-care setting are outlined in Table 9.1. A full description of lean methodology is beyond the scope of this chapter, but exhaustive resources are available for the interested student [30, 31].

Progressing along the ladder from latent defect to harm, the next step closer to patients is error. The most widely accepted criteria for medical error is failure of a planned action to be completed as intended (an error of execution) or the use of a wrong plan to achieve an aim (an error of planning), whether by commission or omission [32]. Thus the drug overdose due to a decimal error may be considered an error of execution by commission, whereas the treatment of mistaken septic shock instead of the actual adrenal crisis might be viewed as an error of planning where proper care was omitted. If the error results in no harm, it is commonly labeled a *near-miss*, whereas if harm occurs, it is considered *a preventable adverse event*. This should be considered distinct from *non-preventable adverse events* for which ways to avoid the known complication are not established—that is, harm occurring as a consequence of medical care but in the absence of an error (such as with the risk of cardiotoxicity from chemotherapy). Because errors resulting in near-misses are far more common than errors resulting in preventable harm, near-misses provide an attractive target for monitoring and measuring quality and safety on a continuous basis. Analysis of near-misses in an iterative manner can help generate hypotheses for root causes more rapidly than if only harmful events are considered. Focusing on errors and can be particularly helpful when related outcome measures are too rare or catastrophic to be acceptable guideposts (e.g., deaths).

Such error-based surveillance (e.g., compliance with a best practice) is particularly helpful when there is good evidence for key steps or processes firmly established in the medical literature. The ability to identify and monitor compliance with important process measures provides actionable data to an improvement team about how to reduce unnecessary variation and close the gap between actual performance and desired best practice (Fig. 9.3). This is the basis of process control—often associated with the Six Sigma management strategy employed widely in many sectors of industry, including health care.

Although there are advantages to error-based quality-performance analyses, there are notable limitations to acknowledge. Error and near-miss rates vary widely depending on the definitions used for error, surveillance methods, and even the safety culture of the reporting unit [33, 34]. Furthermore, measures of errors are most helpful if they can be expressed as rates (errors divided by opportunities for that error type), but often there is no denominator available [18]. Even the numerators can be circumspect because many errors go unreported, and there is an attention bias that favors identification of errors of commission (rather than omission) and errors resulting in harm (rather than near-misses) [22]. Thus, error-based quality assessment may be better applied as a local qualitative and semi-quantitative improvement strategy, rather than an as comparative performance tool [35.]

The measure of actual harm to patients is a final measure of quality and safety in the current hierarchy being discussed (Fig. 9.4). From a high-principled perspective, one can consider harm to the patient a failure to have detected and mitigated the latent system defects and combination of conspiring errors. Although risk and near-miss analyses are more proactive, harm analysis is a more reactive process—there is no putting the genie back in the bottle. From a pragmatic perspective, it is like adding insult to injury to witness harm and not try to learn from it. It is worth noting that all errors are not created equal . . . those resulting in harm may be distinct from those that do not, and harmful errors can implicate defects not necessarily apparent in near-misses [36]. Several organizations have proposed injury-based trigger tools that can be used to provide systematic surveillance measures of harm, such as the Agency for Healthcare Research and Quality's general and pediatric-specific Patient Safety Indicators, the IHI's Global Trigger Tool, and others [18, 37–39]. Some of these metrics and tools focus on types of harm that are assumed to be preventable or largely so, whereas other tools for measuring harm are inclusive of all readily identifiable harm. One key advantage to measuring all harm is that it provides an opportunity to question the boundary between preventable and unpreventable injuries. If the goal in health care is to eliminate or reduce all harm to patients, then including measures of harm considered unpreventable by traditional medical standards can direct our attention toward innovative care or research. Indeed, it is our medical legacy that types of harm formerly deemed to be "a cost of doing business in the ICU" are now considered imminently preventable, as has been the case with many hospital-acquired infections [23, 40, 41].

Harm-based performance metrics have their limitations too, of course—the most obvious being that the patient is injured in some manner. Because unpreventable adverse events and deaths are, to some extent, expected in the PICU, such occurrences do not necessarily raise the specter of preventable error. When they are recognized, attribution may not be accurate. Furthermore, the retrospective nature of harm evaluation provokes numerous kinds of bias toward which human perceptions are prone—such as hindsight and outcome bias [42].

If the analysis and attribution of risk, error, or harm is significantly biased, incorrect, or overly simplistic, then the conclusions not only are invalid but also can lead to unnecessary and possibly counterproductive attempts at remediation. Therefore, all risk, error, and harm analysis—as well as planned responses—must be undertaken with such limits and pitfalls in mind. Finally, because all monitoring and corrective strategies have limitations and none are perfectly suited for all applications, it makes sense to employ multiple simultaneous approaches for a more robust quality-monitoring and safety-monitoring system. Doing so also helps create cross-validation between sources of perceived risks, error patterns, and actual harm—helping to overcome the weaknesses of each individual approach.

The Science of Quality Improvement

When considering QI science, especially late translational research (e.g., T3), an understanding of the "realistic evaluation" model put forward by Pawson and Tilley [43] is helpful to get past some of the constraints inherent in orthodox experimentation [44] (Fig. 9.5). In a nutshell, realistic evaluation seeks to explain variation in outcomes by analyzing the context that may have differentially enabled or disabled an intervention from having the postulated impact. Classic research methods use strategies to wash out variation and isolate the effect of an intervention—such as with randomization, prospective analysis, large sample sizes, blinding, and controlling for known confounders. In the face of conflicted literature on the efficacy of an intervention, a conventional approach might be to perform a meta-analysis—to essentially "lump" the studies together to see what the "true" effect is when analyzed with greater power. However compellingly large the sample size may become, there are many limitations to this method of understanding variation between studies [45–47]. Another way to interpret conflicting studies is to consider the unknown, unmeasured, or unrecognized

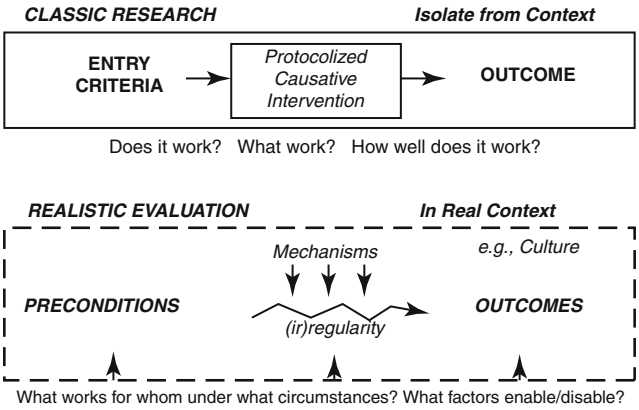

Fig. 9.5 Classic research versus realistic evaluation. Classic prospective interventional research often employs methods to "wash out" the influence of context and bias—such as through strict entry/exclusion criteria, large sample sizes, randomization, blinding, cluster analysis, etc. This largely obviates the need to explicitly understand, measure, or control for contextual variables that may be influencing outcomes—allowing the effects of the intervention to be isolated. This is an effective way to determine whether or not an intervention works and to what degree. In realistic evaluation, greater consideration is given to how uncontrolled preconditions are acted upon by a variety of mechanisms and/or interventions (which may not be carried out consistently) in a context of local idiosyncrasies (such as teamwork climate, staff experience, human factors). In a realistic evaluation of an intended intervention, differences it outcome might be interpreted based on what made the intervention perform as expected in some circumstances but not in others

