Since publication of the Institute of Medicine of the National Academies (IOM) report “To Err is Human” [1], the goals of safe, effective, patient-centered, timely, efficient, and equitable health care have become the mantra for laboratory quality. Although laboratory quality improvement plans have traditionally focused on the analytic phase, most errors occur during the preanalytic and postanalytic phases [2–5]. The greatest opportunities for gains in quality are in the preanalytic and postanalytic phases, but such gains require an integrated approach that engages clinicians, pathologists, couriers, technicians, and supervisors in the quest for improved outcomes. Increasingly, health care purchasers, patients, and payers are demanding accountability. Incentive plans offer financial rewards to organizations that can demonstrate superior outcomes and risk share for those who fall short.

In this issue of *Clinics in Laboratory Medicine*, we address opportunities for quality improvement throughout the testing cycle with the goal of enhanced clinical effectiveness and improved patient outcomes. I discuss opportunities for improvement and quality reporting and the development of quality measures. Robin Stombler addresses the lessons learned from the nation’s first experience with building an institute for laboratory quality. Ana Stankovic discusses errors in the preanalytic portion of the testing cycle, and Giuseppe Lippi, Roberto Fostini, and Gian Cesare Guidi focus on specimen collection. Brian Jackson addresses the dangers of false-positive and false-negative test results. Jay Brooks discusses quality improvement opportunities in blood banking and transfusion medicine, and Steven Raab and Dana Marie Grzybicki discuss the measurement of quality in anatomic
pathology. William Dupree focuses on a macrosystem perspective to quality, and Myra Wilkerson and Greg Strevig discuss electronic data interface and interoperability. Conrad Schuerch, Jay Jones, and Mark Selna discuss clinical effectiveness; Nancy Elder, John Hickner, and Deborah Graham address quality and safety in outpatient laboratory testing. Whitney High focuses on medical malpractice. We hope this issue will serve as a useful reference for your own quality improvement efforts.

Dirk M. Elston, MD
Department of Pathology
Geisinger Medical Center
100 North Academy Avenue
Danville, PA 17821, USA

E-mail address: Delston@geisinger.edu

References


Opportunities to Improve Quality in Laboratory Medicine
Dirk M. Elston, MD

Department of Pathology, Geisinger Medical Center, 100 North Academy Avenue, Danville, PA 17821, USA

Reduction of medical errors remains a high national priority. Since publication of the Institute of Medicine (IOM) report “To Err is Human,” laboratory technicians have strived to contribute to the greater goal of safe, effective, patient-centered, timely, efficient, and equitable health care. Reduction of laboratory errors is an important component of these efforts. Estimates of error frequency in anatomic pathology vary according to the detection method used, but may range as high as 7% to 10% [1]. Effective quality assurance programs can track error rates and provide opportunities for error reduction.

Laboratory errors may originate during any phase of the testing cycle, from the decision to perform the test through the final interpretation of the test and appropriate clinical response. Quality programs in laboratories traditionally have focused on the processes under direct control by the laboratory, particularly the analytic phase. These programs are essential for precise and reliable laboratory tests, but the greatest potential for error occurs during the preanalytic and postanalytic phases [2–5]. Significant gains in quality can be achieved through efforts directed at these phases [6–10].

Important sources of error in the preanalytic phase include errors in patient identification, specimen collection, transport, and accessioning. Additional sources of error include unintelligible requests and the decision to order an inappropriate test. In a study of the accuracy of frozen section margins for head and neck squamous cell carcinoma, errors, final margin reporting redundancy, and waste was caused mainly by lack of anatomic correlation at hand-offs between the surgeon and the laboratory [11]. Effective communication and flow of information are critical.

In the postanalytic phase, bad outcomes may be related to turn-around time, delivery of reports, or interpretation of results [12]. Accurate results
that are received too late or received by the wrong provider may result in harm. Medical errors related to failure of effective communication of a report remain common, even in highly computerized and integrated health care systems. The most common errors relate to transmission of reports for imaging studies, clinical laboratory studies, and anatomic pathology reports [13]. Interpretation of results is critical, especially where the pretest probability is very low or very high.

The analytic phase is the sole province of the laboratory, while the pre- and postanalytic phases involve the laboratory, the clinician, the patient, and many others involved in data entry, specimen collection, and transport. Patient identification errors occur in 1% in anatomic pathology and 0.05% of specimens in transfusion medicine. Multiple specimen transfers and handoffs contribute to the higher error rate [14]. An effective quality program requires active participation from all of those involved in the testing process. Each separate requirement for data entry creates an opportunity for error, as does data transfer between electronic systems [15,16]. Even when technologic improvement such as bar coding or electronic identification are instituted, errors may be related to incorrect codes from hospital transfer or previous hospitalization. Manual verification of scanned data still is required, even in the presence of electronic identification systems. In the postanalytic phase, errors related to communication can be reduced through the use of standardized keyboard macros or bar-coded phrase sheets [17]. Clinical decision support can be built into reporting systems [18]. To be effective, they should be developed and validated through the joint efforts of laboratorians and clinicians. Ideally, decision support begins with the decision to perform the test. Judgment in ordering studies is cited frequently as a deficiency in pathology residency training [19]. There is great opportunity for gains in quality in addressing appropriate test ordering within the laboratory, in the clinic, and in the hospital.

Pathologists can play a critical role in improving patient outcomes through decision support at the time of test ordering. They also play an essential role in decision support in the interpretive phase of a pathology report. Unfortunately, experience in quality assurance procedures, skill in communication, and the clinical knowledge to be effective consultants remain important deficiencies among newly minted pathologists [20]. Although most physicians are satisfied with laboratory diagnostic services, they commonly are not satisfied with communication with the laboratory [21]. These data suggest ample opportunity to improve the pathologist’s role in contributing to and measuring the clinical effectiveness of medical services.

Surgical pathology remains problematic for establishing and measuring quality standards. Standards are being developed to ensure quality in anatomic and autopsy pathology [22]. Progress has been hampered by a lack of uniformity with respect to identification and documentation of errors in surgical pathology [23]. Uniform definitions of errors related to the testing cycle can improve exchange of data and enhance quality assurance
programs in anatomic pathology [24,25]. Measurement of quality can be difficult. For instance, outcomes measurements are difficult in autopsy pathology where the clinical outcome is predetermined. Some laboratories have relied upon benchmarking of turn-around times and staffing levels as an opportunity for quality improvement [26].

In the ambulatory setting, cancer misdiagnosis accounts for most closed malpractice claims with substantiated harm to a patient [27]. The frequency of errors in cancer diagnosis varies between institutions, ranging up to 11.8% in one study. Sampling errors were common, with errors in interpretation accounting for 5.0% to 50.7% of the errors depending on the institution. A large proportion of the errors were associated with harm [28]. In another study, the mean and median frequencies of discrepancy in anatomic pathology were 6.7% and 5.1%, respectively. Almost one-half of the discrepancies related to a change within the same category of interpretation, whereas 21% related to a major change, such as a benign diagnosis that was changed to a malignant diagnosis. Roughly 5% of the discrepancies had an effect on patient care [29]. Intraoperative consultations have been identified as an important source of error in pediatric pathology [30]. In a study of 2839 intraoperative consultations, 115 discrepancies were identified (an error rate of 4%), of which seven (0.2%) were major [31].

An anatomic pathology error database can be used to track and target high-frequency errors and errors with high clinical impact [32]. Further studies are needed to define the value and appropriate uses of double reading, case conferences, and consultation.

Effective quality assurance programs require continued monitoring. Long-term monitoring of frozen permanent section correlation has been shown to result in sustained improvement in performance [33]. Monitoring programs in all areas of laboratory performance improvement can help to sustain the gains in quality that have been achieved.

In this issue, a group of experts addresses specific opportunities for quality improvement in laboratory medicine. It is hoped that the issue will serve as a springboard for an exchange of ideas to improve quality in all aspects of laboratory medicine.

References

Several questions posed during small group sessions at a 2003 conference hosted by the Centers for Disease Control and Prevention (CDC) offered some perceptions of the need for more organized quality-of-care initiatives within the laboratory community: “A new ongoing quality institute is needed to address patient safety and quality issues for laboratory services. Do you agree or disagree?” “Current professional laboratory and clinical organizations adequately address patient safety and quality issues related to the delivery of laboratory services. How strongly do you agree or disagree?” [1].

The unofficial tally showed that most of the participants queried either agreed with or were neutral toward the first question. As for the second statement, most believed that, when examining the full spectrum of quality and safety issues related to laboratory testing, most organizations fell short.

The CDC took the responses from the 2003 conference and initiated plans for a new public-private partnership organization devoted to improving the use of laboratory tests and services. As the Institute for Quality in Laboratory Medicine (IQLM) was being formed, workgroups were established to begin a process for identifying a core set of indicators for quality in laboratory medicine; identifying and demonstrating the feasibility of networks to collect critical information related to laboratory quality; and identifying award categories to recognize the best in quality laboratory practice. Volunteers were assigned to workgroups based on their background and areas of expertise. These first workgroups were chaired by Lee Hilborne, MD, MPH (UCLA Medical Center, Los Angeles, California) and Fred Meier, MD (Henry
Ford Medical Group Laboratories, Detroit, Michigan); Networks—Michael Noble, MD (University of British Columbia, Vancouver, Canada) and Barbara Goldsmith, PhD (Caritas Christi Health Care, Boston, Massachusetts); and Awards—Ana Stankovic, PhD (BD Diagnostics, Franklin Lakes, New Jersey) and Elissa Passiment, EdM, CLS(NCA) (American Society for Clinical Laboratory Science, Bethesda, Maryland). These individuals, along with Robin Stombler (Auburn Health Strategies, Arlington, Virginia) and staff of the CDC, also served as members of an executive committee to provide advice and direction to the IQLM.

Unique audience, unique needs

One of the early decisions made in the formation of the IQLM was to gather a full spectrum of stakeholders in the delivery of laboratory testing and services. The IQLM engaged patients, clinicians, laboratory professionals, manufacturers, information and biotechnology experts, government, and accrediting bodies to contribute to its mission of generating substantial improvements in quality, effectiveness, safety, and appropriateness in the broad universe of laboratory testing. This marked the first time that such a group of multiple stakeholders had been assembled to address quality issues in laboratory medicine.

The stakeholder organizations participated in the IQLM through one of two partnership advisory councils. The organizations are listed in Box 1. As expected, some organizations are more proactive than others. Some joined the effort to watch and listen, whereas others became actively engaged in the mission of the IQLM.

In building broad support from organizations representing these stakeholders, it was clear that the IQLM must target these very individuals to gain acceptance and achieve success in improving quality practices. Initially, the IQLM focused on balanced diversity in its advisory councils, workgroups, and taskforces, and encouraged the wide sharing of information among the organization’s membership. There was a conscious decision not to appoint workgroup members based on the organization they represented, but rather because of professional expertise. This approach helped to mitigate politics and encouraged a new, broad set of volunteers.

Despite its success in bringing disparate parties to the same table, the institute faced the challenge of motivating quality improvement among an equally diverse target audience. What motivates a medical technologist versus a surgeon versus an insurance company versus a manufacturer versus a patient may be quite different, and the IQLM needed to understand these differences so that the appropriate products and services would reach its audience.

No competing with the competition

There are other organizations that work on components of improving laboratory testing and services, but no prior organization focused on
engaging a diversity of stakeholders to encourage advancements in this arena. Existing organizations make valuable contributions to quality, and there was no need to duplicate their efforts. Some organizations offer accreditation of laboratory services (eg, The Joint Commission, COLA, College of American Pathologists); certification of personnel (eg, American Society for Clinical Pathology, American Association of Bioanalysts); regulatory approaches (eg, Centers for Medicare and Medicaid Services, Food and Drug Administration); or advocacy services (eg, American Medical Association, American College of Physicians). The IQLM was not established as an accrediting, certifying, regulatory, or lobbying organization, nor did it ever intend to be such. Other organizations offer those services, and the IQLM was determined not to compete in those areas.

The cry for educational opportunities, scientific research, and scholarly activity in the quality arena was loud and clear. Overuse, underuse, and misuse of care were identified by the National Roundtable on Health Care Quality as quality problems that need to be addressed [2]. A report by the Lewin Group found that the “appropriate use of diagnostics is integral to high quality health care, including informing earlier, more targeted health care interventions and averting negative health outcomes and unnecessary costs” [3]. The National Committee on Quality Assurance issued its 2005 State of Health Care Quality report noting the number of annual avoidable deaths and medical costs if performance measures were appropriately followed [4]. A preliminary analysis by the American Academy of Family Physicians’ National Research Network concluded that errors occur throughout the testing cycle, especially in the preanalytic and postanalytic phases, and that these errors commonly cause consequences for patients [5]. As outlined in Box 2, the IQLM positioned itself to serve as a premier source for information on laboratory practices and as a forum for all users, payers, and providers concerned with the quality and safety of laboratory services.

Independence day

The goal of establishing an independent organization was reached on July 11, 2005, when the IQLM was incorporated. The IQLM was recognized by the Internal Revenue Service as an organization exempt from federal income tax under Section 501(c)(3). As the IQLM promoted activities related to improving the quality of laboratory tests and services, the Dark Report listed the IQLM as one of the top 10 stories for 2005 (it ranked number 4). Olympus featured IQLM on its professional links page noting that it is a “proud member” [6], American Academy of Family Physicians Web site advertised the IQLM newsletter [7], and ARUP Laboratories issued press releases citing the IQLM [8].

Among its successes are the following:

- Over 70 major associations, corporations, and government agencies representing users, payers, and providers of laboratory services became
Box 1. Stakeholders participating in Institute for Quality in Laboratory Medicine advisory councils

Professional Partnership Advisory Council
AdvaMed
Agency for Healthcare Research and Quality
America’s Health Insurance Plans
American Academy of Dermatology Association
American Academy of Family Physicians
American Association for Clinical Chemistry
American Association for Respiratory Care
American Association of Bioanalysts
AABB
American Association of Pathologists’ Assistants
American Association of Physician Offices and Laboratories
American Board of Genetic Counseling
American Clinical Laboratory Association
American College of Medical Quality
American College of Physicians
American College of Preventive Medicine
American College of Surgeons
American Medical Association
American Medical Technologists
American Osteopathic Association
American Registry of Pathology
American Society for Clinical Laboratory Science
American Society for Clinical Pathology
American Society for Cytopathology
American Society for Healthcare Risk Management
American Society for Histocompatibility and Immunogenetics
American Society for Microbiology
American Society for Quality
American Society of Cytotechnology
American Society of Hematology
Armed Forces Institute of Pathology
Association of Public Health Laboratories
Centers for Disease Control and Prevention
Clinical and Laboratory Standards Institute
CLMA
College of American Pathologists
COLA
Federation of American Hospitals
Joint Commission
March of Dimes  
Medical Group Management Association  
National Academy of Clinical Biochemistry  
National Committee for Quality Assurance  
National Institute for Standards and Technology  
National Quality Forum  
National Society of Genetic Counselors  
Society for General Internal Medicine  
Society for Vascular Surgery  

*Technology Partnership Advisory Council*  
Abbott Laboratories  
Affymetrix  
Ambion Diagnostics  
American Proficiency Institute  
Argent Global Services  
ARUP Laboratories  
Bayer Healthcare Diagnostics Division  
BD  
Beckman Coulter  
Bio-Rad  
Cerner Corporation  
Clinical Micro Sensors, A Motorola Company  
Cytyc Corporation  
Dade Behring  
DiagnosisOne  
Digene Corporation  
Gen-Probe  
Genzyme Genetics  
Healthcare Information and Management Systems Society  
Intermountain Health Care  
National Institute of Standards and Technology  
Nexus Corporation  
Olympus Diagnostics Systems Group  
Omnitech Labs  
OraSure Technologies  
Orchard Software Corporation  
Ortho-Clinical Diagnostics  
Premier  
Quest Diagnostics  
Roche Diagnostics  
Streck Laboratories
Box 2. Established goals for the Institute for Quality in Laboratory Medicine

The Institute will provide a forum for communication and networking among professional, consumer, government, industry, and other organizations concerned with the quality and safety of laboratory services.

The Institute will facilitate and provide educational opportunities to advance quality care.

The Institute will produce and facilitate scientific research and scholarly activity to promote improvements in laboratory testing and services to benefit the health of the public.

The Institute will serve as a source for information on laboratory practices.

The Institute will provide mechanisms for generating improvements in the quality and safety of laboratory tests and services throughout health care systems.

---

Courtesy of the Institute for Quality in Laboratory Medicine, Arlington, VA; with permission.

part of the IQLM Professional or Technology partnership advisory councils. These organizations collectively represented over 350,000 physicians, approximately 200,000 laboratory professionals, 1200 health care manufacturing companies, major accrediting bodies, all state public health laboratories, 3 million volunteers dedicated to improving the health of mothers and children, federal government agencies, and a significant share of the nation’s hospitals. The IQLM was the first organization to bring together this diversity of stakeholders dedicated to achieving improvements in quality, effectiveness, safety, and appropriateness of laboratory testing.