contextual factors that altered the impact of an intervention (either positively or negatively) in different studies or sites. If influential contextual factors can be identified, implementing such enablers or eliminating disablers may allow an intervention to perform as intended. A good example of this is the impact of safety culture on certain unit outcomes, such as nosocomial infection rates, throughput delays, medication errors, and staff turnover [48–52]. Teamwork and safety climate have proven to be responsive to ICU-wide interventions, both positively and negatively, making interventional anthropology a growing domain of QI [49, 53].

Not only are the models for QI science often different from traditional biostatistical approaches, but the way measures are viewed and used also differ (Table 9.2) [54, 55]. Where classic research seeks to discover new knowledge in the scientific realm, improvement science seeks to operationalize it in real life. In research orthodoxy, unequivocal hard outcomes (e.g., death) are preferable to surrogate markers assumed to be correlated to them (e.g., multiple organ dysfunction scores). However, measures in improvement science are multidimensional, often with a hard outcome measure as an ultimate verification that improvement is occurring, but critically important process measures that serve as the tools to guide the specific improvement strategies employed. An important illustration of the difference between QI process measures and traditional surrogate measures can be made with hand hygiene compliance. While the hard outcome measure of hospital-acquired infections (HAIs) can be measured, they are hopefully infrequent events and do not lend themselves to rapidly determining efficacy of interventions to reduce them. Hand hygiene compliance, however, can be measured frequently, and rapid-cycle improvement interventions can be built around such key processes contributing to HAIs without waiting for adverse events to occur.

Where classic research strives to have a rigorously consistent management protocol, improvement science constantly tweaks and refines management toward best practice. Whereas classic research seeks to eliminate or minimize biases, improvement scientists try to hold them sufficiently steady during testing to allow for causal inference. Classic research is typically powered a priori to definitively answer a primary question with statistical significance once all the data are gathered, and interim analyses are shunned to avoid spurious signals. Conversely, improvement science seeks to generate real-time and continuous data that can be interpreted and acted on simultaneously—sometimes using data trends to inform confidence in recent interventions and guide next steps from a probabilistic vantage point. For instance, if a QI intervention were inexpensive, safe, and minimally burdensome (e.g., a central line insertion checklist), then one might accept a different level of confidence in the statistical significance of local implementation than, say, if one were considering an intervention that is expensive, cumbersome, and accompanied by potential unintended consequences (e.g., a fast MRI scanner).

A large historical body of QI efforts is blighted in the eyes of many scientists as mere administrative window dressing—the rewriting of policies, the aesthetic revision of a patient portal Web site, the feel-good of patient-centered or family-centered niceties [56]. However, the modern patient-safety movement represents a sustained effort to use rigorous methods driven by data and testing. As the axiom goes, "In God we trust; all others bring data" [57]. At the core of science, be it classic research or QI, is the method. There are many ways to apply the scientific method, and the crucial challenge is to have sufficient knowledge of the tools and methods, to grasp their limitations, and to know which tool is right to apply to a particular problem at hand [5, 12, 55]. To this end, pioneers in the science of improvement have put forward an archetypal model for improvement that can serve as a fairly universal platform well suited for improvement work in health care [58, 59].

The model for improvement begins with a clear expression of aims in measurable and time-specific terms as it applies to a defined population. As the IHI's motto goes, "Soon is not a time . . . some is not a number." The model for improvement calls for clearly defining or developing new

Table 9.2 Comparison of measures in classic research versus quality improvement science

	Classic research	Improvement/safety science
Usual goals	Discovery of new knowledge; providing objective proof or basis; establishing best practice	Operationalize discoveries or best practice into routine care; ensure/monitor performance
Intervention or protocol	Single static protocol; first and last patient in protocol get same management; long timetables	Flexible/dynamic protocol or multiple serial tests; management adjusted freely based on learning; short and responsive timetables
Management of confounders	Identify, eliminate, exclude, and control for biases thru blinding, randomization, cross-over, etc	Identify and understand biases; stabilize biases during tests or interpret findings in bias context
Preferred measures	Hard and unequivocal outcomes; background data to ensure comparability	Blend of outcomes measures, relevant process measures, and possible balancing measures
Power and scale	Powered to definitively answer question and possibly explore post hoc analyses	Minimally sufficient data to meet confidence threshold for action or decision; successful tests scale up
Data interpretation	Data blinding; no interim peeking; data safety monitoring boards; classic biostatistics with significance thresholds	Real-time data; analyze & act on data simultaneously; statistical process control (control charts); data trends influence next steps

Based on data from Refs. [55, 57]

metrics to track progress toward the aim in quantitative ways—some combination of outcome, process, and balancing measures to know whether improvement is actually taking place independent of qualitative opinion. The crucial judgment step in this model is selecting the change to test. All improvement requires making changes, but not all changes result in improvement. Reliance on individuals with keen insight into the system complexities at hand as well as the operational realities is most helpful at this decision point, as an improvement team ideally selects from a host of possible changes the one(s) *most likely* to result in desirable change.

Once a candidate change is identified, the testing process can begin within the Plan → Do → Study → Act (PDSA) cycles. The PDSA structure is just a bare-bones expression of the scientific method as it is applied in the real work setting. PDSA is a decades-old construct offered by William Edwards Deming, PhD (Fig. 9.6), and is very analogous to the Six Sigma terminology for the scientific method as applied to process control—namely: Define, Measure, Analyze, Improve, Control [56]. The PDSA approach is not intended as a one-off pilot study, but instead a repetitive, rapid-cycle, action-oriented learning approach without necessarily waiting for complete stabilization of effect before another PDSA is undertaken. PDSA cycles that fail to meet hypotheses should be scrutinized for reasons, whereas PDSA cycles that suggest or result in clear improvement should be considered for refinement and scaling up. Therefore, all PDSA cycles should generate learning. In fact, failed PDSA cycles, if they are adequately analyzed, often teach improvement teams more about a system than successful PDSA cycles.

When interpreting measures for QI, one must understand the nuances to outcome, process, and balancing measures. Outcome measures indicate whether changes are actually leading to improvement and directly speak to the aim (e.g., central-line infection rates). Process measures relate to key activities or steps that are believed to drive the outcome measure (e.g., compliance with central-line insertion best practice). Balancing measures are sometimes selected to monitor whether changes intended to improve one part of a system are causing new problems in other parts of the system (e.g., procedure cart equipment costs). The key distinction is between "looking good" versus "doing well." Figure 9.7 depicts a model relating outcome measures and process measures. Predicated on the historical pressures of external audits to measure compliance with mandated processes (e.g., Joint Commission mandates), one might be tempted to believe that high compliance with process measures are preferable.