- IQLM survey conducted to help define quality for laboratory testing and services. Ninety-five percent of the respondents noted that laboratory medicine is “important” or “essential” in addressing patient safety and information technology [9].
- Developed a list of over 200 volunteers offering their services and expertise to the IQLM.
- Produced a complimentary monthly electronic newsletter on laboratory quality and safety with more than 2000 subscribers, and established a Web site at www.iqlm.org to serve as a resource center.
- Published a series in Medscape General Medicine on controversies in laboratory medicine.
Collected data to assess evidence-based laboratory indicators to help reduce errors and improve safety and quality.

- Began the development of a network for all laboratories to obtain and share information about quality practices.
- Produced a long-term research agenda.

Based on the importance, diversity, and uniqueness of the IQLM, the CDC stated that the “IQLM is the appropriate and only qualified institution” to compete for a cooperative agreement on the improvement of laboratory testing and services in the United States. An announcement of this sole-sourced cooperative agreement application was issued publicly in December 2005.

Money costs too much

Despite its success and promise, the IQLM faced challenges. It needed funds to exist and operate as a viable organization. Perhaps to its detriment, the IQLM did not charge dues. Because the IQLM was not formed as a membership organization, assessing member fees was not deemed suitable. Also, to maintain the diversity of organizations participating in its advisory councils and to allow equal voice to all stakeholders, the IQLM chose not to levy dues for participation.

On June 22, 2004, the CDC issued a program announcement in the Federal Register entitled “Developing and Implementing the Institute for Quality in Laboratory Medicine.” The awardee of the cooperative agreement would be responsible for a number of quality-related projects and responsible for “manag(ing) a process to incorporate and implement an Institute for Quality in Laboratory Medicine, including the logistics of the formation, legal documents, and structure of the Institute.” This cooperative agreement was awarded to the National Quality Forum, but after months of negotiations between the CDC and the National Quality Forum, the scope of work changed and did not include responsibilities for the management and implementation of the IQLM.

In December 2005, the CDC issued another announcement, but this time it was for a sole-sourced, cooperative agreement opportunity for the IQLM. The funding was to be used to prioritize and manage projects to improve laboratory testing and services. The IQLM proposal was accepted. Yet, like the previous cooperative agreement, months of negotiations also failed to produce funding for these projects. The IQLM Board of Directors examined many alternate sources, but ultimately determined that funding had not been sufficient to support the needs of the IQLM.

Proud legacy

Because of the lack of adequate funding, the IQLM Board of Directors voted to dissolve the IQLM. As recognized by the Internal Revenue Service, this termination became effective after March 17, 2007. Nonetheless, the
presence of the IQLM endures within and beyond the laboratory. The IQLM provided a successful impetus for collegial relationships among users, payers, and providers of laboratory services. A number of organizations increased their commitment to the quality arena by implementing new workgroups and research. Subscriptions to the IQLM newsletter grew significantly to include thousands of subscribers. Reviews of the IQLM series on Controversies in Laboratory Medicine published in *Medscape General Medicine* were glowing. The volume of quality-related inquiries and messages of support from grassroots providers was astounding and attests to the need for quality-related services, programs, and collegiality. Attempts to feed this hunger for improvement and collaboration should continue to prosper in the shadow of the IQLM. We will all be better off for it.

**Resurrecting the Institute for Quality in Laboratory Medicine**

Efforts have been taken to protect the intellectual property of the IQLM should an opportunity to resurrect the IQLM present itself in the future. The IQLM was based on three principles:

1. Embrace all stakeholders, because it takes the combined efforts of all (from users to providers to payers) truly to impact change and improve the quality, safety, and effectiveness of laboratory tests and services.
2. Treat all stakeholders equally by creating a place where all voices may be heard and all advice may be rendered regardless of ability to pay.
3. Pursue an educational and scientific mission that offers an agenda of progress.

This philosophy, woven into the infrastructure of the IQLM, is essential for this diverse constituency to effect true change.

If the same questions asked in 2003 were posed to the IQLM stakeholders today, would their responses change? The IQLM has enlightened many to the possibilities that such an entity brings, but is the entity itself really needed? The ultimate goal of the IQLM was to advance health care through improved laboratory testing and services. The pursuit of this goal should be foremost.

**References**


Patient Safety: A Macrosystems Perspective

William B. Dupree, MD*, Fan Lin, MD, PhD

Department of Anatomic Pathology, Geisinger Medical Center, 100 North Academy Avenue, Danville, PA 17822, USA

It is estimated that gross revenues for anatomic pathology (AP) in the United States will exceed $11.3 billion in 2008 [1,2]. An aging population, new laboratory use trends by managed care providers, anticipated growing consumerism, and rapidly developing new technology combine to generate great optimism among economic analysts. Against this backdrop of financial optimism, AP laboratories are seen more commonly as businesses capable of generating large positive revenue streams. Viewed from this perspective, AP macrosystems, like all businesses, operate under the universal dictum, “no margin: no mission” [3]. An almost irrefutable corollary is that businesses cannot develop and maintain a sustaining margin through cost containment and cost reduction alone [4]. To survive in a free market, businesses must grow purposefully. Growth of a laboratory, through acquisitions, mergers, strategic partnerships, new tests, or recruitment of doctors and patients, engenders competition for limited resources. Competition necessitates adaptation to an ever-changing external environment. For AP laboratories, this milieu of competition creates a need to perform services better, faster, and less expensively. This triangle of forces creates a potential for increased errors in the AP test cycle (Fig. 1). Those who direct the AP process must understand and accept this risk and be prepared to manage it.

In essence, AP processes may be grouped into work units comprised of task-oriented microsystems staffed by knowledge workers commonly under the direction of managers who often have little or no background in the knowledge workers’ discipline [5]. Patients’ needs are at the center of any AP microsystem. It is in these frontline microsystem units where patients’ needs and care teams meet. Most times, patients’ needs are met best

* Corresponding author.

E-mail address: wbdupree@geisinger.edu (W.B. Dupree).

0272-2712/08/$ - see front matter © 2008 Elsevier Inc. All rights reserved.

labmed.theclinics.com
when the Microsystems of a laboratory, including business, technical, informatics, and professional modules, are bundled conceptually into a macrosystem [6]. Close coupling of AP work cells with a macrosystem construct (Fig. 2) creates a decision-making perspective that can improve service quality, reduce risk, and create a service delivery system that enhances patient safety. W. Edwards Deming, a guru of industrial quality improvement, proffered that only when the theories of workplace, worker, and work are linked intimately into a macrosystem can an outcome of ultimate quality
be expected [7]. Many of Deming’s guiding principles for macrosystems in search of quality apply to AP when the process is considered as a macrosystem [7].

In the past, the focus of AP laboratory actions aimed at patient safety were based primarily in programs of quality control and quality assurance [8]. More recently, quality-improvement initiatives have addressed microsystem issues, such as the definition and standardization of diagnostic criteria, standardization of AP processes, standardization of diagnostic lexicons, synoptic reporting, and even such basics as simply defining error in AP [9,10]. With the advent of a new-model AP paradigms centered on personalized pathology and business concepts, such as creating customer value propositions, differentiation in the marketplace, definition of best practice algorithms, and change management guidelines, the macrosystem perspective is evermore important in assuring patient safety. This article examines the engines of change that have an impact on AP macrosystems, the evolving macrosystem tactics being spawned to ensure the maintenance of patient safety in the face of revolutionary change, and macrosystem change theory. The usefulness of high reliability organization theory and normal accident theory, which are successful in reducing errors in nonmedical, high reliability macrosystems, is explored. Finally, the potential for enhanced patient safety through macrosystem consolidation, standardization, and communication in a surgical pathology best care algorithm based on outcome analysis is explored.

Evolution or revolution in anatomic pathology macrosystems?
Infrastructure seeds of change

Competition and changing AP laboratory expectations in health care delivery systems have led to dramatic changes in a short period of time. As late as the 1990s, a simpler paradigm existed in AP in which the matrix was that of an independent laboratory sector populated with many local independent laboratories. Major commercial laboratory companies were limited in number. Most hospitals were independent and operated single-site AP laboratories. AP services were provided primarily by private pathology group practices. Even publicly traded commercial laboratory companies outsourced nearly all AP to local pathology groups. By 2000, consolidation of hospital ownership triggered creation of an enlarged number of consolidated hospital laboratory organizations serving multiple hospitals. Hospital outreach programs were limited in number. More intermediate-sized AP regional and national laboratories began to emerge. With the consolidation of regional laboratories forming more powerful national organizations, AP as a profitable growing business sector caught the attention of national providers. In addition, there was a steady growth in the number of hospital laboratory outreach programs providing AP services to office-based
physicians in surrounding communities. Currently, two dominant national laboratories employ more than 10% of the nation’s 13,000 pathologists [11]. There is a steady growth in the number of specialty testing companies with proprietary or patent-protected diagnostic molecular technology. These firms want total control of AP specimens so that the protected tests can be performed only in their laboratories. Their business goal is to be the exclusive provider of these diagnostic technologies. Specialty physicians, such as urologists, dermatologists, and gastroenterologists, are establishing in-house AP services (technical component/professional component [TC/PC] arrangements, pod laboratories, and in-house AP laboratories) to bolster declining clinical revenues. Some pundits believe that change in AP not only will continue but also accelerate to a pace bordering on revolution. From a macrosystem perspective, poorly managed rapid change can lead to errors that jeopardize patient safety. To minimize change-induced errors resulting in compromise of patient safety, a macrosystem perspective must be adopted.

**Macrosystems must create customer value propositions to ensure patient safety**

The AP sector has achieved an unenviable consensus: nearly everyone is unhappy with it [12]. Payers complain about lack of access; high cost of services; the inability to determine differential quality among providers because of insufficient, easily accessible computerized quality data; and alleged excessive error rates [13]. Many believe that the current tort system neither deters negligence nor compensates adverse events adequately yet drives up costs [14]. Disenchanted clinicians in many areas view AP merely as a complex laboratory test and anatomic pathologists as no more than supertechnologists. Anatomic pathologists, increasingly forced into the posture of knowledge workers, are angry about having responsibility without commensurate authority: heightened stress, loss of respect, declining incomes, and lack of overall professional satisfaction [11,15,16]. Disenchantment extends beyond what is felt by payers and pathologists. Boards and class B pathology managers many times generally view pathologists as uncooperative, impending government regulations as excessive, and AP costs in the molecular era as potentially uncontrollable. Payers, such as insurance companies, large employers with self-insurance, and government agencies, complain about the lack of an easily accessible long-term laboratory database that can be focused on outcome analysis to establish best practice algorithms and take cost out of delivery systems. Further confounding the mission of AP macrosystems is a larger fundamental question: What is the primary goal of health care systems in general and AP macrosystems in specific: well being of individual patients or stewardship of limited resources and profitability [17]?
Customer value propositions: engines of change

If a man can write a better sermon or make a better mousetrap than his neighbor, though he build his house in the woods, the world will beat a path to his door.

—Ralph Waldo Emerson

Against this situational backdrop, to assure quality of care and to compete successfully for limited resources in the AP marketplace, AP macrosystems must create evermore attractive customer value propositions.

“Customer value proposition” has become one of the most widely used terms in business markets [18]. The irony is that there is no agreement as to what constitutes a customer value proposition or what makes one persuasive. At this juncture in AP, however, most experts agree that in addition to price, convenience, relationships, connectivity, and quality, the psychologic bias of decision makers becomes an important issue [19]. Businesses long have assumed that clients adopt new products that deliver more value or usefulness than existing ones. Therefore, it follows logically that businesses only need develop innovations that objectively are superior to incumbent products and that consumers will have sufficient incentive to purchase them. This perspective of relative advantage articulated in the 1960s by Everett Rogers and his antecedent, Ralph Waldo Emerson, fails to capture psychologic biases that affect decision making [19]. Daniel Kahneman explained the issues of why and when individuals deviate from rational economic behavior (Box 1) and for his efforts won a Nobel Prize in Economics [20]. Based on this construct, laboratory representatives, in a frantic effort to compete successfully for limited resources by attracting

---

**Box 1. Characteristics of human responses to alternatives in the market place**

1. People evaluate attractiveness of an alternative based not on its objective or actual value but on its subjective or perceived value.
2. Consumers evaluate new products relative to a reference point, usually a product they already own or consume.
3. People view any improvements relative to this reference point as gains and treat all shortcomings as losses.
4. Losses have a far greater impact on customers than similarly sized gains.

accounts governed by biased decision makers, may promise a “have it your way” service, which can increase tensions between sales and operations significantly and jeopardize patient safety. Although it is necessary to guard against this trend, differentiation in the marketplace is necessary to compete successfully. Recent trends of merit in the differentiation of AP macrosystems include

1. Near totally subspecialized models for delivering AP services
2. Pay-for-performance paradigms
3. Disease-oriented reporting formats
4. The promise and risk of new technology

Near totally subspecialized models for delivering anatomic pathology services: an engine of change

Pathologists apply a confusing array of diagnostic criteria that are evidence based and empiric to a small number of difficult cases randomly mixed with a large number of routine cases, often with incomplete or misleading clinical information. Within this milieu, generalists infrequently may be exposed to a particular lesion type as a function of disease prevalence, caseload, or chance. In AP, subspecialization means seeing a higher volume of a particular lesion, which leads to additional practice or experience, increased skill level, and, hence, more chances for improving quality of care, reducing error frequency, and ensuring patient safety. Further confounding the practice of generalists is the shift in practice paradigm from a primary focus on morphologic diagnosis and prognostication toward personalized pathology, which includes molecular profiling and monitoring of individualized therapy. With this expanded mission comes the need for a multitude of skill sets, including molecular pathology, cytogenetics, protein biochemistry, advanced immunohistochemistry, advanced molecular techniques, and bioinformatics. These skill sets must be coupled with the ongoing need to remain current in an ever-growing classic AP database spanning at least 12 organ systems. At what point is the capacity of a generalist saturated? At what point is patient safety sacrificed? At what point should logistical convenience and a generalist’s desire for a variegated caseload producing intellectual satisfaction and fleeting new job potentiality be trumped by the higher goal of patient safety?

From a macrosystem perspective, anatomic pathologists share the economists’ dilemma that the resources to produce services are finite but consumers’ desire for services seems infinite. The macrosystem perspective is inescapable in that the more efficiently resources are used in the production process, the more consumers’ wants can be satisfied. The macrosystem challenge is to identify ways to maximize patient safety, meet clinicians’ desires, improve the standard of care, and grow the book of business in the face of limited resources. Economists have known for more than 200 years that specialization
can benefit a macrosystem by improving efficiency and increasing production without increasing resources. Adam Smith, in *The Wealth of Nations*, argued that the key to improving the wealth of individuals, organizations, states, and countries (macrosystems) is “for each to produce what each produces most efficiently among the array of things desired by a society” [21].

With consolidation of smaller AP practices into larger macrosystems, the need for general anatomic pathologists diminishes and the ability to support a near totally subspecialized model increases. From a macrosystem perspective, this model differentiates the delivery system, increases efficiency, and, although still unsubstantiated by data, likely improves turnaround time, quality, and patient safety. Careful integration of this macrosystem construct into the microsystems of the delivery process, however, currently remains a challenge. Size and cost constraints may dictate that pathologists cover more than one organ system service, producing a potential dilutional effect on benefits. Nonalignment of pathologists more comfortable in the generalist model may create anxieties and tensions within the system, sapping valuable energy and time. In systems with service needs at remote sites lacking telepathology capabilities, coverage of frozen sections can be a safety issue and demands maintenance of overall generalist skills. Prudent integration of macrosystem constructs into microsystem processes is mandatory to avoid outcomes that compromise patient safety.

**Pay for performance paradigms: engines of change**

Quality is defined by Webster as “a degree or grade of excellence”. Traditionally, the analysis of quality in AP has been considered in terms of quality control and quality assurance. Quality-control programs evaluate the uniformity of a specific process to assure that it is operating within acceptable parameters [22]. In essence, it involves a comparison between actual performance and performance levels espoused in departmental procedure manuals. Quality assurance is a term used to indicate a system designed with internal quality checks that encompasses a higher level of oversight and relies on the collection of outcome data [22]. The outcomes monitored usually involve multiple processes or procedures. Only recently, by many standards, has the AP laboratory sector as a whole begun to adopt concepts of quality improvement aimed at reducing diagnostic errors, shortening turnaround time, and proactively improving clinical outcomes [10]. Too often, errors are considered microsystem failures rather than assessed for their root cause from a macrosystem perspective. Forces brought to bear by tensions in a macrosystem may produce errors that on first analysis may seem to be microsystem problems.

The solutions to challenges of quality in anatomic pathology remain the holy grail. Quality, patient safety, and evidence-based medicine share a common link in that each requires careful measurement of individual work processes and patient outcomes. Integration of health care informatics
technology is making this task increasingly easy and less costly. It is anticipated that increased scrutiny and public ranking of providers by quality of outcomes will motivate third-party payers and clinicians to select AP macrosystems as a function of their demonstrated performance metrics. AP laboratories that have quality metric modules incorporated into their everyday macrosystem processes not only will be able to improve their quality of service consistently and, hence, increase patient safety but also will have a significant advantage in the marketplace.

**Disease-oriented reporting formats: engines of change**

A disease-oriented laboratory report not only will differentiate the macrosystem but also serve as a bridge to the future by assimilating traditional AP knowledge with new millennial molecular perspectives. This function has the potential to reduce errors and improve patient safety. As early as 1968, Dr. Lawrence Weed described the beginnings of his problem-oriented record concept, which envisioned a medical record that in addition to being a documentation of the events of a patient’s clinical course would “guide and teach” [23]. In this spirit, a disease-oriented laboratory report in many instances will serve to convert more traditional AP data into information and, where possible, knowledge coupled with molecular studies usable at the bedside by members of a background-diverse medical care team. This report will serve as a one-stop shopping center for AP outputs necessary for the care of patients who have a given disease process. In essence, a disease-oriented laboratory report not only will organize relevant microsystem outputs around a specific disease process but also, by using visuals, such as digital pictures, graphs, and charts, help convert data into information and, in many instances, information into knowledge. The data fields in toto will create subsets of information that are transformed into knowledge by linking all relevant laboratory information to visuals that correlate patient status with the collective experience of the world’s literature. The practice of medicine is about resolving doubt and making decisions by processing data into subsets of information and linking these to experience acquired in an a posteriori fashion. A disease-oriented AP report facilitates this process, improving quality of care and ensuring greater patient safety. The execution of a disease-oriented reporting format is beyond the scope of most AP microsystems and requires the resources of a macrosystem as a whole.

**New technologies: engines of change**

Progress in AP most times is driven by technologic advances [24]. The mission of AP is expanding from being predominantly diagnostic to including prevention, prognostication, therapy monitoring, and evaluation of outcomes [24]. This more personalized approach makes obsolete the paradigm
that divides medical laboratories into the arbitrary divisions of clinical pathology and anatomic pathology. Commoditization, molecular pathology, and a disease-oriented reporting format will blur this artificial separation further. Rapid development of new technology will drive radical market changes and redefine underlying economic and clinical pathways of health care [11]. More and more companies are shopping outside their organizations for innovation. Innovation relevant to a given macrosystem’s needs may be found as raw ideas, market-ready ideas, or market-ready products [25]. Each category of innovation has its own reach, cost, risk, and speed variables. AP macrosystems will need to assess their level of acceptance for each of these variables. Some technologies will fulfill their promise; others will not. Skillful shopping in the technology bazaar of innovation will determine success or failure. Macrosystems that choose technologies that do not fulfill expectations will be at financial risk. Successful macrosystems will be those that consistently can predict which new technologies will grow and then can execute them rapidly. A nimble macrosystem centered around a management culture of execution with a well thought-out new technology infrastructure, business plan, implementation algorithm, and metric by which to determine when and why a new test procedure should be divested in not only will be a leader in the market but also will enhance quality and patient safety. Consolidation of macrosystems, such as radiology and AP, will enhance patient safety further through shared new technologies. In vitro and in vivo diagnosis supported by new imaging technologies molecular constructs and powerful computers already is the vision of change on the drawing boards of many large diagnostic biotechnology companies [11].

A proliferation of change engines

Change in AP is occurring so rapidly that many question if evolution or revolution is underway. Information/knowledge is the end product of all AP macrosystems and must be transmitted and stored in an easily useable fashion. Health care systems are working to integrate AP information/knowledge into real-time patient electronic medical records [26]. Office-based physicians are implementing electronic medical records that require connectivity enabling electronic ordering and electronic reporting back to the office. Each upgrade of service may introduce a point of entry for error. Software aimed at monitoring AP processes (middleware) is being developed rapidly. Overdependence on these packages may lead to undetected error jeopardizing patient safety. Payers are in search of ways to contain escalating AP costs while maintaining a high standard of care. Evidence-based medicine algorithms offer the best construct to date for obtaining this goal [27].

To develop these algorithms, meta-analysis of long-term AP information is a requisite. Regional health information networks are evolving to meet
this need [28]. A requisite for contracting with many managed care providers is easy access to macrosystem AP data. Health care in general and AP specifically, with a push from employers, is becoming more comfortable with the use of quality management systems. Tools, such as Lean and Six Sigma, provide competitive advantages for AP macrosystems while enhancing quality and ensuring better patient safety [29]. With such widespread, rapidly accelerating change, effective change management becomes more than a corporate boardroom construct. Effective change management in AP macrosystems may be a cornerstone of patient safety.

**Macrosystem change management: a cornerstone of patient safety?**

Academicians and practitioners seeking to comprehend challenges and opportunities presented by rapid change within a complex macrosystem increasingly are consulting the writings of political and military strategists of the past. Machiavelli’s most famous work, *The Prince*, is used widely as a map for directing modern corporate macrosystem change. Although Machiavelli wrote approximately 500 years ago, his political world has much in common with AP macrosystems of today, which are struggling with the turmoil created by challenges to the status quo. This, in part, is the result of what Parel describes as Machiavelli’s fundamental axiom: “human nature is the same always and everywhere” [30]. Macrosystems are comprised of microsystems populated by people forming in aggregate a society that must be governed. Jay aptly argues that the new science of macrosystem management is only a continuum of the old art of government [31]. Macrosystem management and political theory may be but two similar branches of the same tree.

One application of Machiavelli to management and business studies is within the realm of leading change in a macrosystem. The pillar of this application is individuals as the level of analysis. Machiavelli posited that it is mandatory for any leader of change to be aware of human nature and thus “judge and calculate” whether or not it is better to be loved or feared in executing the change process. A contingency approach characterizes Machiavelli’s view of individual modus operandi in making decisions. Berlin contends that Machiavelli’s approach to leadership disregards Tayloristic notions of one best way in favor of the existence of multiple realities [32]. It is Berlin’s contention that Machiavelli believed that individuals are autonomous beings not entirely bound by norms and conventions, which allows for consideration of individuals as enriched by historical and cultural artifacts surrounding them. One absolute, however, is that one can “Trust people to serve their own interests.” Machiavelli argued that there is a social covenant that regards self-interest as a form of weakness. The will to serve self-interests is so strong, however, that most people are blind to the strength of this urge. To lead change in a macrosystem, this blindness must be recognized and exploited by creating points of alignment. When malignment is
allowed to occur, disequilibrium is created, which can sap a microsystem’s energy and, in the case of AP systems, lead to errors that have a negative impact on patient safety.

Adapting to change is one of the principal theories permeating Machiavelli’s work. Machiavelli accepted the reality that change is a tumultuous process resisted by some and welcomed by others. He maintained that structural change, which follows from customary tradition, requires less drastic action to secure than transformative change. Fig. 3 shows a model of organizational change adapted to AP developed from a review of contemporary organizational change literature. The model posits that the success of organizational change intervention is premised on the state of equilibrium between the opposing forces of change within an organization. This dynamic serves as a central theorem of Machiavelli’s work and remains the cornerstone of leading macrosystem change in AP from a human resource perspective. Long before Cummings and Worley (1993) identified strong leadership as critical to the successful implementation of internal change, Machiavelli declared that leaders need to be perceived as having strong personal attributes and the ability to instill loyalty in their followers [33,34]. Often overlooked in AP management discussions is the importance of strong macrosystem leadership in ensuring patient safety. Macrosystem change management, the cornerstone of patient safety in the new AP paradigm, may require Machiavellian insights.

Communication is a central challenge in bringing about successful change in a macrosystem [35–39]. Traditional best practices of change
communication (ie, involving employees in all aspects of a program and cascading information through layers of management) may have contributed to the downfall of many change implementation initiatives in the past. Many experts in change leadership communication are rethinking this best practice algorithm and now suggest an assertive humane approach, in which communication combines speed with clarity and defines excellence in action as a more effective strategy. Under this new paradigm, faster communication and action seem superior. The business case for every redesigned work process is based on improving patient safety, making money, saving money, or adding value. Stakeholders appreciate speedy execution. Clients are pleased by safer, better, less expensive, and faster service. Equally important is that speed makes the overall change more palatable and humane. Employees who adapt to the new work behaviors are reinforced and rewarded sooner, eliminating their apprehensions more quickly, hence dispelling human entropy within the system, which can lead to errors and jeopardize patient safety. Employees who choose not to change their behaviors, or who are unable to, can be made “available to the industry” more quickly. The longer process allows for more unsafe discord, which can jeopardize patient safety.

Creating an anatomic pathology culture of safety: the macrosystem perspective

Much attention recently has focused on patient safety in AP. To date, much of the attention on reducing errors and improving patient safety has focused on microsystem issues. To create a culture of safety, however, especially in the wake of radical change, consideration must be given to macrosystem approaches. Ackoff and Emery [40] have emphasized the need for systems thinking in regards to quality, suggesting that optimizing a microsystem or components of a macrosystem, in particular a purposeful system, such as AP, actually may harm final performance. For example, if a laboratory wants to improve the function of its histology operation by increasing use and reducing excess capacity, it might pare down the histotechnology budget by reducing full-time equivalents. The net system-wide effect would see delays in turnaround time, lengthening hospital stays, and increasing overall costs to a greater level than the savings accrued by the initial microsystem action. Hammer and Champy produced a study in which they offered an analysis similar to Ackoff and Emery’s, which recommended close consideration of a macrosystem rather than focusing exclusively on microsystems in isolation when quality improvement is the goal [35]. Systems thinking often is disregarded in attempts to create a culture of safe change in an AP macrosystem. At the macrosystem level, two paradigms used by nonmedical, high reliability organizations to understand error include normal accident theory and high reliability organization theory. These macrosystem approaches to understanding error can be useful in AP.
High reliability organization theory contends that humans who participate in and manage complex macrosystems themselves are not sufficiently complex to sense and anticipate the problems generated by the system. It seems likely that the more recent surge in the development of AP computer middleware will continue to address this issue.

Safety problems can be traced to errors. Many have become accustomed to viewing errors as aberrant developments. Perrow was one of the first to reject this belief and contended that dispelling this myth is the first step in developing a safety culture [35]. Perrow posits that outcomes that result from multiple and unexpected errors are intrinsic to most activity and even inevitable in some settings—they are normal and to be expected. In essence, normal accident theory argues that errors are a consequence rather than a cause of a problem and unless this is accepted and incorporated into a macrosystems culture, progress cannot be made. Many errors arise from systems failure and the key is to learn how to design systems better to minimize future errors. The goal is minimization rather than absolute avoidance, because anticipation of everything that can go wrong in an ever-changing macrosystem, like AP, is impossible.

A key to improving patient safety is minimization of errors by understanding their precursors better. Several systemic causation models are advanced in the literature. Reason postulated [38] that patient safety–compromising errors occur when human, organizational, and technical defenses are inadequate or lacking. These defenses must be in equilibrium with two processes common to all macrosystems, production and protection. Productive macrosystems, such as AP, expose people and assets to danger and, hence, require various forms of protection to intervene between errors and their possible victims. To understand how these defenses can fail, Reason proposed a construct that he called the Swiss cheese model of defense. Some defenses are structural, others rely on people, and others depend on procedures and administrative controls. Although most defenses are functional, they all have weaknesses. Optimally, each defense layer is intact. According to Reason, however, the defenses are more like slices of Swiss cheese, having many holes. Unlike the cheese, however, these holes continually are opening, shutting, and shifting their location. The presence of holes in any one slice normally does not jeopardize patient safety. Most times, a compromise in patient safety happens only when holes in many layers momentarily line up to permit a trajectory of accident opportunity, bringing error into damaging contact with patients. Reason categorized defense holes further as acute failures and latent conditions. Active failures are the unsafe acts committed by people who are in direct contact with the systems production forming a human/system interface. Active failures include slips, lapses, fumbles, errors, and procedural violations. Latent conditions may be likened to resident pathogens within a system. They arise from decisions made by designers, builders, or managers. Latent failures also may be inherent in current procedures and may lie dormant for many years before they
combine with active failures to create a violation in patient safety. Impending changes in AP have the potential to create holes in defenses by (1) penetrating latent conditions previously lying dormant, (2) creating new latent factors, and (3) introducing active failures through nonalignment of personnel.

Under the guidelines of normal accident theory, most times, compromise of patient safety is the result of macrosystems failure rather than individuals perpetrating errors. Reason notes that the best people can make the worst errors as a result of latent conditions. Understanding the interplay between active failures and latent conditions in the ever-changing milieux of AP macrosystems could lead to an improved risk management program and a more comprehensive patient safety program aimed at the person, the team, the task, the workplace, and the institution as a whole.

**Market forces regulation and standardization: the future of anatomic pathology macrosystem safety**

AP macrosystems are part of the larger universe of United States health care. In the United States, the health care system often fails to deliver on the promises of science and new technology [41]. It is estimated that five to seven people die as a consequence of medical error every 15 to 20 minutes and another 85 to 113 patients are harmed significantly [42]. Health care safety expert Lucian Leape compares the risk of entering an American hospital to that of parachuting off a building or bridge [43]. A growing number of experts contends that the problem with American health care is rooted in regulatory and market failures [44]. Their contention is that institutions and processes mandated by law and custom are preventing demand for health care from matching efficiently with those most capable of providing it. Reimbursement fee schedules for AP services may reflect this disconnect between market forces and regulation. Many pundits advocate strengthening market mechanisms; rewarding doctors according to patient outcome, rather than the number of patients they treat; and providing information about health care provider effectiveness.

Although far-reaching market reconfigurations and regulatory change likely will take professionals and legislators time to craft, Geisinger Health Systems has embarked on an experiment to test a macrosystem standardization safety/outcome construct. More recently, Geisinger Health Systems has gone so far as to offer the equivalent of a 90-day warranty on elective cardiac bypass surgery [45]. Geisinger essentially guarantees its “workmanship,” charging a flat fee that includes 90 days of follow-up treatment. Even if a patient suffers complications or must return to the hospital, Geisinger promises not to send the insurer another bill. This radical new way to encourage hospitals and doctors to provide higher quality care and avoid costly mistakes rests squarely on macrosystem standardization.
In analyzing how cardiac bypass surgery is performed at Geisinger, 40 essential steps were identified and procedures were put in place to ensure the steps always would be followed in a standardized way regardless of which surgeon or which one of the systems three hospitals was involved. Geisinger’s 40-step system makes sure every patient gets the recommended standardized approach. Although the experiment is in its incipient stages, preliminary data suggest patients are less likely to return to intensive care, spend fewer days in the hospital, and are more likely to return directly to their own homes instead of going to nursing homes. Should the data continue to show improved outcomes trends, it is hard to know if this approach, which Geisinger calls Proven Care, will catch on with payers. Will similar approaches evolve for the medical laboratory?

Modern nonmedical industries long have understood the importance of standardizing the way they manufacture their widgets to improve quality and reduce costs. AP macrosystems have been slow to focus their attention on standardizing the way they deliver care. This lack of standardization may contribute to the gap that exists between AP macrosystem performance and the skills and intentions of the people who work in it. AP macrosystems are highly complex systems with many opportunities for ambiguities in terms of how the work of many individuals should be coordinated successfully into an integrated whole. Unless everyone is completely clear about the tasks that must be done, exactly who should be doing them, and just how they should be performed, the potential for error always will be high. Fully acknowledging that AP is part science and part art, there is a generous proportion of process at the microsystem and macrosystem levels that contributes to reduction in errors, increased patient safety, and optimal outcomes. The anticipated trends in pay-for-performance reimbursement models well may be the catalyst that sparks the development of standardized best care and outcome-based algorithms, which can be applied at the microsystem and macrosystem levels in AP. The time for “extended warranties” on workmanship in AP is at hand.