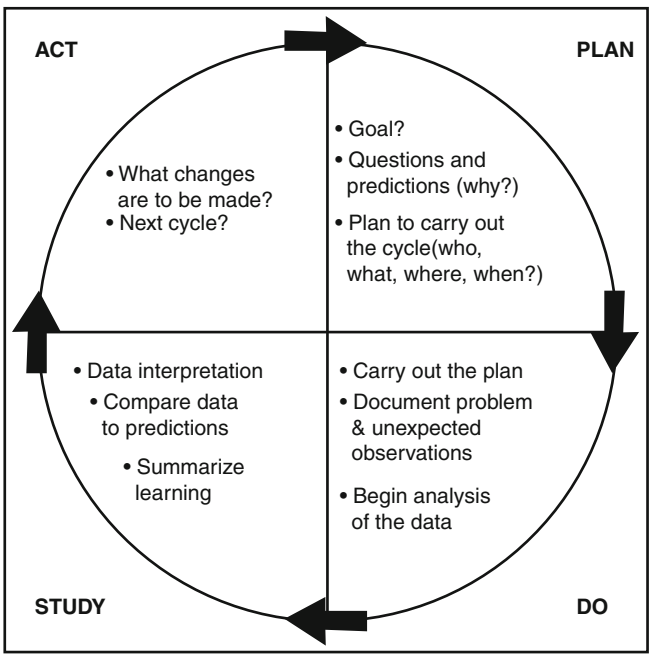

Fig. 9.6 The "Plan → Do → Study → Act" Cycle

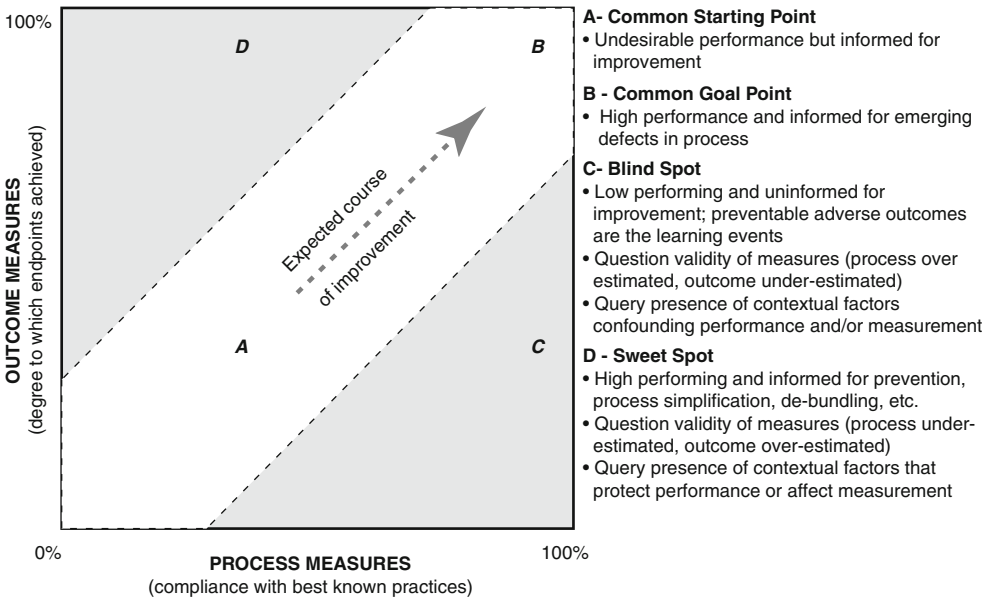

Fig. 9.7 Process measures versus outcome measures

However, such a state can cripple an improvement team's ability to target interventions at defects in the real drivers of an outcome. If the outcome measure is perfect (e.g., zero central-line infections), then high compliance in a process measure may, in fact, reflect the etiology of the high performance (point B of Fig. 9.7). However, if the outcome measure is less than perfect, but the process measures (e.g., compliance with central-line insertion) show ideal practice, then either the wrong process measures have been selected or the measurement process is fallible – and the improvement team is blinded as to what to do next (point C of Fig. 9.7). When the process measures are less than perfect and the outcome measure is poor, the improvement team has insight into how to improve (point A of Fig. 9.7). Similarly, if the process measures are imperfect but performance is good for the outcome measure, an improvement team may be well poised to make improvements before additional adverse events occur (point D of Fig. 9.7). The outcome measure is the proof, whereas the process measures serves as informants that help an improvement team close the gap between actual and best practices. However, either can be miscalibrated, depending on surveillance techniques, staff members' perceptions of psychological safety, local response to error, safety culture, and so forth. In essence, an improvement team wants process measures that are "graded" as critically as possible to reveal all possible areas for improvement. Nonpunitive response to error (e.g., noncompliance) can help generate the honesty, accuracy, and transparency necessary to improve outcome measures.

One additional model worth discussing relates to high-reliability organizations (HRO). In industries with particularly complex and high-risk characteristics, some organizations have managed to contain errors and harm with remarkable effectiveness [60–62]. Examples include nuclear power plants, aircraft-carrier flight decks, and commercial aviation. Characteristic features of the context where development of an HRO is salient include: environments rich with the possibility for errors to occur, high-stakes error potential that can result in significant harm, high-pressure psychology with an unforgiving social or political milieu, opaque and delicate operations where learning through experimentation is difficult, and system complexity such that no one person can have mastery of the numerous and nuanced processes or technology involved. Although this may seem to describe health care in general (and pediatric critical care specifically), there are no known examples of health-care units achieving the kind of safety and reliability performance as HROs in other industries—but that does not mean that the health-care industry cannot learn from these examples. What, then, are the characteristic features of an HRO? They were outlined succinctly by Weick and Sutcliffe as follows [60]:

1. *Preoccupation with failure*: Small, seemingly inconsequential errors are regarded as a symptom that something serious is wrong. There is a commitment to finding and analyzing the "half-events" and to treating all failures as learning opportunities in a nonpunitive and transparent manner.
2. *Sensitivity to operations*: Attention is given to what is happening on the front line by all levels in the organization. Effective teamwork and a culture of safety are considered crucial. Data on processes and system performance are built into daily work and emerge visibly to promote situational awareness.
3. *Reluctance to simplify*: Excusing or explaining away errors is avoided. Diversity in experience, perspective, and opinion are encouraged. The questioning of assumptions is respected and supported, being viewed as a form of loyal dissent.