References


Informatics Tools for Quality in Anatomic Pathology

Jeffrey W. Prichard, DO, Gregory Strevig, Adam M. Brown, Myra L. Wilkerson, MD *

Department of Pathology and Laboratory Medicine, Geisinger Medical Center, 100 North Academy Avenue, Danville, PA 17822, USA

In clinical laboratories, pathologists are accustomed to complying with regulations and accreditation standards related to quality control and test reproducibility; however, these parameters are not so easily defined in anatomic pathology, which is still largely subjective and interpretive. Geisinger Medical Laboratories (GML) and many other pathology groups began developing quality planning initiatives following the release and publication of reports from the Institute of Medicine in 1999 [1] and the Rand Corporation in 2003 [2] that focused on system errors in health care delivery and poor disease management in the United States and the serious consequences to the patient population. The anatomic pathology group at GML decided to design a program for quality assurance and quality improvement that would encompass the entire work unit, starting with clinician ordering and ending with final coding and billing. The group started by breaking down all workflow components into steps and listing what currently could be measured at each step that is captured electronically in the anatomic laboratory information system, as well as what the group would like to measure or track at each step that would require additional information technology resources. The group next broke the workflow down into pre-analytic, analytic, and postanalytic phases, similar to the practice in the clinical laboratories. This article presents three projects that have resulted from the quality assurance/quality improvement program that is still being developed and refined. The first project is a tracking monitor that follows specimens and data through the entire testing process and provides a snapshot at any given time of where specimens and data are in the workflow throughout

* Corresponding author.
E-mail address: mwilkerson@geisinger.edu (M.L. Wilkerson).
all three testing phases. The second project is an analytic phase assessment of Her-2/neu immunohistochemistry (IHC) testing using digital imaging. The third project is postanalytic and concerns results reporting.

**Project I: surgical pathology specimen tracking**

*Background*

Anatomic pathology traditionally is a series of manual processes through which specimens flow and commonly are tracked with manual logs. This system makes it difficult to know where a specimen is within the workflow at a specific point in time, to know how long the specimen spends at a given point or process, and how to identify bottlenecks within the workflow.

*Goals*

The defined goals for the GML surgical pathology specimen-tracking project were to

- Identify steps in the workflow process at which a specimen can be tracked
- Develop a real-time Web-based specimen-tracking monitor that can be viewed by all pathology personnel

*Project details*

The group began this project by identifying clear steps in the workflow process at which a specimen could be tracked electronically using data currently collected in the pathology information system (CoPath Plus version 2.7, Cerner Corporation, Kansas City, Missouri). These steps are called “tracking events.” Although a specimen may not proceed through every tracking event, each specimen has a current status and is at a current step. The surgical specimen workflow was divided into nine tracking events that could be monitored electronically:

1. Accessioned surgicals: These cases have been accessioned in the surgical pathology laboratory and are awaiting gross dictation.
2. Report entry (gross): These cases currently are having gross description text entered or edited in the pathology system.
3. Histology: These cases have had gross description text entered in the pathology report but still are waiting for slides to be completed and released, or “verified,” by histology in the pathology information system.
4. Pathologist: At this step, the cases have had gross description text entered in the pathology report, and all slides have been verified by histology. The slides are in transit from histology to the pathologist, are in the pathologist’s inbox, or are in the pathologist’s office.
5. Report entry (microscopic): These cases currently are having microscopic description text entered or edited in the pathology report.
6. Awaiting sign-out: At this step, cases have had a final diagnosis entered into the pathology report and are awaiting final pathologist review and electronic sign out.

7. Signed out: These cases have been signed out electronically by the pathologist and are waiting for electronic report distribution.

8. Report distributed: These cases have been distributed electronically to patient data repositories and are awaiting billing review.

9. Billing reviewed: These cases have been reviewed and coded by billing personnel and are in the process of being billed.

Next, the group developed a tracking monitor to display the tracking events graphically on an interactive bar graph. This tracking monitor is deployed via an intranet Web page, and the tracking event data are updated every 5 minutes. The Web page was developed using ColdFusion (Adobe Systems Inc., San Jose, California), and the database was created in Microsoft SQL Server 2000 (Microsoft Corporation, Redmond, Washington). Each 5-minute snapshot contains data about every specimen that falls within the accessioned tracking event through the billing reviewed tracking event. If a specimen has been billed successfully, it falls off the tracking monitor. The user has the ability to mouse over any bars on the graph to see how many cases currently are at that tracking event. The user also may click on any of the tracking event bars to display specific information on cases currently at that step, including accession numbers and the primary pathologist assigned to those cases. Therefore, users can determine at a glance the status of a specific specimen, what steps it has already passed through and where it needs to go, as well as how much work may be coming their way. For example, histology can see that a large number of specimens has been accessioned and will be arriving shortly. A pathologist can see that histology has just verified slides because the pathologist bar has just increased, signaling that specimens are ready for pathologist review.

Because it is possible to determine what specimens are at which tracking events, it also is possible to calculate how long a specimen has been at each step. On the graphical tracking monitor, each tracking step has three bars associated with it, “Normal,” “Caution,” and “Overdue.” There are rules for how long a specimen should spend at each tracking event. For example, the electronic results distribution for cases that have been signed out should run every 5 to 10 minutes. Thus, cases that are in the signed-out status less than 5 minutes are in the green-colored “normal” bar. Cases that have been in a signed-out status between 5 and 10 minutes are in the yellow-colored “caution” bar. Finally, cases that have been in a signed-out status for longer than 10 minutes are flagged as “overdue,” the red-colored bar. Therefore, if a user notices a sudden increase in cases in the caution and overdue bars in the signed-out tracking event, it would be wise to verify that the electronic method of results delivery is functioning. Similar rules may be implemented for each of the tracking events to help identify bottlenecks in the overall
specimen workflow and to allow users to check the overall system status at a glance. An example of the tracking monitor output is shown in Fig. 1.

Conclusions

The group is continuing to develop and refine rules for this tracking event monitor. In addition to tracking surgical specimens, the group will develop a more granular tracking monitor for specific histology tracking events through the use of barcodes on tissue cassettes and slides. It also is developing a tracking ticker that can run on a user’s desktop, providing easier access to the graphical monitor. It will add e-mail notification for cases that have entered an overdue status. Large-screen monitors will be placed in surgical pathology, transcription, and histology so that the overall system status can be checked instantly by any worker in those areas and any red status bars can be noted immediately and addressed as needed.

Project II: applying statistical quality control to immunohistochemistry through digital imaging

Background

The FDA, in 1998, approved trastuzumab (Herceptin, Genentech, San Francisco, California) for use in treating breast cancer, ushering in the
age of targeted molecular therapy. The approval came with the requirement that a companion quantitative IHC test be used to assess whether the staining of the target oncoprotein reaches a quantitative threshold value of intensity that is associated with drug efficacy. Pathologists have been struggling since 1998 to overcome the challenges of IHC to produce an accurate and reproducible test with quantitative results. These challenges include inconsistent fixation and antigen retrieval, absent or nonstandardized internal and external controls, and reliance on subjective visual interpretation. Oncologists have recognized the shortcomings of the performance of quantitative IHC, and in 2007, the American Society of Clinical Oncology and the College of American Pathologists co-released guidelines for the performance of HER2 protein IHC testing in an attempt to produce a more accurate test to guide therapy [3].

This guideline document addresses the need for standardization throughout the total IHC testing process, reflecting the increased demands for quality in the performance of IHC. The College of American Pathologists has stated that all predictive cancer factor testing is deemed highly complex, requiring stringent quality standards and biannual surveys to assess adherence to these standards. The College of American Pathologists requires documentation of initial and biannual validation of a predictive IHC assay with 95% concordance, review of positive and negative controls, a mechanism to ensure daily that the test is working reliably, external proficiency testing, and ongoing quality assessment. At this time, the tools to measure and control for all the variables of IHC are limited, most notably in their inability to measure tissue fixation and adjust antigen retrieval appropriately.

Variability in tissue fixation remains a confounding factor for IHC, but automated retrieval and staining systems have largely eliminated variability in the staining portion of the process. Still, antibodies and reagents can degrade, resulting in variation that needs to be detected and controlled. The HercepTest (Dako, Glostrup, Denmark) provides breast cancer cell lines (MDA175, MDA-231, and SKBR3) with known HER2 oncoprotein expression. The use of standardized cell lines as multilevel positive controls provides a control target for assessing the accuracy of the HER2 staining process, although this process still does not control fully for variability of specimen tissue fixation.

Typically, these control results have been verified visually, a process that is subjective and not always reproducible. Digital imaging of IHC and its inherent numeric quantification of visual IHC results for controls and specimens provides an opportunity to leverage the statistical quality-control tools of the clinical laboratory in the performance of microscope-based IHC testing. The visual results of the control assays can be quantified, and the statistical tools commonly used in the clinical laboratory, such as Levy-Jennings charts and Westgard rules, can be used to expose errors in this step [4]. A Levy-Jennings chart is a graph on which quality control data are plotted to give a visual indication whether a laboratory test is
working well. On the x-axis the date and time or, more commonly, the number of the control run, are plotted. A mark is made indicating how far the actual result is from the mean expected value for the control. The distance from the mean is measured in SD. Lines cross the graph at the mean and at 1, 2, and sometimes 3 SD either side of the mean, making it easy to see how much the result varies from the target mean. This point is where informatics currently is most useful in improving the quality control of IHC procedures.

These statistics are made possible through the numeric data provided by digital images. A digital image file is a numeric representation of the color information present in different areas of an image. As an analogy, one can visualize a “color-by-number” picture in which areas are labeled with numbers corresponding to paint colors. When the appropriately numbered paints are placed in the corresponding areas of the image, the original artwork can be reproduced. The more colors used and the more detailed the image, the more accurately the reproduction imitates the original artwork. A digital image is similar to “color-by-number” in that it divides a visual image into a grid of pixels laid out on x- and y-axes. Numbers representing the color present in each area of the grid are stored in a bitmap file. The numeric code for digital colors varies but typically distinguishes among 256 hues, saturations, and shades of color in a typical Internet image, resulting in millions of colors that can be represented and reproduced accurately.

For quality control in IHC, digital images of IHC slides contain numeric data of the visual results of the process, and these data can be used to do calculations. Using computerized algorithms, one can measure numerically exactly how much of any of millions of colors is present in an IHC slide. This process is known as “colorimetry.” In addition, more complicated algorithms can use the data relating the location of the colors to analyze the shapes present and to localize the staining to the nucleus, cytoplasm, or cell membranes. This process is known as “morphometry.” Numerous commercial image-analysis systems now are available to digitize and perform these algorithms to extract the numeric data from IHC results.

As with any step in a test, it is necessary to control for variable conditions in acquiring and analyzing a digital image. These variables range from the brightness of the bulb to the gain of the photovoltaic chip used to sense the light coming from a tissue specimen. This is the point at which most of the quality differences among the various commercial products for digital image acquisition in pathology are seen. Systems that apply calibration steps for standardizations are preferable to those that do not. There also is significant variation in the largely proprietary algorithms used to measure and localize the colors in the specimen images in commercial image analysis systems. These variations greatly limit standardization among the systems, although the best of these systems allow optimization, validation, and standardization between runs.
Goals

The goals for applying image analysis to HER2 IHC controls included:

- Quantitation of HercepTest kit controls using image analysis
- Application of common laboratory control methods to the results
- Establishing quantifiable acceptable limits for control results to monitor variability between runs and to detect process errors

Project details

GML uses digital imaging to measure the results of the multilevel positive controls of the Dako HercepTest for HER2 oncoprotein expression in breast cancer specimens. The target mean and SD are established from valid control results for each level of control over a 6-month period to include most sources of testing variation. From these calculations, acceptable limits for control results are established. The group performed 84 runs of HER2 stains using the HercepTest kit over a 6-month period and analyzed the slides on the ACIS II Automated Cellular Imaging System (Dako North America, Inc., Carpinteria, California). The statistical analysis of results for the SKBR3-overexpressed HER2 protein control are provided in Box 1, and all 84 runs are displayed graphically in Fig. 2.

Measurements of the controls are plotted on a Levy-Jennings chart, and Westgard rules are applied to detect errors in the staining process and to see whether the results from the patient samples run concurrently with the controls are acceptable for release or, alternatively, need to be rerun. Control results exceeding 2 SD are repeated. If the repeated control is within 2 SD, the run is accepted. If a control result exceeds 3 SD or repeatedly exceeds 2 SD, an investigation is performed to identify the source of error.

The results of all three levels of controls for Her-2/neu IHC are tracked using software developed by the group (pictured in Fig. 3). This screenshot demonstrates a control failure as denoted by the arrow. At GML, the most common reason for a control failure is a focus failure in the robotic microscope caused by dust or fingerprints on the cover slip.

<table>
<thead>
<tr>
<th>Box 1. Quality control target values for SKBR3 control for HER2 immunohistochemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean:</strong> 3.12 units</td>
</tr>
<tr>
<td><strong>95% confidence interval for mean:</strong> 3.08–3.17</td>
</tr>
<tr>
<td><strong>SD:</strong> 0.2</td>
</tr>
<tr>
<td><strong>95% confidence interval for SD:</strong> 0.18–0.24</td>
</tr>
<tr>
<td><strong>2 SD range:</strong> 2.71–3.53</td>
</tr>
<tr>
<td><strong>Coefficient of variation:</strong> 6.50%</td>
</tr>
<tr>
<td><strong>Number of specimens:</strong> 84</td>
</tr>
</tbody>
</table>
**Conclusions**

HER2 protein expression is the first of an ever-growing number of biomarkers that will require complementary assays by laboratories. For IHC to serve this new purpose, stringent validation and quality control practices need to be employed to assure accuracy, precision, and clinical relevance of the results. Major challenges remain in standardizing IHC, most notably the inability to measure tissue fixation and adjust antigen retrieval techniques appropriately. To control better for variability in tissue fixation, there is a need for a quantifiable internal reference standard of fixation and retrieval for IHC for all target molecules [5]. Fortunately, the advent of molecular therapies has coincided with that of digital imaging, and informatics can provide tools to address some quality concerns in quantitative IHC using time-proven statistical methods of quality control from the clinical laboratory.

**Project III: results reporting**

**Background**

During the past 4 years GML has seen an explosion in electronic results reporting, including electronic medical record (EMR) systems, physician office practice management systems, patient Web portals, and electronic interfaces to external reporting repositories. Traditional paper reports distributed manually or by the postal service or faxing were easily lost, misplaced, or sometimes misfiled into the wrong patient’s record. Electronic distribution processes for results reporting eliminated some of the failure points for the delivery of results, but electronic transmission has introduced its own set of issues that now keep laboratories from delivering results consistently and in a timely fashion.
Issues with delivery are no longer tied to printers, fax machines, and human factors; instead they are tied to electronic algorithms, program coding, manipulation, extractions, and interfacing. The points of contact for an electronic result include the source system, processing agents, interface engines,
image repositories, and destination systems. Because the manual delivery of paper results has been removed from the process, and everything now is transmitted electronically, it is easy to assume that results always are delivered quickly, correctly, and to the appropriate place.

The first warning sign that results are not filing correctly are calls to the laboratory from physicians who are unable to find results in the patient’s EMR. If calls are being placed to various areas within the laboratory, and there is no tracking mechanism, these calls can seem like sporadic, isolated instances; however, the failure of patient’s results to reach the EMR or other destination repositories affects patient safety, treatment, and outcome. Consequently, providers and patients lose confidence in the laboratory’s ability to provide results in a consistent and timely manner and may experience panic, wondering if they are missing results critical to care delivery.

Goals

The defined goals for the GML results reporting quality project were to

- Ensure that all results from both anatomic pathology and clinical laboratory information systems reach their respective repositories
- Detect, correct, and resend any filing errors within 24 hours
- Reduce the number of filing errors to within a 6-sigma standard

Project details

GML electronically sends anatomic pathology results to both the Geisinger EMR system and GML’s outreach portal system. The electronic transmission originally was developed as an HL7 interface that sent textual data to both systems. This process worked, but it had weaknesses in providing rich, formatted pathology results that included supporting data such as images, graphs, and table formatting. To provide enhanced anatomic pathology reports, the group developed a second-generation interface to send report images to the patient data repositories as PDF documents. The PDF interface uses the source pathology system to generate the file for conversion to PDF, the PDF engine for creating the document, an image repository for storing the PDF document and creating a filing message to the patient repositories, the interface engine for transmitting the filing message, the patient repositories, and the network connectivity between the points. Although this change still involved an electronic interface, it increased the number of possible connection points of failure over its predecessor HL7 interface. Fig. 4 shows a block diagram of the possible connection failure points for the electronic transmission of pathology results.

The blocks highlighted in green in Fig. 4 represent the steps in transmitting pathology results to the patient repositories. A process log is created for each step and is monitored by various departments within the information technology department. On any given day, 400 to 600 results are sent through the process. Because there are approximately 19 logs to monitor,
the job of supporting and investigating the failures in any given day can be resource intensive. Missed errors and communication issues were problematic because of the cross-departmental makeup of the process. After approximately 6 months of supporting this process, the group realized that pathology results still were not reaching their respective destinations every time. A new approach that would guarantee the filing of pathology results correctly to the patient data repositories was needed. Because all pathology results started with the pathology system and ended with the patient repository, the group decided to use a “black box” theory. Fig. 5 represents the new concept in monitoring the transmission of pathology results.