4. *Commitment to resilience*: There is an assumption that errors will occur, and there is a commitment to detect, contain, and recover from errors that do happen. The concept of fail-safe is not that a person or system will not fail but that something will inevitably fail and everyone will still be kept safe regardless.
5. *Deference to expertise*: Decision making and problem solving deeply engages people with the most related knowledge and expertise—typically the frontline workers, as opposed to top-down leadership.

Understanding Variation Through Statistical Process Control

As mentioned in Deming's system of profound knowledge, it is crucial for QI teams to understand variation in data to be able to avoid making errors in interpretation. However many QI teams do not, practically speaking, have sufficient biostatistical training to have confidence in real-time interpretation of continuous data streams generated by their efforts. This can stymie aims-oriented, data-driven, rapid-cycle PDSA improvement work. One accessible solution to this dilemma is for improvement teams to develop a working knowledge of statistical process control (SPC) [63–65]. The basic component of SPC is the control chart (also known as a Shewart chart) which serves as a graphical heuristic. It is not a hypothesis test but rather is constructed to generate insights into temporal signals in a complex system under a wide range of unknowable circumstances, both future and past.

To grasp the essence of a control chart, imagine a conventional bell-shaped curve turned 90°on its side to give a horizontal set of lines corresponding to the mean and standard deviations (Fig. 9.8). Time is represented on the x-axis, and the performance metric on the y-axis. Sigma is similar to standard deviation, but depends on the type of control chart being used and is generally more sensitive to detect outlying data as signals. The plus and minus three-sigma boundaries are called the upper and lower control limits, respectively. Accepting distributions of data within plus-or-minus three sigma of the mean (i.e., within a total range of six sigma) affords a rational way to minimize type I and type II errors. When data are outside this range, the system is considered *out of control*.

Unlike a bell-shaped curve where data are collapsed into one time bin, a control chart plots data over the additional axis of time (Fig. 9.8), providing insight into signals that emerge from the sequence of data being measured. The practical power of SPC is that people who are not statisticians can bring significant statistical rigor to their quantitative data in an intuitive format by understanding just a handful of simple rules to distinguish special-cause variation (i.e., signal) from common cause variation (i.e., noise) (Fig. 9.9) [20]. These rules are fairly intuitive for anyone with a basic grasp of probabilities. For instance, any single data point more than three sigma (roughly equivalent to three standard deviations) from the mean should stand out as a signal to which an attributable cause should be sought. So should a series of nine points on the same side of the mean line—tantamount to flipping nine consecutive heads with a coin.

SPC is rooted in venerated time series analysis and is an available function in most advanced biostatistical software packages. Yet such data can be readily maintained in simple spreadsheets, typically by entering numerators and denominators on whatever data-collection cycle is appropriate (e.g., weekly, monthly, quarterly). This allows QI teams to collect, interpret, and act on data in real time, without the bottleneck created by relying on a statistician to intermittently decipher