Using the “black box” theory, one ignores all of the processes that are highlighted in blue in Fig. 5; the only areas of concern are the output and the destination. This concept allowed the group to spend more time investigating and correcting the issues instead of reviewing the logs looking for possible failures. A report scheduled to run at 1:00 AM each day captures all the cases signed out on the previous day. A second report is generated from the patient data repositories listing all the pathology cases that filed for the previous day. This report is shipped to the error-monitoring process at 2:00 AM each day. Both reports are loaded into a database and are compared with each other. The cases that do not match are kicked out and are considered failures. This list is presented on an error-monitoring Web page for the transcription services department to investigate, correct, and resend.
through the process. The reports then are verified manually to make sure they have filed within the 24-hour time period. If transcription services is unable to determine why a case did not file into the repository, that case is shipped to the pathology department information technology support team, which uses the logs that are created within the steps defined in Fig. 4 to determine if there was a failure within the process.

The database that stores the daily reports used in generating the error-monitoring process also stores the resolution to issues in the electronic filing process. Each case is displayed separately on the error-monitoring Web page (Fig. 6) with a link to enter the reason and resolution of the failure, along with the verification that the case filed appropriately into the patient data repository. The database then is used to generate quality assurance reports for use in the quality improvement process. Fig. 7 shows the pilot analysis used for the initial quality improvement process. Based on the rules for Pareto chart analysis, all issues are listed in descending order by count, and a cumulative percentage is attached to the chart. Based on cumulative percentages, issues that are above 80% should be assigned the highest priority for resolution, because the Pareto principle states that the majority (80%) of variation in processes can be explained by a minority (20%) of the causes [6]. The chart in Fig. 7 represents the findings after the 30-day pilot. There are six categories of issues in the above-80% rule: Incorrect

Fig. 5. Black box analysis diagram. The shaded area represents the processes that are being treated as one process for monitoring.
Patient Name, Patient Repository Error 8651, Patient Repository Error 8653, Patient Repository Processing Error, Invalid Requisition Received, and Incorrect Repository Order Number. Following is a brief explanation of each issue and associated findings.

1. Incorrect Patient Name: Most patient name issues are linked to the anatomic pathology department being the first instance of service within the GHS to see a patient after a name-change event in the patient’s life (eg, marriage). The new name is entered correctly into the pathology information system, but when the result is sent to the patient repository,
it does not match the medical record number, because the health system’s patient data repository has not been updated. The reverse of this scenario can happen when an update to the patient name in the GHS does not update the records in the pathology system, or when the name is changed after the pathology requisition is generated. The first part of this issue was addressed by updating the internal registration system, which in turn updates the pathology system, so that all systems have matching patient names and medical record numbers when the patient’s results are signed out. The other corrective action is to ensure that the clinics update the patient’s demographic history with the patient before ordering the test.

2. Patient Repository Errors 8651 and 8653: These errors are linked to ordering practices within the repository by physicians or their staff. The corrective action for resolving these issues was order entry training for the physicians and their staff.

3. Patient Repository Processing Error: These errors are results that have filed to the correct patient medical record but to the wrong place within that record. The patient’s EMR contains various pages where results file, including a laboratory page where all clinical and anatomic pathology results should file. The results in this instance have filed to a miscellaneous page, an image page, or another incorrect page. There are various reasons for this misfiling, and the corrective actions range from holding retraining clinics on ordering in the patient repository to a programming change in the repository to lock in laboratory report types so they file only within the laboratory results category.

4. Invalid Requisition Received: The pathology department receives requisitions for pathology orders that do not match the accompanying specimen. The pathology department accession the specimen with the order number as received on the requisition, and the results file back to the incorrectly placed order. Again the corrective action is training clinic personnel how to order correctly. The pathology department will be looking into the types of pathology orders available in the patient repository to see if the order process can be streamlined, making it easier to place an anatomic pathology order.

5. Incorrect Patient Repository Order Number: The anatomic pathology accessioning staff places orders from electronically generated requisitions. These requisitions contain barcode order numbers. There are times when the accessioning staff receives orders that are order history reprints instead of the actual current requisition. The order history shows all laboratory orders placed with multiple order numbers. The accessioning staff enters the order number manually and selects the wrong order number. Corrective actions include training the accessioning staff to inspect orders without barcodes and to use the correct order number and training ordering staff in clinics to submit only the electronically generated requisitions with a barcode order.
Conclusion

Since moving to the new error-monitoring process the group has completely eliminated calls from the physician staff about missing results. It also has realized the goal of filing all results within 24 hours on cases that have failed during the electronic transmission process. The next step was to bring the electronic transmission processes within 6-sigma control. At the end of the first year the total number of signed-out cases was 146,288, with 2024 failed reports or defects. These results calculate to a process sigma of 3.7. Since putting the improvements described previously into place, the group has moved its process sigma to 5.16, approaching the desired 6-sigma level of 3.4 defects per million opportunities.

Although there still is room for improvement, the group is realizing its goals by using the “black box” theory for monitoring results that have failed to file into the patient data repository. They have been able to provide a guaranteed turnaround time of all pathology results filing appropriately within 24 hours from time of sign-out, building back the providers’ confidence level in the laboratory’s reporting and speeding up the delivery of care to the patient community.

Final conclusions and future directions

These three projects represent the first informatics components of a total quality assurance/quality improvement program GML is developing in its anatomic pathology department. The group is developing informatics tools to track all steps in the flow of both specimens and associated data throughout the entire testing process from specimen collection through billing. Barcode labeling of specimen collection containers, specimen requisitions, tissue cassettes, and slides will play a major role in making it possible to track this flow and identify bottlenecks, look at process improvement, and set rules for specimen flow to flag cases falling outside the rules so they can be identified and dealt with quickly. Producing a reliable, readable barcode every time on tissue cassettes and slides is a current project. It also will be possible to identify quality markers at each step in the process flow that can be measured and tracked in pre-analytic, analytic, and postanalytic testing phases. The group will continue to use quantitative IHC data provided by digital imaging to expand statistical quality control, quality assurance, and quality improvement. It will expand the use of voice recognition software for dictation, templating as many gross and microscopic descriptions as possible in a synoptic format to decrease transcription errors and, it is hoped, producing reports that are better understood by providers. The group is moving toward the use of disease-oriented reports that incorporate clinical laboratory, radiologic, pharmacologic, and clinical history data. All of these parts of the total quality program require the development of new informatics tools that are largely unavailable today from pathology laboratory information system vendors.
References


Laboratory Clinical Effectiveness: Pathologists Improving Clinical Outcomes

Conrad Schuerch, MD\textsuperscript{a,}\textsuperscript{*}, Mark Selna, MD\textsuperscript{b}, Jay Jones, PhD\textsuperscript{a}

\textsuperscript{a}Division of Laboratory Medicine, Geisinger Health System, 100 North Academy Avenue, Danville, PA 17822-0131, USA

\textsuperscript{b}Division of Clinical Innovation, Geisinger Health System, 100 North Academy Avenue, Danville, PA 17822-0131, USA

Key points

The roles of the pathologist in laboratory clinical effectiveness are as follows:

2. Provide reliable laboratory measures for clinical outcomes.
3. Establish and use standardized laboratory database for outcomes research and health care improvement.
4. Participate in design of standardized practice algorithms, including laboratory test ordering, interpretation, and therapeutic recommendations.
5. Develop patient health information tools to support care.
6. Improve and extend laboratory reporting to include interpretive aids and clinical recommendations.
7. Use information systems tools to improve reliability and quality of care in all health care settings: hospital, clinic, home, and long-term care.
8. Provide clinical consultations when appropriate.

Background, framework, and definition of “laboratory clinical effectiveness”

One of the most important challenges to the American health care community is to improve overall medical performance. Often-quoted figures show

\textsuperscript{*} Corresponding author.

\textit{E-mail address: cschuerch@geisinger.edu} (C. Schuerch).
American health care to be less effective but more expensive than health care in other developed countries [1]. The public is well aware of highly published medical errors, and several important reports have brought forward objective data to show the great need for improvement in the quality of medical practice. The results of the Institute of Medicine’s report “To Err is Human” extrapolated a figure between 44,000 and 98,000 unnecessary hospital deaths per year in the United States [2]. The study of the Rand Corporation by McGlynn and colleagues [3], which was published in the *New England Journal of Medicine*, compared the care actually received by patients to the standard of practice in a broad selection of medical situations. The results showed that patients received all of the recommended care only approximately 55% of the time. Notably, most errors were errors of omission. For example, recommended laboratory testing or radiography was performed only 61.7% of the time. As a result of these and other studies, the Institute for Healthcare Improvement conducted its 100,000 Lives campaign, which aimed at eliminating in-hospital errors. The health care environment has changed, and new approaches to hospital and laboratory inspections are being used by the Joint Commission, College of American Pathologists, and state agencies that look at quality from new angles. The Centers for Medicare and Medicaid Services is experimenting with pay-for-performance approaches. Transparency of performance data and publication of quality and outcomes measurements for consumers and state regulators are becoming frequent.

Laboratory medicine is a highly regulated part of health care, and clinical laboratories, which have had a long history focusing on quality, continue to have significant error rates in the range of 0.012% to 0.6% [4]. There is no standard error definition or methodology in the various reports, which accounts for the wide range reported. Quality in anatomic pathology has been more in focus in recent years, and the error rates are reported in the range of 0.08% to 5.3% [5]. No standard definition of significant error is in use. Given the frequency of clinical laboratory tests and the high likelihood of a serious outcome from an anatomic pathology diagnostic error, these error rates make it imperative to improve laboratory performance.

Laboratory medicine, as a discipline, has redoubled its efforts at eliminating errors, more than half of which occur in the preanalytic phase in clinical pathology. Tightening the rules on specimen identification, barcoding, and automating multiple steps in the laboratory cycle are recent changes implemented to reduce laboratory errors. Many clinical laboratories are tracking their errors in various phases of the testing cycle and initiating process improvements. Anatomic pathology departments have been introducing new quality improvement approaches, including routine double reading of slides, synoptic reports, error tracking, and process improvements (eg, barcoding). These changes are important in moving the laboratory industry in an error-free direction. From an overall health care perspective, however, eliminating errors in laboratory services is likely to affect comparatively few patients. Although important, they have relatively little potential for impacting the
huge proportion of patients who receive less-than-optimal care, as reported in the Rand study. How can laboratory medicine effectively participate in the movement to improve health care delivery for all patients?

The laboratory plays a critical role in diagnosing and managing most acute and chronic illnesses and screening for disease. Laboratory interdigitates with nearly every specialty and at all levels of the patient encounter. As the medical community responds to the challenge of raising the standard of health care, it is intuitive that the laboratory, through its integral position in health care delivery, should be able to contribute to the improvement of overall health care performance. Ensuring that accurate laboratory results appear in a timely manner on a patient’s chart must be a baseline competency that is executed with near-perfect reliability. To further impact the health of the population, laboratorians must expand their domain of activity in ways in which they can add new value. Fortunately, new tools are available to laboratorians that allow them to impact the larger domain of health care delivery. This article explores the laboratory’s opportunities to join the urgent movement of improving patient- and population-based outcomes.

The total testing process or testing cycle as characterized by George Lundberg [6] involves all that happens from the time the physician makes an order until the result is reported back to the physician. Although the laboratory is responsible for clinically significant errors that need to be eliminated, a much larger number of significant errors occur in the provider’s ordering process, the interpretation of results, the physician’s follow-through actions, communications to the patient, and the patient’s response. The activities of laboratory clinical effectiveness, for the purpose of this article, are those that laboratorians can take outside of the traditional laboratory segment of the testing cycle to impact clinical outcomes (Fig. 1). This
article outlines various ways in which laboratories can integrate their discipline and skills into programs of health care systems to improve care delivery and clinical outcomes. Laboratorians can improve the quality, usefulness, and accessibility of the information they deliver for patients and busy health care providers. There are opportunities for pathologists to improve care for individual patients by serving as effective consultants and even caregivers. Other opportunities pertain to the laboratory’s role in assuring quality data as a foundation of health care decision making and quality measurement. Finally, there is a collaborative leadership role in effecting changes of provider and patient behavior, enforcing or reinforcing best practices, and measuring outcomes.

Several philosophical tenets underpin the advocacy of this direction. One is that standardization of practice and elimination of variation improve quality and outcomes. Another is that good health care involves teamwork among specialties. Cooperation at specialty interfaces, good communication, and sharing of knowledge are critical to success, as is the sharing of responsibility across the team for clinical outcomes. Pathologists must share accountability for the larger process, extend themselves outside their traditional boundaries, and engage in activities that improve clinical outcomes. Pathologists often have special knowledge to contribute in designing clinical pathways, and they have informatics and communication tools at their disposal that may be used to improve clinical performance.

Although physicians have good intent and ready access to the “best practice guidelines,” for various reasons patients often fail to receive care that is consistent with such guidelines. Reliability is the nemesis of modern medicine. The authors believe that further engagement at the clinical interface is the laboratorian’s most effective way of contributing to improved population health.

Decision support for test ordering and optimized regimens

Clinical pathologists traditionally have tried to impact ordering practices through educational efforts; however, the literature indicates limited effectiveness of educational activities [7,8]. Usually a focused campaign can change practice habits for a given test over a short period of time; however, even with ongoing reminders there is often regression to bad habits. Providing comparative use or cost data to clinicians sometimes can have impact, but in most cases, persuading physicians to change ordering habits through education is a discouraging exercise.

Motives to change ordering practices often have been to eliminate unnecessary testing and cost and substitute more effective test choices. There is concern that excessive test ordering increases cost by causing additional evaluation of erroneous or outlier results; therapeutic misadventures also could result. In recent years, concerns of overuse have been addressed partly by the federal government in Medicare payment practices, which have eliminated large test panels and require clinical diagnostic codes to justify ordering
practices. From a laboratory’s cost standpoint, the incremental cost of individual routine tests are small, and a small percentage reduction of automated testing (eg, 5% of complete blood count or routine chemistries) saves relatively little in operational cost [9]. The control of expensive esoteric testing and unnecessary transfusion expenses is low-hanging fruit, however. Stewardship of the medical dollar remains a responsibility of laboratorians.

Administrative measures are the most effective approach to change laboratory test use. Limiting the test menu, redesigning the request form, requiring documentation of the indications for test or transfusion, and requiring approval or review before sending out expensive tests can help rationalize the use of costly resources [7,10]. The implementation of reflex testing algorithms approved by a medical policy committee can improve and streamline the test-ordering process [11]. This approach can reduce unnecessary testing, eliminate return office visits, reduce phlebotomy procedures, and shorten the time to diagnosis. Pathologists engaged in reflex testing design and implementation undoubtedly can have a positive impact on patient care.

If laboratorians wish to respond to their moral imperative to improve health care, it is likely that they will get much greater effect expending effort on other aspects of improving clinical effectiveness than working at the difficult task of decreasing routine laboratory testing. In general, the delay or failure of a practitioner to order an appropriate laboratory test or to act upon it is a much more common danger to a patient than an adverse outcome after an unnecessary laboratory test. Following this line of thinking, laboratorians should turn their energies to ensuring that patients are tested and managed according to best practice guidelines, tests are interpreted correctly and acted upon, and information is communicated effectively to patients to elicit appropriate patient responses.

In recent years, the emergence of the electronic medical record (EMR) and practice management systems for physician offices has provided opportunities to present best practice guidelines, preferred test pick lists, specialized disease order sets, and reminders to clinicians. These tools are used to standardize practice, decrease variability, and improve reliability and quality across physician groups. A group must design the content of these systems, and pathologists, who supply a substantial proportion of the data, can contribute to designing and building tools for clinicians and patients.

The practice parameters, clinical protocols, and patient communication tools must be developed and collaboratively accepted as professional standards of practice. A physician group must prioritize messages that compete for physicians’ attention and projects that compete for information technology resources. Laboratory tests may be part of a set of health care reminders that appear as flags at the time of opening up a patient’s EMR. For example, in the Geisinger Health System, the EMR is configured to identify women who are due for a Pap smear and, at the time of the next encounter, to signal the need for the test and present preconfigured order/documentation templates. This reminder to the physician, in combination with
automated letters to patients and electronic tracking, has facilitated a great improvement in obtaining Pap smears. The systematic tracking by cytopathologists of patients who have abnormal Pap smears to ensure that clinical follow-up action has been taken is an example of laboratory engagement in clinical effectiveness. Other examples include reminders for periodic lipid screens and prostate-specific antigen in men aged 50. Other examples of ordering support are discussed in the section on improving health outcomes.