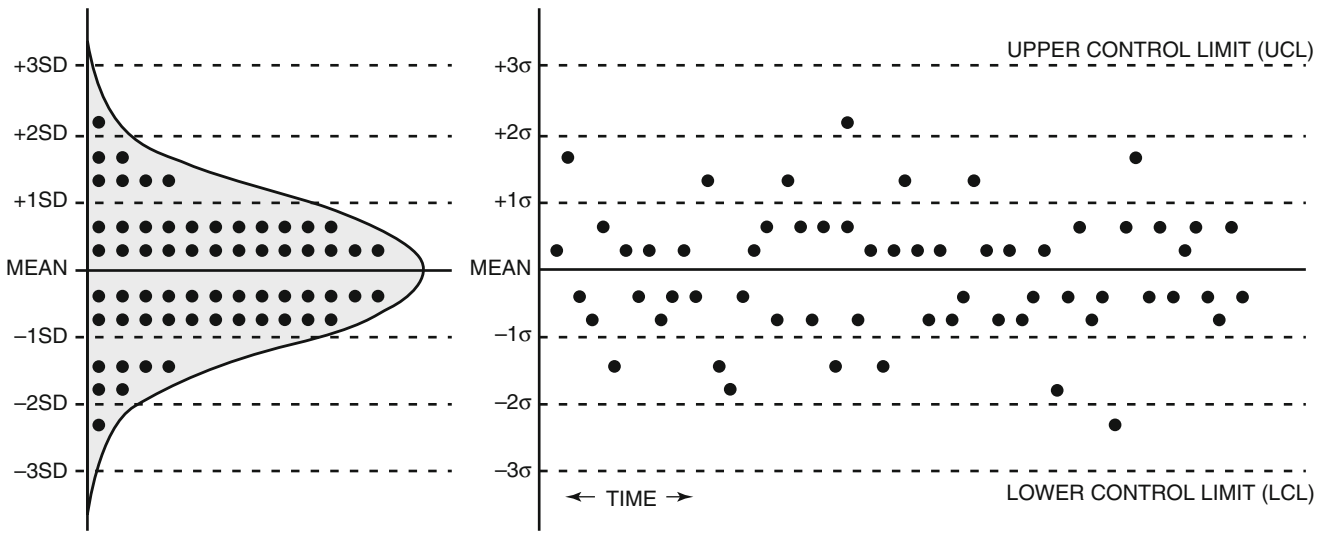

Fig. 9.8 Analogy of a bell-shaped distribution to control chart data

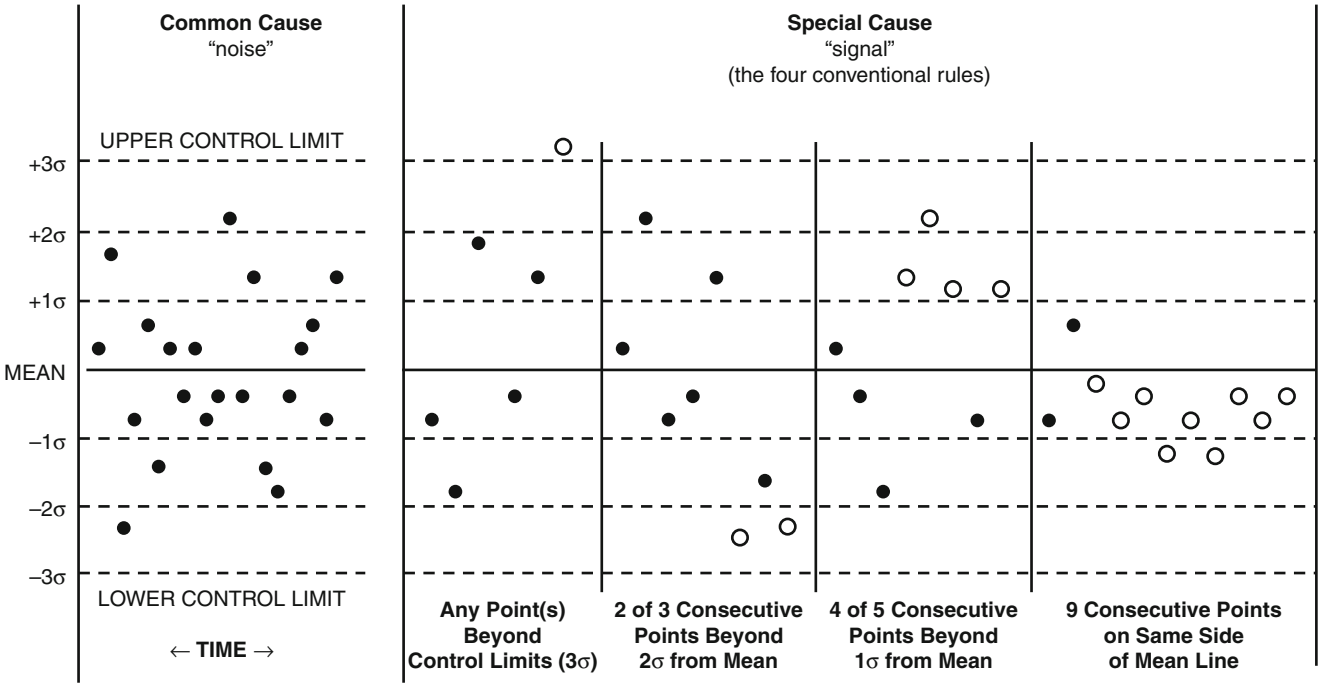

Fig. 9.9 Conventional rules for special cause variation in a control chart

signals and noise. Some expertise is required in selecting the correct type of control chart when getting started, because just as with traditional biostatistics, the nature of the measure (e.g., continuous, integer, categorical) determines the proper control chart to use—although one control chart visually appears much the same as the next one. Knowing when to reset mean lines also requires judgment, but generally speaking, break-points in the data are suggested by special cause variation in the data rather than by arbitrary cut points (e.g., fiscal year) or before-versus-after intervention timeframes (as is more common in classical time series analysis).

SPC can help "tell the story" of an improvement project, showing if a performance measure is in control, is exhibiting instability, is consistent over time, is showing improvement in response to interventions, or is showing decay in the context of neglect. SPC can help keep QI teams from mistaking noise for signal (i.e., false positive or type I error) and signal for noise (false negative or type II error). Figure 9.10 provides an example using catheter-associated bloodstream infection rates. Data point 'A' might seem high, but it does not meet the rules for special-cause variation; therefore reacting to this signal may represent energy wasted on random variation. Conversely, the mean line change at point 'B' represents a small but statistically significant change in the outcome measure that might have gone unrecognized without a control chart or periodic statistical testing.

Some traditionalists prefer *P* values or confidence intervals to describe variation; however, this often collapses time as a dimension (e.g., pre- versus post-intervention analyses) and reduces rich graphic information into a few simple numbers. Confidence intervals and *P* values can still be derived from the primary data using classic biostatistical methods for additional verification, but determining the equivalent *P* values of a control chart "misses the point," according to Deming, who asserted that the intent and purpose of continuous QI is distinct from other forms of experimentation. Among other things, the *P* value of a control chart would depend on gross variability, the number of points plotted, and the number of rules used for special-cause variation (more than the conventional 4 can be used). However, a very rough estimation of the statistical rigor of a typical control chart can be made. If a control chart with normally distributed data were interpreted with the four conventional rules, the chances of either missing a real signal or seeing a false signal in 30 plotted points would be roughly 1 in 50–200 (or a *P* value of 0.02–0.005).

A Few Key Tools and Change Concepts for Quality Improvement

Many organizational change concepts, improvement tools, and safety strategies can be applied in an ICU—and it can certainly be helpful to have a wide variety of approaches at one's disposal, because some quality and safety problems lend themselves better to particular types of approaches. Although a description of the many possible tools and concepts is beyond the scope of this chapter, some approaches

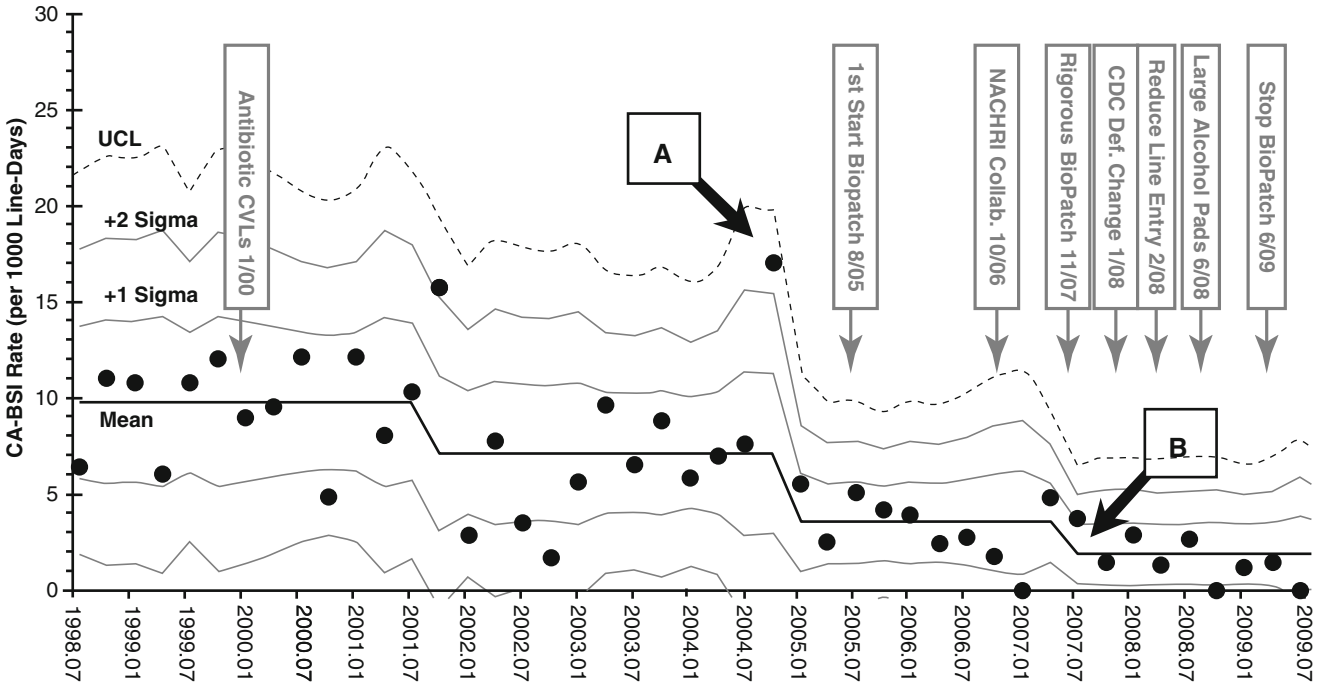

Fig. 9.10 Example of a control chart of catheter-associated bloodstream infections, annotated with key tests of change. *CABSI* catheter-associated bloodstream infection, *CDC* Centers for Disease Control and Prevention, *NACHRI* National Association of Children's Hospitals and Related Institutions

are more common and readily accessible than others. A few of the more practical ideas are reviewed in the following paragraphs.

One central strategy to improve quality and safety is error-proofing. Error-proofing methods have been categorized and ranked in a number of ways, including a binary codification. Level 1 error-proofing effectively achieves total prevention of an error, whereas level 2 simply makes errors less likely. Level 1 error-proofing is clearly preferable but unfortunately more difficult to identify or devise. An example of level 1 error-proofing is a forced function achieved by anesthesia-machine medical-gas connections, whereby a pin system allows only the oxygen hose to be connected to the oxygen source, air-to-air and vacuum-to-vacuum, thereby disallowing a misconnect that would be possible if hose connectors were universal. Level 2 error-proofing methods are much more prolific yet individually insufficient. Often multiple level 2 strategies will be deployed for an additive or synergistic effect. Examples of level 2 error-proofing include color coding, decision-support tools, and pop-up warnings in computer order-entry systems. In contrast to this binary classification, error-prevention strategies can also be considered along a continuum of general effectiveness (Table 9.3). Unfortunately, the easiest and most reflexive strategies often gain the least traction by themselves—even when they provide a necessary foundation for higher-order error-proofing strategies.

Gawande and Pronovost have championed the use of checklists to help overcome the incredible complexity of modern medicine [66–68]. When checklists were first used in aviation, it was with a growing understanding that even the most experienced and venerated pilots could be confounded by the huge number of variables and sequential steps in a complex procedure (e.g., taking off in a World War II bomber) and that failure to execute seemingly minor steps (e.g., releasing the lock on the flaps) could result in catastrophe (e.g., crashing at takeoff). Fundamentally, checklists identify a minimum set of high-value practices immediately before an action or in real time during a process. Checklists primarily help mitigate errors of ineptitude (i.e., the failure to apply knowledge) but can also combat errors of ignorance (i.e., the absence of knowledge). Pronovost, for instance, demonstrated that use of a central-line insertion checklist helped correct preprocedural hand washing by physicians dramatically – in spite of the fact that physicians know that hand hygiene is an important practice. Furthermore, the effectiveness of rounding checklists to address daily goals is enhanced with explicit verbal prompting by team members rather than by passive completion by one rounding team member – reinforcing the QI tenant that it is not merely *what* is implemented, but *how effectively* it is implemented [69].

Evidence-based protocols or pathways (commonly manifested as order sets) help to guide care toward best or preferred practices and can provide cues to remember easy-to-forget nuances and exceptions. Somewhat distinct from checklists, protocols are often more detailed and may not be necessarily by used in real time. Protocols primarily

Table 9.3 Error reduction hierarchy

More common, less effective	**Education & encouragement**: providing information to staff and cultivating intentions
	**Clear rules & policies**: setting explicit expectation consistent with best practice in concrete/written/available form
	Audits: creating data to detect defects in care that can promote situational awareness and/or be acted upon
	Simulation: creating opportunities for staff to practice and leadership to assess adherence to rules & policies
	Standardization: implementation of protocols/pathways to reduce unnecessary variation and improve predictability
	Checklists, double-checks, closed-loop communication: promotes attention and adherence to critical steps real-time
	Simplification, lean engineering: eliminates unnecessary steps or distractions where errors can be introduced
	Making the easy way the right way (or hard to do it wrong): capitalize on human factors/nature to do it the easiest way
	Automation & computerization: machines can perform more reliably than humans (if effectively designed/implemented)
Less common, more effective	**Forcing functions**: a physical constraint that makes a misuse nearly impossible without deliberate modification or override

help mitigate errors of ignorance, but like checklists, they can also reduce errors of ineptitude. Checklists and protocols both help reduce unnecessary variation, which can promote predictability and thereby reduce errors of communication and teamwork. Both checklists and protocols can also combat errors of supervision (i.e., the lack thereof) and have been touted to help focus the mind on important opportunities for customization of care by leaving unspecified modifiable factors outside the scope of the tool. Checklists and protocols, alongside more advanced electronic health record systems, can also function as decision-support tools that effectively improve the quality of care while simultaneously reducing costs [70–72].

There are, however, substantial challenges to the acceptance of standardized care, use of checklists, implementation of protocols, and effectiveness of automated decision support. Experienced by many physicians as "cookbook medicine," such tools admittedly do not always accommodate the wide range of practitioner experience nor facilitate innovation or creative problem solving. As the old NASA saying goes about astronauts, "There are two ways to die in space: (1) not following the procedure exactly as written and (2) following the procedure *exactly* as written." It is probably fair to say that one has to know the rules to know when to break them – but not knowing the rules is a recipe for failure.

Another common tool in the improvement-science arsenal is optimizing the visual workspace [73–75]. The Five-S mnemonic is often invoked, standing for (1) *sort*/*scrap* what is needed versus not needed; (2) *straighten* (or *set in order*) to make a workspace organized, labeled, ergonomic; (3) *scrub*/*shine*/*sweep* to eliminate messes that obscure the organization; (4) *standardize* to make it easy for everyone to maintain and anticipate; and (5) *sustain order* with clear role responsibilities and discipline. Reducing unnecessary clutter and putting things in consistent order reduces search time for key materials and eliminates the wasteful warehousing of unnecessary items. Reported benefits have included improved efficiency, situational awareness, fewer lost items, streamlined supply chains, and improved staff and consumer satisfaction.

Root cause analysis (RCA) is, at its simplest level, a problem-solving methodology. Numerous formal RCA approaches have been well described in the literature, including fishbone diagrams and the "Five Whys" method of Toyota [76]. A robust RCA often illustrates the "Swiss cheese" model of error, where multiple latent defects (holes in the cheese), none of which alone are hugely problematic, all align to allow an error to propagate through the system and reach the patient. As already alluded to, an analysis of recognized errors or harm can provide constructive organizational learning. Such analysis is often undertaken in a fairly informal way, such as through qualitative discussion at morbidity and mortality reviews. However, formal RCA tools undertaken in an iterative manner can improve the likelihood of discovering latent defects in complex systems while simultaneously reducing some of the shortcomings of anecdotal analyses previously cited [42]. The overarching goal is to identify and eliminate causes of problems rather than to address immediately obvious "symptoms." RCAs can, but do not always, reveal system problems that will cause harm again if root causes are not recognized and addressed. Clearly, not all system problems are revealed by single events, so RCA is best undertaken as an iterative process, where recurrent themes are sought and contextualized. Often, categorical contributors are plotted on a Pareto chart

(e.g., a histogram ordered from high to low prevalence) to visually depict where the preponderance of errors or harm seem to arise. This can provide an area of focus for consideration of risk from latent defects in such domains.