Pathologists as clinical consultants

Pathologists used to take pride in being the doctor’s doctor—able to consult on laboratory tests and their clinical significance and sometimes advise on therapy. With the explosion of medical knowledge and the rise of subspecialization, the clinical pathologist can no longer know everything in the field. He or she is most often limited in expertise to one or two subspecialty fields and has a broad general knowledge of laboratory medicine. Clinicians may be more knowledgeable about the clinical interpretation of laboratory tests in their own subspecialty. When the greatest expertise or the greatest availability for consultation belongs to the laboratorian, he or she must provide an active, consultative role and sometimes even recommend therapy. Clinical and laboratory guidelines and “best practice” are moving targets and ever more complicated. Medical practitioners are expected to remember and order in more detail than is humanly possible. Laboratorians can take responsibility for improving the lot of practitioners by making laboratory information easily understandable, linked to clinical interpretations, and actionable. Doing this effectively often requires advanced clinical information systems and collaboration with clinical specialists. This consultative and facilitative role along the laboratory–clinical interface will have increasing importance for clinical pathologists in the future.

Historically, the role of clinical consultant pathologists has been most exercised by blood bankers, who have the oversight of pretransfusion evaluations, selection of appropriate blood products, evaluation of transfusion reactions, enforcement of safety measures, and rationalization of blood product usage. In many institutions, therapeutic apheresis is administered under transfusion services, and the pathologist plays a direct clinical role. At Geisinger Medical Center, blood bankers have been activists beyond the bounds of the laboratory in recent years. The transfusion medicine director proposed and obtained support for transfusion guidelines from heavy user departments—the Transfusion Committee and the Medical Executive Committee. In ordering blood components, the indications for their usage must be selected from choices on an order form. This system is available in electronic and paper formats, and there is tight monitoring for compliance. As building clinical evidence in the literature demonstrates the detrimental effects of transfusion on patients in intensive care units and surgical venues [12], there is more urgency in bringing discipline to
transfusion practice. In addition to better blood component usage, the use of Factor 7 concentrate and Factor 9A has been subjected to guidelines at Geisinger, and the protocol requires involvement of a pathologist before use of these drugs. Further activism by the blood bank staff ultimately stimulated interest in a blood conservation program (bloodless medicine) that was established with its own support staff outside of the laboratory.

A visible advocate for pathologists expanding their roles as clinical consultants is Dr. Michael Laposata, director of laboratories at Massachusetts General Hospital. Using staff and residents, he developed round-the-clock coagulation consultation service and issued reports using a specially developed computer system (AMEDx) [13]. This service has been popular with clinicians, and he has grown a successful reference service in the Boston area. He also formalized consultative rounds and services at Massachusetts General Hospital in autoimmune disease, hemoglobinopathy/anemia, transfusion reactions, and complex cases of serum protein analysis.

At Geisinger, the Department of Laboratory Medicine provides a coagulation consultation service, including 24-hour availability by one of four pathologists between two hospitals with reports generated on the same AMEDx system used by Dr. Laposata. The laboratory has supported (as part of its point-of-care testing role) the installation of six thromboelastograph analyzers for the operating rooms and intensive care units in two hospitals. The same pathologists provide consultative interpretation of the thromboelastograph histograms that have been networked and can be seen from remote sites or home. Pathologists have developed a standard interpretive algorithm for evaluation of microvascular bleeding and use it to advise clinicians on the appropriate blood components in complex cases. This approach is a significant step in the rationalization of blood component usage in difficult operative bleeding cases. The effects of several concurrent quality and use improvement activities have been impressive: a 3.7% absolute decrease in red blood cell transfusions, a 14.7% decrease in total transfused blood products, a 20% decrease in transfusion reactions, and a 46.3% decrease in discarded blood products over a 3-year period in which operative cases have increased by 14.3%.

Another service the Geisinger hematopathologists provide is an anemia consultation aimed at helping triage anemia patients for the primary care physician. In this case, several tubes of blood are obtained and, depending on initial blood indices and the results of other laboratory tests available in the laboratory computer system, a selection of reflex tests is ordered. The blood smear, clinical history, and laboratory results are reviewed by a pathologist, and an integrated report is issued that can include treatment or triage recommendations.

Pathologists’ engagement in clinical decision making can be valuable in several critical areas, such as transfusion medicine and coagulation, which provide help in the efficient interpretation and management of several medical problems. However, pathologists can have much broader impact on routine clinical practice by integrating the knowledge of good clinical/laboratory
practice into the EMR and other clinical support systems for use by large
groups of physicians.

**Improving the delivery of laboratory information**

Many aspects of laboratory clinical effectiveness—appropriate ordering,
effective reporting, prioritization and interpretation of laboratory data, de-
cision support, and motivation of patients to do the right thing for them-
selves—can be supported and improved with electronic tools. The
challenge is that each system needs to be bought, designed, interfaced,
and often customized. On the reporting side, there are enhancements,
such as graphic data for laboratory tests, that are facilitated by electronic
data systems. Reports with age- or therapy-specific reference ranges, some-
times depicted graphically, can make laboratory data more immediately
comprehensible. Written interpretations for immunoprotein analyses, West-
ern blot, electrophoresis, platelet aggregation, and molecular and cytoge-
etic tests are required of laboratorians today because most clinicians lack
the knowledge to interpret these data correctly.

Enhanced reports with photographs, color graphics, anatomic drawings,
multiple laboratory markers, and risk stratification charts have been devel-
oped and marketed successfully in commercial settings for specialty surgical
pathology, urology, and cardiac risk. All of these reports facilitate communi-
cation of information to clinicians and often assist clinicians in communicat-
ing with patients. Additional patient materials, even patient-specific reports,
have been developed by laboratories to supplement official laboratory re-
ports. To the extent that laboratories move in this direction, they improve
their clinical effectiveness. A futuristic concept not yet developed would in-
volve national or commercial Web-based tools for improving the presenta-
tion of laboratory data for clinicians and patients. In this concept, hospital
laboratories of any size could subscribe to a service that would enable
them to package anatomic and clinical pathologic results with relevant clin-
ical research data in an information-rich disease and patient-specific report.

**Electronic patient access to laboratory**

Enhanced reporting can be taken to a higher level when patients have ac-
cess to their own medical record online as they do in several health care or-
ganizations. A survey of Geisinger patients ranked access to laboratory data
the number one for using “MyGeisinger,” which is the patients’ Internet-
based view of their own medical records. The nearly 100,000 MyGeisinger
users also can schedule appointments, request pharmacy refills, email their
physicians, and pay bills online. Laboratory test explanations initially
were written by pathologists but later were supplemented by commercially
available laboratory explanations for lay persons [14]. Cardiac risk can be
accessed via linkage to a commercial Web site (EBSCO Publishing, Ipswich,
Massachusetts). Patient Internet access systems have the potential to present graphically enhanced reports. The future of this technology may allow us to prompt patients to get their recommended laboratory tests at a draw station or to enter their home point-of-care laboratory results into the EMR online. The use of electronic connectivity with patients will become increasingly important in the future and enable a more efficient and clinically effective mode of health care delivery.

**Clinical effectiveness reporting**

Health care delivery is a team endeavor. No matter how well laboratories provide traditional laboratory service, how well they use computers to prompt clinicians to order the right tests, or how accessible and direct their reports, even beyond providing patient-directed reports, laboratorians can make a more profound contribution to improvement in health care. Laboratorians as team members can provide data and rationale to measure quality and outcomes, change models of health care delivery, and impact the structural problems inherent in the current health care delivery system.

For more than 15 years, Geisinger’s laboratories have developed a standardized laboratory database for a system that includes three hospitals and 38 distant clinic sites. Together, the laboratories perform more than 5 million billable tests per year—nearly 70% from outpatient settings. Key tests that are used in “pay for performance” (eg, low-density lipoprotein [LDL] cholesterol, hemoglobin A1c [HbA1c], microalbumin) were proactively centralized, validated, and held consistent for 10 years for peer comparison studies to prepare a clinical archive that would be homogeneous and suitable for population studies. Although quality laboratory data archives exist in several integrated health care systems, the “breadth” (ie, homogeneity of test methods in a geographically diverse, multiple site enterprise) and “depth” (use of “best of breed” test methods, validated by data mining) of the archive in the Geisinger Health System are unusual. In addition to being a tool for improving quality in the distributed laboratory, the database is a primary tool for measuring clinical outcomes for many diseases. It allows one to compare clinical performance of one part of the organization to another and allows tracking of clinical effectiveness over time. There is even greater potential value when laboratory data are mined in conjunction with and correlated with the EMR-derived clinical data.

Geisinger’s Center for Health Research is performing research on chronic diseases in primary care clinic settings with experimental and control clinics. New clinical care models, which may yield improved outcomes and be implemented across the health care system, are under investigation. The solid laboratory infrastructure is essential for credible research. The laboratory’s roles in this health services research range from providing convenient phlebotomy and quality point-of-care or home testing to providing outcomes data.
Laboratory test method selection for clinical effectiveness

In creating an interpretable database, few and preferably one method per test should be selected and established as a primary method for future population and database review and for reporting clinical effectiveness. Operational, financial, and medical and scientific considerations impact this decision. Across multiple facilities in a given health care system, based on scalability in different-sized laboratory operations, different methods and analytic platforms are chosen. Primary selected methods must harmonize with secondary point-of-care methods, because the aggregate of reported results becomes relatively indistinguishable in the EMR and clinical repository. Administrative and unit cost savings that result from group purchasing are but one benefit of such consolidation and standardization. This inherent efficiency also encompasses information technology that has fewer interfaces, uses test codes and standardized reference ranges, and associates similar test results with one another.

Establishing a database in the laboratory information systems to support clinical effectiveness

No matter where a standardized method is performed, for it to be valuable in establishing and monitoring clinical effectiveness, it must be connected to an enterprise database and extractable for future studies. Most laboratory information systems (LIS) contain software applications to extract numerical fields and patient demographics as a simple, comma-separated value file (.csv file). If fields exist for standardized textual comments within the tabular database structure of the LIS, this information also may be easily extractable. Conversely, free text comment fields are difficult to extract; once extracted, they are hard to compare across multiple patients and patient encounters. As large regional health care systems aggregate to form a regional health care information organization, it is important to recognize that LIS information almost assuredly will be heterogeneous because of laboratory site differences. With the use of Logical Observations Identifiers Names and Codes (LOINC) assignment to different laboratory methods in these regional health care information organizations, individual test results in diverse LIS theoretically can be normalized and merged.

Data mining

Databases that are established with standardized analytic methods in a well-structured and extractable format are of great value in evidence-based medicine and studies of clinical effectiveness [15]. Concise information may be extracted from such databases with simple and readily available “data mining” tools. Fig. 2 shows a data mining pathway used at Geisinger for
extracting numerical data from an LIS with Microsoft Access and Microsoft Excel. In the LIS, a software application does the initial extraction of a .csv file and exports this file (also known as an ASCII or flat file) to a secure network drive location. It is important that these data, once removed from the LIS, be stored and manipulated in a secure environment and not removed to nonsecure PCs or storage devices, which may result in patient privacy being compromised. This .csv file is imported into Microsoft Access to sort required numerical data and patient demographics. The data table allows pasting of Microsoft Access rank ordered data arrays to the companion Microsoft cell program, which is used for presenting large database files. This exported graphic may be a population or a correlation histogram, along with statistical parameters to allow for interpretation, based on demographics or other test information originally sorted in MS Access. Data also may be pasted into a statistical program (SASS or EP Evaluator) for statistical analysis or graphic presentation. Simple accessible data mining tools will lead to peer comparisons of population data and drive standardization of analytic methods, not only within a given health care enterprise but also possibly between peer health care enterprises.

A good example of how this is already occurring is seen in the desire of the National Kidney Foundation to make serum-creatinine measurements traceable to a National Institute of Standards and Technology isotope dilution mass spectroscopy reference method. Because the enzymatic serum-creatinine method is most closely traceable to this isotope dilution mass spectroscopy method, it is emerging as a preferred analytic method that can yield a reliably calculated estimated glomerular filtration rate (which then can be validated in a population distribution). Clinical effectiveness
measured parametrically in given disease populations, which drives analytic method selection, is a relatively new phenomenon.

Validation of laboratory parameters applied to clinical effectiveness

The measurement of effectiveness in diagnosing and treating many disease states lies largely with the accuracy and predictive value of a few parametric laboratory tests. Clinical decision points should be validated based on population distributions of a given laboratory parameter in the population being studied. Fig. 3 demonstrates the value of data mining in population validation of estimated glomerular filtration rate, the reporting of which is currently required in several states in the United States [16,17]. If clinical decision points for estimated glomerular filtration rate are defined at 60 or 30 mL/min, the population percentile targeted for intervention can be determined. The factors that contribute to this parametric value, including the test method, patient demographics (ie, age, gender, ethnicity), and means of calculation, should be defined clearly.

Laboratory clinical effectiveness test parameters in disease states

The current number of parametric laboratory values used in evaluating clinical effectiveness for populations is relatively few. In this fairly new field, a few laboratory tests have been verified for use in population studies, and

![Graph showing estimated glomerular filtration rate distribution](image)

Fig. 3. Example of Excel graphic export of data-mined estimated glomerular filtration rate.
those population studies concern relatively few diseases. One of the most prevalent disease states currently being addressed by clinical effectiveness studies is diabetes. Clinical effectiveness measures for diabetes currently focus on four parameters, which are discussed in the following sections:

1. HbA1c
2. urine microalbumin
3. LDL cholesterol
4. glucose as measured in tight glycemic control protocols

**Hemoglobin A1c testing**

HbA1c is ubiquitous as a surrogate measure for describing clinical outcome and monitoring effectiveness of diabetes treatment. This test has been standardized over the years to include the specific HbA1c subfraction of glycosolated hemoglobin [18]. Because the landmark Diabetes Complications and Control Trial demonstrated a significant avoidance of disease complications by patients achieving target values of HbA1c [19], focused efforts at standardizing HbA1c measurements among chromatographic, electrophoretic, and enzyme immunoassay and methods have been underway. The glycosolated hemoglobin standardization program has gone so far as to certify different methods and even lot numbers within methods to glycosolated hemoglobin standardization program accuracy guidelines [20]. Although these efforts have improved disparity among various HbA1c test methods, biases between different methods (eg, immunoassay methods and high performance liquid chromatography methods) still exist. Health care systems should strive to implement a single standard method that is judged accurate for clinical effectiveness monitoring purposes and remain with that method for serial determinations within individual patients and within populations. Method-related shift of a few tenths in a population of patients who have diabetes may invalidate clinical effectiveness measures that also hope to achieve such shifts. If relative clinical effectiveness among different health care systems is to be judged under current and proposed pay-for-performance programs, a given HbA1c method should be verified to understand bias to other methods to avoid artifactual differences in performance.

**Microalbumin testing**

The measurement of trace amounts of albumin in urine is used to determine risk of progressive renal disease in patients who have diabetes [21,22]. Unlike HbA1c, few, if any, serious efforts have been launched to standardize methods for its measurement. Some laboratories also measure only albumin, whereas others measure albumin as a ratio to urine creatinine, an endogenous internal standard. These methods, although encouraged by various
clinical guidelines, have not been codified [23]. Because of differences in hydration state and urine concentration during diurnal variation, differences may be exhibited within individual patients and patient populations unless compensated by urine creatinine ratioing. Validation studies of random urine albumin versus albumin-creatinine ratio show a high rate of discordance (Fig. 4), presumably because of the state of urine concentration.

**Low-density lipoprotein cholesterol testing**

Although not directly related to glucose absorption or metabolic abnormalities, LDL cholesterol testing is frequently included in outcome studies of populations who have diabetes and may be used as a pay-for-performance parameter for clinical effectiveness. Populations are most often initially screened for risk assessment by a calculated LDL cholesterol level that includes embedded risk parameters of direct high-density lipoprotein cholesterol and triglycerides. Subsequent LDL tests that track effectiveness of individual therapy (eg, with statin drugs) should be accomplished with directly measured direct LDL cholesterol [24]. The data from the two approaches should not be merged because there are significant differences in the value of calculated and direct LDL [16,25].

As LDL cholesterol testing continues to grow in importance as a clinical effectiveness measure, it will likely evolve as a highly standardized measure similar to HbA1c, with a standardization body similar to glycosolated hemoglobin standardization program [26].

![Fig. 4. Correlation of albumin (ALBU) to albumin/creatinine ratio (ACR) from x-y paired data exported from MS Excel to EP Evaluator software. Med Dec Pt, medical decision point.](image-url)
Use of point-of-care glucose in monitoring tight glycemic control

An important example of the use of laboratory tests in judging clinical effectiveness is protocol-based bedside glucose monitoring [27]. Implemented by adult intensive care units at Geisinger in 2003, this tight glycemic control protocol has concurrently yielded greatly improved rates of readmission and wound infection, days to discharge, and vent weaning time. In this case, the need to establish a tight glycemic control protocol [28] (e.g., the Portland protocol or Yale protocol) supersedes the importance of method precision to support the protocol. Measuring whole blood glucose levels quickly and with appropriate quality control by critical care nursing staff at the point of care using bedside monitoring devices has supplanted central laboratory plasma glucose testing. Tight glycemic control protocols prescribe monitoring frequency protocols for insulin dosage and target ranges for glucose that avoid even moderate hyperglycemia ($<140\, \text{mg/dL}$).

It is important when selecting bedside glucose monitoring devices that there be means to integrate numerous glucose results into information systems [29]. In selecting point-of-care devices with connectivity, one must be aware that most of them use glucose oxidase glucose methods with reflectance or biosensor measuring technology. Although accessible, convenient, and
inexpensive for monitoring, these methods have slightly higher coefficients of variation and may be prone to some interference with medications that do not affect central laboratory methods. The bedside glucose methods have peer-compared proficiency testing and some quality control peer-comparison programs; hence, it is not critical to reference them back to a primary method within the health care enterprise (hexokinase in the central laboratory).

At Geisinger, the laboratory provides oversight to clinical care testing at all sites. In the intensive care units, the laboratory provides standardized glucose meters that could download their data into the LIS through a Telcor workstation. The laboratory validates the point-of-care results by parallel testing with the central laboratory on a subset of patients and provides monthly reports to clinicians pertaining to use and average glucose levels. Fig. 5 shows trends that occurred at Geisinger Medical Center after implementation of the tight glycemic control protocol. As numbers of bedside glucose tests increased in adult intensive care units, mean glucose values dropped substantially.

An integrated approach to clinical effectiveness using informatics tools

Since 2005, Geisinger has used a formalized approach to evaluate, redesign, deploy, and monitor its “high reliability” care models. This approach, called ProvenCare, has been rigorously applied to selected types of procedural care (eg, coronary artery bypass, cataract removal, primary hip replacement) and chronic disease care (eg, diabetes, coronary artery disease). Inherent to the ProvenCare performance improvement process is selecting performance measures that are clinically meaningful, operationally actionable, and broadly available. These “component” measures, aggregated as disease-specific “all-or-none composite” measures, are tracked for each primary treating provider. When aggregated for each clinic site, they serve as the basis for provider incentive compensation. For the disease-specific measure sets, laboratory test ordering and laboratory test values comprise a significant proportion of the required data. For example, the composite measures for diabetes (implemented in early 2006) (Table 1) include ordering specific tests (HbA1C, lipids) and achieving desired therapeutic objectives (HbA1C <7; LDL <100). Primary care physicians are rewarded based on the percentage of their clinic site’s patients who have diabetes who achieved 100% achievement/compliance on all components.

To ensure effectiveness of clinics in satisfying all of the component measures for as many patients as possible, several systems were put in place, including disease-specific registries, scheduled patient reminder letters, nurse “rooming protocols and documentation templates,” and alert-linked order sets for physicians. Notably, most routine orders are preconfigured in the EMR and provisionally executed by the nurses. For Geisinger’s approximately 20,000 adult patients who have diabetes, the sustainable effect of
this redesigned process has been dramatic (Fig. 6). The approach to improving care using “all-or-none” achievement of composite measures aims directly at fixing the unreliability of health care delivery and it is off to a promising start at Geisinger.

**Interdisciplinary collaboration for clinical effectiveness**

One successful interdisciplinary collaboration to improve patient outcomes pertains to glucose monitoring of patients in intensive care units. Another example from our experience at Geisinger includes the establishment of patient-friendly coagulation clinics run by the pharmacy department, with

![Graph showing percent of diabetic patients with different number of components received or achieved](image-url)

**Table 1**

<table>
<thead>
<tr>
<th>Component measures</th>
<th>Quality standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>HgbA1C measurement(^a)</td>
<td>Every 6 months</td>
</tr>
<tr>
<td>HgbA1C control(^a)</td>
<td>&lt;7</td>
</tr>
<tr>
<td>LDL measurement(^a)</td>
<td>Yearly</td>
</tr>
<tr>
<td>LDL control(^a)</td>
<td>&lt;100</td>
</tr>
<tr>
<td>Blood pressure control(^a)</td>
<td>&lt;130/80</td>
</tr>
<tr>
<td>Retinal examination</td>
<td>Yearly</td>
</tr>
<tr>
<td>Urine (protein) examination(^a)</td>
<td>Yearly</td>
</tr>
<tr>
<td>Foot examination</td>
<td>Yearly</td>
</tr>
<tr>
<td>Influenza immunization(^a)</td>
<td>Once</td>
</tr>
<tr>
<td>Pneumococcal immunization(^a)</td>
<td></td>
</tr>
<tr>
<td>Smoking status(^a)</td>
<td>Nonsmoker</td>
</tr>
<tr>
<td>Use of ACE/ARB for microalbuminuria/DM nephropathy</td>
<td>Yes</td>
</tr>
<tr>
<td>Use of ACE/ARB for hypertension</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Composite: Patients who received/achieved ALL (*) components within the prior 12 months.

*Abbreviations:* ACE, angiotensin converting enzyme; ARB, angiotensin receptor blockers.

\(^a\) A second immunization for persons who have diabetes and are older than age 65 is indicated only if the first immunization was before age 65 and was 5 or more years ago.
testing support by the laboratory and medical guidelines developed by clinical departments. The establishment of coagulation clinics has become popular since the late 1990s in centers around the country and has proved to be a great patient and physician satisfier while yielding more reliable quality care [30,31]. Nine different coagulation clinics, which were established in clinics over a wide geographic region and within the hospitals, currently serve more than 20,000 patients. The hospital laboratories guaranteed a 30-minute turn-around time for prothrombin time with international normalized ratio (PTINR), and the clinics were equipped with i-STAT instruments for rapid point-of-care PTINR analysis. The pharmacists adjusted dosages of coumadin according to a preapproved algorithm. There has been improved patient compliance with having their blood tested compared with what occurred in the usual clinic model. In the first year for a patient population of 1754 using coagulation clinics, the population PTINR improved. Compared with historical controls, an estimated 53 potential strokes were avoided.

Who pays for laboratory clinical effectiveness?

The benefits of improving clinical effectiveness ultimately accrue to the patient as improved outcome. Insurers and payers for service benefit if total health care expense is decreased. Providers at risk in capitated or diagnosis-related group models benefit similarly. Because the benefit of a laboratory improving its clinical effectiveness falls outside the laboratory in general, there is usually no new revenue stream associated with these activities and they are considered “value added” to a laboratory’s product.

Exceptions to the foregoing statement are few. Clinical pathology consultations receive professional compensation but often at inadequate levels to justify the activity on an economic basis. Reports that translate laboratory data to effective clinical information—well packaged for busy clinicians—have had proven commercial value (eg, integrated reports for cardiac risk, cancer prognostic reports). This area of opportunity exists in anatomic and clinical pathologic areas. Unfortunately, the packaging of multiple tests in specialty panels sometimes has been structured to promote overuse to benefit the providing laboratory at the expense of the payer and patient.

Organization of laboratory clinical effectiveness

Laboratory clinical effectiveness has no standard organizational structure. Most efforts are made by laboratory directors or individuals with motivation to improve efficiency and effectiveness and decrease cost. Laboratory user demands for guidance and help in interpretation of laboratory tests motivate these efforts, as do regulatory requirements for transfusion committee oversight (State Department of Health, Joint Commission) and review of use changes annually (Medicare). Currently the urgent demand for improved outcomes and effectiveness is pushing health care organizations to invest in
information technologies and hire specialists in clinical outcomes and medical management. The owners of the information technology tools and the medical management staff are positioned to help implement and monitor the programs of laboratory clinical effectiveness. They—pathologists, pharmacists, and others—are necessary collaborators to improve patient care. Wise organizations establish committees and organizational structures that design and make decisions about the EMR and clinical practice guidelines.

To keep the laboratory doctoral staff focused on laboratory use and clinical effectiveness, Geisinger Medical Laboratory established a Laboratory Utilization Committee in the late 1990s. Renamed the Laboratory Clinical Effectiveness Committee in 2005, the Committee has encouraged, monitored, and provided leadership and advice on several projects (Box 1). Members have included a medical director of the Geisinger Health Plan, physician leader of the Department of Clinical Effectiveness, and many laboratorians. The committee has sometimes been impeded in keeping momentum behind projects because of competing organizational demands. A recent focus on clinical effectiveness espoused in Geisinger’s 5-year program, “Striving for Perfection,” has stimulated much more interest in projects to improve the execution of clinical best practices, however, including laboratory ordering and follow-up on results.

Innumerable opportunities exist for improving clinical effectiveness. A simple one is illustrated by a probe of the Geisinger pharmacy information system. Of 7319 consecutive patients on digoxin in a 4-month period, the average time since the last drug level report was 2.4 years. In 36% of patients the time was more than 3 years and in 14% the time was more than 5 years. Digoxin toxicity causes or complicates a certain number of admissions in our institution each year. A guideline (eg, one digoxin level per year) could be established by expert agreement and built into the EMR. Either physicians or patients could be issued reminders to ensure that ordered tests occurred. Follow-up systems could be created for individuals who are not caught in the first intervention. Outcomes could be measured in decreased

<table>
<thead>
<tr>
<th>Box 1. Laboratory clinical effectiveness committee activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Monitor blood product use</td>
</tr>
<tr>
<td>• Manage sendout tests</td>
</tr>
<tr>
<td>• Provide laboratory test use reports</td>
</tr>
<tr>
<td>• Outline clinical guidelines, standing orders</td>
</tr>
<tr>
<td>• Provide graphic reports for prostate-specific antigen and free prostate-specific antigen</td>
</tr>
<tr>
<td>• Guide clinical consultations in coagulation, anemia</td>
</tr>
<tr>
<td>• Oversee laboratory–pharmacy collaborations</td>
</tr>
<tr>
<td>• Provide diabetes program support</td>
</tr>
</tbody>
</table>
admission and complications. This approach could be taken for several therapeutic drugs, including antiepileptics.

The future of laboratory clinical effectiveness

The future of clinical practice is often said to entail more patient involvement in medical decisions, more self-care, and more out-of-hospital care. It is likely that laboratories will be involved in supporting and tracking results from patients’ self-testing, which may involve a new paradigm of collaboration among laboratory, patient, and clinician. Another new paradigm is that of self-directed testing—patients purchasing laboratory tests to look for disease, understand their health risks, or support their own wellness programs. Information and clinical advice may be demanded of laboratories that provide this testing. Unfortunately, it is possible to envision advertising driving overuse of these tests, but there will be opportunities for turning the population toward greater health.

From a population management standpoint or a payer standpoint, it is advantageous to identify individuals at high risk for disease so that they can be screened earlier and more often and treated promptly. Genetic testing likely will be a component of future risk assessment. The message to patients likely will be complicated, and the laboratory or health system may need to develop new resources for communication. Population genetic testing is a research area for Geisinger’s Weis Research Center. A repository for blood cells and serum is being established; collection and specimen processing involves clinic staff and the clinical laboratory working in collaboration. New genetic markers for chronic disease identified in similar research settings may enable early identification of individuals at risk for preventive measures or periodic screening, and genetic findings may lead to new therapeutics. The laboratory can be a part of a team that drives improved clinical effectiveness, this time through basic science, outcomes research, test development, and the application of new solutions to selected populations.

In the future, laboratorians and clinicians in an integrated delivery system with an EMR will build systems that automatically order population screening tests on behalf of physicians and prompt physicians to order appropriate tests in specific clinical conditions. Selected reflex testing protocols can be automated based on test results and clinical parameters in the EMR. Finally, recommended clinical actions can, in some cases, be incorporated in the laboratory report and be available at the time of clinical decision making. As designers of the artificial intelligence incorporated in electronic systems, laboratorians and clinical colleagues together will serve as “information managers” to improve the consistency and quality of patient care.

Summary

The concept of “laboratory clinical effectiveness” implies a change in emphasis for laboratorians from improving the quality of laboratory services to
improving population outcomes. This emphasis demands engagement and collaboration outside of the laboratory and acceptance of a shared accountability for medical outcomes. Although pathologists have used their energies outside of the traditional laboratory testing cycle, involvement in this broad arena is of increased importance as medicine strives to improve its performance. The measurement of improved patient outcomes involves first building a standardized and validated laboratory database for clinical effectiveness. Second, pathologists can serve as clinical consultants in certain disciplines in which the pathologist may have the greatest available expertise. Most importantly, however, as members of multidisciplinary teams and through the use of clinical computer systems, pathologists can influence the consistency and reliability of test ordering, and assure that appropriate care is given based on clinical laboratory information. Pathologists can play a part in the future of clinical medicine by designing systems in which populations consistently receive care according to best standards of practice.

Further readings


References

Measuring Quality in Anatomic Pathology

Stephen S. Raab, MD*, Dana Marie Grzybicki, MD, PhD

Department of Pathology, University of Colorado Denver, 12605 East 16th Avenue, Aurora, CO 80045, USA

Quality

In anatomic pathology, quality is the product (ie, the diagnostic information) or service that meets the requirements of a wide number of individuals or groups, including patients, clinicians, pathologists, pathologist extenders, organizations, and regulatory agencies [1,2]. Quality assurance (QA) is a system of control activities that promotes the higher level functioning of specific processes within the anatomic pathology laboratory. These processes usually are composed of multiple steps or subprocesses, each of which is subject to quality control (QC) activities to ensure that these steps or subprocesses meet acceptable parameters [1,2]. QC activities often entail the development of standards of acceptable performance. In an ideal system, QA/QC activities are used to guide quality improvement (QI) activities that attempt to improve the quality of a specific product or service beyond its current status [1,2]. Well-designed QA programs by themselves may lead to improved product or service quality [1].

In industry, standardization is a key component of quality, and considerable effort is expended on standardizing work processes and materials [1, 3–5]. One rationale for standardization is that there is a best method of performing work, which should be adopted to produce work of optimal quality. In medicine, standardization is highly variable within and across areas of care delivery, including anatomic pathology laboratories [6–9]. For example, one anatomic pathology laboratory may have well-documented, standardized protocols for the gross examination of specific specimen types, but
the actual work processes of gross examination may not be standardized, so individuals within that laboratory grossly examine specimens differently. A second laboratory may have standardized policies and work processes. Different approaches do not necessarily imply that one anatomic pathology laboratory provides higher quality services compared to another, because strengths may be seen in alternative approaches [10]. The lack of standardization implies that an optimal level of product or service is not being delivered across the population of all anatomic pathology laboratories, however.

In anatomic pathology, standards for acceptable performance have been adopted for several quality metrics, and other quality metrics are closely examined despite the lack of performance standards [1,11–15]. Examples of traditional measures of anatomic pathology quality are diagnostic accuracy, customer (clinician) satisfaction, and specimen turn-around time. This article focuses mainly on diagnostic accuracy, recognizing that measuring any quality metric is complex and demanding [11]. Laboratories use several methods to standardize quality metrics, and the College of American Pathologists, the American Society for Clinical Pathology, and the Papanicolaou Society of Cytopathology have taken leadership roles in developing these quality metrics [12]. Laboratories traditionally have created quality standards by benchmarking current practice or using expert opinion [12] rather than developed standards based on evidence-based outcome assessments. The separation between laboratories and clinical services and the lack of health information technology tools are two reasons why linking clinical outcome to testing services is difficult.

No diagnostic or screening test has 100% sensitivity and 100% specificity [1,2]. Much to the chagrin of health care personnel involved in diagnostic testing and screening, specific processes in testing services often are held to a high level of performance, which leads to the impression that laboratory services must achieve near perfection. Pap test screening and interpretation is one example in which laboratories are held to exceptionally high levels of accountability. In reality, measures of diagnostic accuracy depend on laboratory and clinical performance, and for some diagnostic or screening tests, the literature reports wide accuracy variations, which partly reflects a lack of clinical or laboratory process standardization. Consequently, although laboratory QA/QC activities evaluate metrics of diagnostic accuracy, laboratories can only address and control domain-specific subprocesses that contribute to the overall test accuracy [10]. Measuring the diagnostic accuracy of individual subprocesses is difficult because multiple factors from preceding testing steps bias assessments [10,16] (see later discussion). Although we intuitively know what constitutes good and poor diagnostic or screening test performance, a lack of standards limits our ability to evaluate true test performance for individual clinical and laboratory services.