Summary

At no other time have health-care quality and patient safety recognition been of greater import than they are now. The modern patient-safety movement continues to grow at an unprecedented pace, both at pragmatic and academic levels. The PICU is an environment rich with the potential for risk, error, and harm. It is also full of dedicated, bright, vigilant people who have a wealth of clinical and operational information at their fingertips. This creates a very fertile environment to be able to engage in and pioneer quality-improvement science. PICUs across the United States have reduced historical rates of central-line infections, ventilator-associated pneumonias, harmful medication errors, unplanned extubations, and other adverse events [41, 53, 77, 78]. Sustainability of such improvements is patchy, and enhancements of positive dimensions of care more elusive [79, 80]. Often, the ingredients lacking for QI to flourish in a PICU are (1) local know-how, (2) leaders committed to a quality-improvement vision, and (3) the time and resources of frontline staff members to execute such a vision. Table 9.4 offers a few practical tips for local PICU leaders who want to run QI initiatives. The development of clinical leaders with a quality-improvement vision and the procurement of resources to undertake "optional" initiatives remain real challenges for many PICUs, yet one can hope that the growing safety movement and national health-care reforms will continue to attack these voids with a constructive mixture of carrots and sticks [81].

Table 9.4 Tips for leading quality & safety initiatives from the trenches

Help set the proper scope/frequency of testing:
Usually smaller, more often, more variety of tests
Be a part of the testing:
Get first-hand experience, be wary of over-delegation
Distinguish bad ideas versus poor implementation:
Don't let good ideas fail because of poor execution
Allow operations to evolve:
Question habits/assomptions, prevent ruts/stagnation.
Engage higher order improvement strategies:
Think sustainability & reliability (not brute force)
Good leaders don't hoard information:
Create situational awareness for teams/frontline/staff
Be transparent, share data, network:
Never ask frontline for info that you don't feed back
Strive to have content and improvement expert in same body:
Use "quality coaches" to acquire skills, not substitute for them
Grow quality improvement leaders in your unit (instead of recruiting);
Support and educate frontline staff; identify who can actually "do" it
Don't wait until you're "ready":
Have teams meeting regularly to "do"; start before you're ready

References

1. Kohn LT, Corrigan J, Donaldson MS. To err is human: building a safer health system. Washington, DC: National Academy Press; 2000. p. xxi, 287.
2. Institute of Medicine (U.S.). Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001. p. xx, 337.
3. Audet AM, Doty MM, Shamasdin J, Schoenbaum SC. Measure, learn, and improve: physicians' involvement in quality improvement. Health Aff (Millwood). 2005;24:843–53.
4. Milstein A, Adler NE. Out of sight, out of mind: why doesn't widespread clinical quality failure command our attention? Health Aff (Millwood). 2003;22:119–27.
5. Berwick DM, Nolan TW. Physicians as leaders in improving health care: a new series in Annals of Internal Medicine. Ann Intern Med. 1998;128:289–92.
6. Improvement IfH. The imperative for quality: a call for action to medical schools and teaching hospitals. Acad Med. 2003;78:1085–9.
7. Kachalia A, Johnson JK, Miller S, Brennan T. The incorporation of patient safety into board certification examinations. Acad Med. 2006;81:317–25.
8. Rowe JW. Pay-for-performance and accountability: related themes in improving health care. Ann Intern Med. 2006;145:695–9.
9. Epstein AM. Pay for performance at the tipping point. N Engl J Med. 2007;356:515–7.
10. Miller SH. American Board of Medical Specialties and repositioning for excellence in lifelong learning: maintenance of certification. J Contin Educ Health Prof. 2005;25:151–6.
11. American Board of Medical Specialties. About ABMS: maintenance of certification. www.abms.org/Maintenance_of_Certification. Accessed 11 Apr 2012.
12. Berwick DM. The science of improvement. JAMA. 2008;299:1182–4.
13. Reason JT. The human contribution: unsafe acts, accidents and heroic recoveries. Farnham/Burlington: Ashgate; 2008. p. x, 295.
14. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication guidelines for quality improvement in health care: evolution of the SQUIRE project. Qual Saf Health Care. 2008;17 Suppl 1:i3–9.
15. Speroff T, James BC, Nelson EC, Headrick LA, Brommels M. Guidelines for appraisal and publication of PDSA quality improvement. Qual Manag Health Care. 2004;13:33–9.
16. Shaller D. Implementing and using quality measures for children's health care: perspectives on the state of the practice. Pediatrics. 2004;113:217–27.
17. Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. Crit Care Med. 1995;23:294–300.
18. Scanlon MC, Mistry KP, Jeffries HE. Determining pediatric intensive care unit quality indicators for measuring pediatric intensive care unit safety. Pediatr Crit Care Med. 2007;8:S3–10.
19. National Quality Forum issue brief: strengthening pediatric quality measurement and reporting. J Healthc Qual. 2008;30:51–5.
20. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348:2635–45.

21. Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. N Engl J Med. 2007;357:1515–23.
22. Taylor JA, Brownstein D, Christakis DA, et al. Use of incident reports by physicians and nurses to document medical errors in pediatric patients. Pediatrics. 2004;114:729–35.
23. Winters B, Dorman T. Patient-safety and quality initiatives in the intensive-care unit. Curr Opin Anaesthesiol. 2006;19:140–5.
24. Box G. Robustness in the strategy of scientific model building. In: Launer RL, Wilkinson WG, editors. Robustness in statistics. New York: Academic; 1979.
25. Carder B, Ragan P. Measurement matters: how effective assessment drives business and safety performance. Milwaukee: ASQ Quality Press; 2005. p. ix, 221.
26. Sessions R. How a 'difficult' composer gets that way. New York Times. New York, NY, 1950. p. 89.
27. Jimmerson C, Weber D, Sobek 2nd DK. Reducing waste and errors: piloting lean principles at Intermountain Healthcare. Jt Comm J Qual Patient Saf. 2005;31:249–57.
28. Kim CS, Spahlinger DA, Kin JM, Billi JE. Lean health care: what can hospitals learn from a world-class automaker? J Hosp Med. 2006;1:191–9.
29. Chalice R, American Society for Quality. Improving healthcare using Toyota lean production methods: 46 steps for improvement. Milwaukee: ASQ Quality Press; 2007. p. xiv, 302.
30. Liker JK. The Toyota way: 14 management principles from the world's greatest manufacturer. New York: McGraw-Hill; 2004. p. xxii, 330.
31. Liker JK, Meier D. The Toyota way fieldbook: a practical guide for implementing Toyota's 4Ps. New York: McGraw-Hill; 2006. p. xx, 475.
32. Reason JT. Human error. Cambridge/New York: Cambridge University Press; 1990. p. xv, 302.
33. Milch CE, Salem DN, Pauker SG, Lundquist TG, Kumar S, Chen J. Voluntary electronic reporting of medical errors and adverse events. An analysis of 92,547 reports from 26 acute care hospitals. J Gen Intern Med. 2006;21:165–70.
34. Wilmer A, Louie K, Dodek P, Wong H, Ayas N. Incidence of medication errors and adverse drug events in the ICU: a systematic review. Qual Saf Health Care. 2010;19(5):e7.
35. Savage I, Cornford T, Klecun E, Barber N, Clifford S, Franklin BD. Medication errors with electronic prescribing (eP): two views of the same picture. BMC Health Serv Res. 2010;10:135.
36. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. Qual Saf Health Care. 2003;12 Suppl 2:ii39–45.
37. Agency for Healthcare Research and Quality Guide to Patient Safety Indicators, Version 3.0a (May 1, 2006). http://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V30/psi_guide_v30a.pdf. Accessed 5 May 2011.
38. McDonald KM, Davies SM, Haberland CA, Geppert JJ, Ku A, Romano PS. Preliminary assessment of pediatric health care quality and patient safety in the United States using readily available administrative data. Pediatrics. 2008;122:e416–25.
39. Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf. 2006;32:585–90.
40. Sandora TJ. Prevention of healthcare-associated infections in children: new strategies and success stories. Curr Opin Infect Dis. 2010;23:300–5.
41. Miller MR, Griswold M, Harris 2nd JM, et al. Decreasing PICU catheter-associated bloodstream infections: NACHRI's quality transformation efforts. Pediatrics. 2010;125:206–13.
42. Cook RI. A brief look at the new look at complex system failure, error, safety, and resilience. Chicago: Cognitive Technologies Laboratory; 2005.
43. Pawson R, Tilley N. Realistic evaluation. London/Thousand Oaks: Sage; 1997. p. xvii, 235.
44. Godfrey MM, Melin CN, Muething SE, Batalden PB, Nelson EC. Clinical microsystems, Part 3. Transformation of two hospitals using microsystem, mesosystem, and macrosystem strategies. Jt Comm J Qual Patient Saf. 2008;34:591–603.
45. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? J Clin Epidemiol. 1997;50:223–9.
46. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol. 1995;48:71–9.
47. Bailar 3rd JC. The promise and problems of meta-analysis. N Engl J Med. 1997;337:559–61.
48. Huang DT, Clermont G, Kong L, et al. Intensive care unit safety culture and outcomes: a US multicenter study. Int J Qual Health Care. 2010;22:151–61.
49. Pronovost PJ, Berenholtz SM, Goeschel C, et al. Improving patient safety in intensive care units in Michigan. J Crit Care. 2008;23:207–21.
50. Pronovost P, Sexton B. Assessing safety culture: guidelines and recommendations. Qual Saf Health Care. 2005;14:231–3.
51. Mohr DC, Burgess Jr JF, Young GJ. The influence of teamwork culture on physician and nurse resignation rates in hospitals. Health Serv Manage Res. 2008;21:23–31.
52. Jain M, Miller L, Belt D, King D, Berwick DM. Decline in ICU adverse events, nosocomial infections and cost through a quality improvement initiative focusing on teamwork and culture change. Qual Saf Health Care. 2006;15:235–9.
53. Abstoss K, Shaw B, Owens T, Juno J, Commiskey E, Niedner M. Increasing medication error reporting rates while reducing harm through simultaneous cultural and system-level interventions in an intensive care unit. BMJ Qual Saf. 2011;20:914–22.
54. Davidoff F. Heterogeneity is not always noise: lessons from improvement. JAMA. 2009;302:2580–6.
55. Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability, and research. Jt Comm J Qual Improv. 1997;23:135–47.
56. Speroff T, O'Connor GT. Study designs for PDSA quality improvement research. Qual Manag Health Care. 2004;13:17–32.
57. Lloyd RC. In God we trust; all others bring data. Front Health Serv Manage. 2007;23:33–8; discussion 43–5.
58. Langley GJ. The improvement guide: a practical approach to enhancing organizational performance. San Francisco: Jossey-Bass; 2009. p. xxi, 490.
59. The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement. IHI Innovation Series white paper. Institute for Healthcare Improvement. Cambridge, MA; 2003.
60. Weick KE, Sutcliffe KM. Managing the unexpected: resilient performance in an age of uncertainty. San Francisco: Jossey-Bass; 2007. p. xii, 194.
61. Nance JJ. Why hospitals should fly: the ultimate flight plan to patient safety and quality care. Bozeman: Second River Healthcare Press; 2008. p. ix, 225.
62. United States. Agency for Healthcare Research and Quality, Lewin Group. Becoming a high reliability organization: operational advice for hospital leaders. Rockville: Agency for Healthcare Research and Quality; 2008. p. 33.
63. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. Qual Saf Health Care. 2003;12:458–64.
64. Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. Qual Saf Health Care. 2008;17:137–45.
65. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. Qual Saf Health Care. 2007;16:387–99.

66. Gawande A. The checklist manifesto: how to get things right. New York: Metropolitan Books; 2009.
67. Gawande A. The checklist: if something so simple can transform intensive care, what else can it do? New Yorker. 2007. p. 86–101.
68. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med. 2006;355:2725–32.
69. Weiss C, Moazed F, McEvoy C, Singer B, Szleifer I. Prompting physicians to address a daily checklist and process of care and clinical outcomes. Am J Respir Crit Care Med. 2011;184:680–6.
70. Amarasingham R, Plantinga L, Diener-West M, Gaskin DJ, Powe NR. Clinical information technologies and inpatient outcomes: a multiple hospital study. Arch Intern Med. 2009;169:108–14.
71. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330:765.
72. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. Arch Intern Med. 2007;167:1400–5.
73. 5S: A lean method to cut the clutter. OR Manager. 2007;23(3):15. Retrieved 11 Apr 2012, from ProQuest Nursing & Allied Health Source. (Document ID: 1244435041).
74. Hubbard R. Case study on the 5S program: the five pillars of the visual workplace. Hosp Mater Manage Q. 1999;20:24–8.
75. Waldhausen JH, Avansino JR, Libby A, Sawin RS. Application of lean methods improves surgical clinic experience. J Pediatr Surg. 2010;45:1420–5.
76. Gupta P, Varkey P. Developing a tool for assessing competency in root cause analysis. Jt Comm J Qual Patient Saf. 2009;35:36–42.
77. Brilli RJ, Sparling KW, Lake MR, et al. The business case for preventing ventilator-associated pneumonia in pediatric intensive care unit patients. Jt Comm J Qual Patient Saf. 2008;34:629–38.
78. Sadowski R, Dechert RE, Bandy KP, et al. Continuous quality improvement: reducing unplanned extubations in a pediatric intensive care unit. Pediatrics. 2004;114:628–32.
79. 5 Million Lives Campaign. How-to guide: sustainability and spread. Cambridge, MA: Institute for Healthcare Improvement; 2008. Available at www.ihi.org
80. NHS Modernisation Agency. Improvement leader's guide to sustainability and spread. Ipswich, England: Ancient House Printing Group; 2002.
81. Levy FH, Brilli RJ, First LR, et al. A new framework for quality partnerships in Children's Hospitals. Pediatrics. 2011;127:1147–56.