Because testing within the anatomic laboratory is composed of numerous steps [10,16], laboratories may use QA/QC activities to target the quality of any individual step to improve the quality of the overall laboratory process.
QA/QC activities may be applied to processes that occur in any one of the individual anatomic pathology laboratory components or testing areas, such as an accessioning area, gross room, frozen section area, histology or cytology laboratory, transcription area, laboratory information service area, autopsy suite, ancillary testing laboratory, fine-needle aspiration clinic, or pathologist’s domain [10]. For example, histologic slide staining quality influences the overall laboratory quality metrics of turn-around time and diagnostic accuracy, because poorly stained sections may necessitate restaining (which affects specimen turn-around time) and interfere with diagnostic interpretation (affecting diagnostic accuracy). Histology section quality may be monitored through QC activities that track the quality of hematoxylin and eosin–stained slides.

Medical error

In 1999, the Institute of Medicine published its famous report “To Err is Human: Building a Safer Health System,” which estimated that between 44,000 and 98,000 patients die each year as a result of medical error. Many more patients suffer from morbidities associated with medical error [17]. The Institute of Medicine estimated that medical errors result in total costs (including the expense of additional care, lost income and household productivity, and disability) of between $17 billion and $29 billion in United States hospitals annually [17]. Medical errors that occur in ambulatory care settings also result in considerable costs.

The Institute of Medicine defined a medical error as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim, although others have advocated alternative definitions of error [17]. The determination of whether a medical error occurred is performed separately from patient outcome assessment or root cause analysis, although medical errors may be subclassified after outcome assessment or root cause analysis. When examining outcomes, errors may be classified as “no harm,” “near miss,” or “harm” events, and some have labeled errors associated with harm as “adverse events” [17]. Most medical errors are not associated with poor patient outcomes. Root cause analysis shows that diagnostic testing or screening errors may be caused by process failures outside of the laboratory domain and do not necessarily imply that a pathologist misinterpreted a specimen [10]. Unfortunately, many health care personnel tend to protect their domain, which limits the ability to determine how multiple factors compound error in diagnostic testing and screening.

Based on the Institute of Medicine definition of error, anatomic pathology diagnostic testing or screening error is the failure to diagnose correctly the disease process in a patient [18]. Pathologists have long known about failures in diagnostic testing and screening processes, and QA/QC activities document these failures using diagnostic accuracy metrics [1,2]. False-negative and false-positive diagnoses are examples of errors of interpretation.
Some patient safety researchers also consider that indeterminate (eg, atypical or suspicious) diagnoses are forms of error, because this category of diagnosis does not accurately convey if a patient does or does not have disease [11]. Unfortunately, the term “error” often connotes a value judgment that is highly negative and pejorative for many health care professionals. The fact that pathologists use indeterminate diagnoses reflects the current state of practice, especially for less-than-optimal specimens, because indeterminate diagnoses provide probability estimates that limit errors of greater severity (eg, false-negative or false-positive diagnoses) [19]. In the ideal diagnostic testing or screening scenario, the frequency of less-than-optimal specimens would be considerably lower, thereby limiting the use of indeterminate diagnoses.

In some pathology circles, there has been an avoidance of the term “error” partly because of the negative connotations and because of the fear of the consequences of reporting. Alternatively used terms include discrepancy, defect, flaw, deficiency, and variance.

Diagnostic testing and screening error taxonomies

Practitioners and researchers use different taxonomies to classify diagnostic testing and screening error, and these taxonomies have different strengths and weaknesses. A few of these taxonomies are discussed in these sections.

Error classified by testing phase

Diagnostic testing and screening are composed of multiple steps that comprise the total testing cycle, as defined by Lundberg [20]. The five phases of the total testing cycle in the Lundberg model are as follows:

1. Pre–pre-analytic: Determining to test and choosing the specific test
2. Pre-analytic: Procuring and transporting the specimen
3. Analytic: Processing and interpreting the specimen
4. Post-analytic: Reporting the test results
5. Post–post-analytic: Acting upon the test results

The anatomic pathology (or analytic) processes, including the diagnostic interpretation, play a critical but only partial role in the overall care of the patient. QA/QC activities may be difficult to undertake in diagnostic testing and screening because the processes in diagnostic testing and screening cross multiple domains of ownership [1]. Because tests may be viewed as parcels of information that are transformed at each step, problems (or errors) in any step may result in compromised data. In some instances (partly because of a lack of standardization), it is impossible to determine when (or if) information has been compromised, and a probabilistic approach may be used to determine the level of error after a diagnostic or screening test. Some clinicians use Bayesian methods to determine the post-analytic probabilities of
disease based on the pre-analytic probabilities of disease and the test results (expressed probabilistically).

Stroobants and colleagues [21] estimated that errors occurred in approximately 20% of clinical pathology tests and that the proportion of error within the pre–pre-analytic, pre-analytic, analytic, post-analytic, and post–post-analytic testing phases was 12.0%, 2.0%, 0.2%, 2.2%, and 5%, respectively. Estimates of the error proportions in anatomic pathology testing and screening cycle have not been published, although most likely they are at least the same as those reported by Stroobants and colleagues [21]. Most anatomic pathology testing errors likely occur in the pre–pre-analytic and the post–post-analytic phases. The lack of standardization of test ordering, choice, and follow-up clearly are major sources of error. The pre–pre-analytic evaluation of patients with a lung nodule suspicious for cancer is an example, because the choice of test depends on the entry point (rather than evidence based analysis of patient outcomes of different combinations of diagnostic pulmonary tests) of patients [22]. Primary care health care providers are more likely to order sputum cytology, pulmonologists are more likely to order bronchoscopy, and surgeons are more likely to opt for wedge excision for patients with the same clinical findings [22]. These test choices affect the types of specimens received by anatomic pathology laboratories (and the types of errors that occur within them). Anatomic pathology-based studies of diagnostic testing error generally have not evaluated the correctness of test choice for many patient scenarios.

Most anatomic pathology QA/QC activities evaluate the errors that occur in the analytic phase of testing, although some QA/QC activities report specimen collection and other pre-analytic or post–post-analytic failures. Within the anatomic pathology laboratory, errors may be classified as

- Accessioning (eg, specimen identification switch with a second patient, wrong physician entered, wrong patient information entered)
- Grossing (eg, specimen not properly sampled, tissue blocks mislabeled, tissue blocks too thick, tissue not properly fixed)
- Histology/processing (eg, sections cut too thickly, tissue not properly processed, floater picked up in waterbath, tissue not adequately stained)
- Transcription (eg, misspelling, incorrect format of report, case assigned to the incorrect pathologist, information omitted from report)
- Ancillary testing (eg, wrong ancillary test performed, failure of ancillary test, ancillary test reported to the wrong patient)
- Sign-out (eg, pathologist misinterpretation, disease process incorrectly described, report dictated in a confusing manner)
- Reporting (eg, wrong report issued, report sent to the wrong physician, report lost)

Some anatomic pathology laboratories separate errors into the categories of laboratory-related and pathologist-related errors [1]. In this scheme,
laboratory-related errors are secondary to failures in all the analytic testing phases, excluding the interpretation phase (or the grossing phase if pathologists perform the gross examination). This separation is helpful in laboratories with separate domains of hospital and pathologist control. At the University of Pittsburgh Medical Center, laboratory information system personnel created a dictionary of anatomic pathology laboratory-related errors, which contains more than 200 distinct deficiencies attributable to different laboratory sections.

Active and passive methods and timing of error detection

Anatomic pathology error detection may be separated into active and passive methods. Active methods of error detection collect more errors than passive methods. In a preliminary study of active error detection in anatomic pathology specimen accessioning areas in multiple hospitals, Grzybicki [23] used an observational method of error detection in which errors were recorded by a third party. Grzybicki [23] reported that this active method detected errors in the accessioning phase in more than 70% of specimens, whereas passively detected errors, measured through anatomic pathology QA logs, occurred in less than 3% of specimens. The active error detection method markedly increased the frequency of specific accessioning error types, such as accessioning cases (1) without two patient identifiers, (2) without sufficient patient information on the requisition form, and (3) with improperly matched requisition form and specimen container [23]. Most of these errors presumably resulted in no patient harm, although one can imagine that extremely rare instances of these errors could lead to catastrophic consequences. The accessioning personnel generally performed “work-arounds” to correct these problems because they occurred so frequently and were never fixed [1]. In a second active error detection study of the accessioning phase, Zarbo and D’Angelo [16] reported that 28% of anatomic pathology specimens were defective. These data indicated that the frequency of anatomic pathology error may be far higher than that assessed by Stroobants and colleagues [21] in the clinical laboratory and that defects are present in most anatomic pathology specimens. This finding should not be surprising given the large number of steps in anatomic pathology testing.

Error detection also may be classified as prospective or retrospective [11]. Prospective methods generally are active in nature and are aimed at limiting the number of adverse events. An example is a secondary slide review of all cases diagnosed as malignant by a first pathologist. The number of errors detected in this fashion is relatively unknown, because some laboratories do not track these errors and some pathologists do not even consider these events as true errors. In conventionally accepted patient safety terms, these errors are classified as near-miss events, because they are corrected before having an affect on patient outcome.
Retrospective error detection methods may be active or passive. Active retrospective error detection is helpful when these error data are used to guide QI activities.

*Traditional anatomic pathology error detection methods*

Error detection methods in anatomic pathology are presented in Box 1. These error detection methods are examples of secondary case review and often serve as the basis for much of the assessment of error frequency reported in the pathology literature [11]. Errors detected by these methods mainly occur in the interpretation process [24], except in the correlation reviews in which errors are more likely to occur as a result of pre-analytic failures. The proportion of anatomic pathology cases with an interpretation error depends on the specific review method. Much of the pathology error literature is based on single institutional studies, which contain biases caused by the lack of pre-analytic and post-analytic process standardization [25,26].

Comparing institutional error proportions using secondary case review methods is challenging because of the lack of QA/QC process standardization. Vrbin and colleagues [27] studied the level of standardization in cytologic-histologic correlation. Vrbin and colleagues sent a survey to 162 American laboratories requesting copies of the materials they used in their cytologic-histologic correlation process. They developed a checklist (derived from the College of American Pathologists Laboratory Accreditation Cytology-Pathology Checklist) to classify the minimum expected (15) and additional data points that laboratories collected when they performed cytologic-histologic correlation. No laboratories collected the exact same data, and 17.3% of laboratories did not record any data on forms, logs, or tally sheets. The mean number of minimum expected and additional variables recorded

<table>
<thead>
<tr>
<th>Box 1. Error detection methods in anatomic pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frozen section–permanent section correlation</td>
</tr>
<tr>
<td>Cytologic-histologic correlation</td>
</tr>
<tr>
<td>Fine-needle aspiration immediate and final diagnosis correlation</td>
</tr>
<tr>
<td>Conference case review (eg, tumor board, unknown conference, daily difficult case conference, subspecialty conference)</td>
</tr>
<tr>
<td>Random or pseudorandom case review</td>
</tr>
<tr>
<td>Focused review of specific case types</td>
</tr>
<tr>
<td>Second opinion or consultation (eg, retrospective or prospective, internal or external), performed outside of conference case review</td>
</tr>
<tr>
<td>Amended report review</td>
</tr>
<tr>
<td>Autopsy case review</td>
</tr>
<tr>
<td>Clinician driven review</td>
</tr>
</tbody>
</table>
on forms was 6.5 and 8.7, respectively. Vrbin and colleagues concluded that laboratories recorded data from the cytologic-histologic correlation process in several ways, which indicated the lack of standardization of the data collection process. Raab and colleagues [18] confirmed the absence of standardized cytologic-histologic error detection processes in four laboratories that shared correlation data for QI activities.

**Error detection by amended report review**

Zarbo and colleagues [11] classified errors into four categories based on the review of amended reports:

1. Interpretation errors, further subclassified as false-negative diagnoses (undercalls), false-positive diagnoses (overcalls), and misclassifications (not altering primary or secondary diagnostic classifications; eg, small-cell carcinoma versus non–small-cell carcinoma of the lung)
2. Specimen errors, including lost specimens, inadequate or nonrepresentative specimens, and improperly handled specimens (eg, specimens not processed appropriately in the analytic testing phase or specimens that did not receive the proper ancillary test in the analytic phase)
3. Identification errors, including incorrectly identified patient material (eg, wrong patient), tissues (eg, stomach versus colon), or anatomic location (eg, right versus left lung)
4. Reporting errors, including erroneous or missing nondiagnostic information, dictation or typing errors, and report formatting errors

Zarbo and colleagues [11] used the term “defect” instead of “error” and based this classification scheme on the type of change made in an amended report and included changes in the (1) primary diagnostic characteristics, (2) secondary diagnostic characteristics, (3) diagnostic reclassification, (4) patient or specimen identification, (5) specimen characteristics (eg, resampling the specimen leading to a changed diagnosis), and (6) other edits not including the changes made in categories 1 to 5. Essentially, this classification scheme is a form of root cause analysis (other methods of root cause analysis are presented later) because it links errors to specific process failures.

An example shows the use and overlap of these different error classification schemes. As mentioned in the preceding paragraph, cytologic-histologic correlation is a method of error detection that may be performed retrospectively or prospectively with conjoint cytologic and histologic specimens (eg, Pap test and cervical biopsy procured during colposcopy or bronchial brush and bronchial biopsy procured during bronchoscopy). In the frame of errors classified by phase of total testing cycle (including the phases within the anatomic pathology laboratory), cytologic-histologic correlation detects errors mainly in the pre-analytic (ie, specimen sampling) and analytic (ie, specimen processing and interpretation) testing phases. Using the amended report error classification scheme advocated by Zarbo and colleagues [11],
cytologic-histologic correlation errors may be subclassified as secondary to
defective specimens or defective interpretations. Cytologic-histologic corre-
lation may be performed actively but generally is performed retrospectively.

Diagnostic disagreement and error

Physician interpretation is a complex, cognitive task, and physicians un-
dergo long periods of training and evaluation before becoming credentialed
to make these interpretations. Many areas of medicine exhibit high levels of
variability in physician judgments [28–30], with pathology being no excep-
tion. In internal medicine, for example, physicians may disagree regarding
the interpretation of chest radiographs, EKGs, and the signs and symptoms
of patients. These disagreements reflect not only the complexity of human
physiology and disease processes but also the complexity of medical decision
making. Some of these disagreements reflect erroneous decisions caused by
human cognitive error and occur commonly in all fields of medicine.

In most scenarios, diagnostic disagreements are not associated with
harmful outcomes. For example, two pathologists may differently subclas-
sify the same high-grade sarcoma. These differences may not affect patient
prognosis or clinical management; however, the patient either has one ma-
lignancy or the other, and one diagnosis fails to describe the disease process
in the patient. In some situations, we currently may lack the knowledge base
to make the distinction among tumor types, and research may be needed to
address this issue further. Some pathologists have argued that most interpre-
tive disagreements are not true errors and only interpretive disagreements
associated with harm are errors. The weakness of this argument lies in its
necessary association of error with an adverse event, when only a small sub-
set of the total population of errors results in a clinical adverse event. Many
patient safety scientists argue that focusing on adverse events is important,
because detecting and preventing these errors reduce patient harm. A chal-
lenge in this stage of patient safety investigation is determining which
anatomic pathology errors are associated with harm so that they may be tar-
geted for further study.

Root cause analysis

There are several well-accepted methods of performing root cause analy-
sis. The amended report method presented by Zarbo and colleagues [11] is
useful because it specifically examines error causes from a diagnostic testing
and screening perspective. This method ignores specific causes of error (eg,
system error), however, and does not delve deeply into other causes (eg,
causes of cognitive error). Patient, specimen, provider, and system factors
cause diagnostic testing and screening errors [1]. A root cause analysis
method that has been applied effectively to anatomic pathology is the Eind-
hoven Classification Model for the Medical Event Reporting System for
<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latent errors</td>
<td>Errors that result from underlying system failures</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Technical: Physical items, such as equipment, physical installations, software, materials, labels, and forms</em></td>
</tr>
<tr>
<td>TEX</td>
<td>External</td>
<td>Failures beyond the control of the investigating organization</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Knowledge-based behaviors</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The inability of an individual to apply his or her existing knowledge to a novel situation</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>(continued on next page)</em></td>
</tr>
</tbody>
</table>
Transfusion Medicine [31–33]. This method focuses on three domains: technical (equipment, forms, and software), organizational (procedures, policies and protocols), and human (knowledge based, rule based, and skill based). These three domains are useful in classifying contributing factors and organizing causes of error. They allow for error investigation to focus on system factors rather than entirely on human factors. Limitations in our current QC measures of diagnostic testing and screening error are the excessive focus on interpretation error and the inability to determine contributing factors to diagnostic misinterpretation. Table 1 shows a more detailed list of error causes in this classification model [31–33].

Raab and colleagues [34] performed root cause analysis by examining the overall and individual QA/QC diagnostic test performance data and determining causes of error based on less-than-optimal test performance. The limitation in this method was that root cause analysis was performed a long time after the error occurred, and aspects of the testing process could not be evaluated in retrospect [31–33]. A benefit of studying overall test

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRQ</td>
<td>Qualifications</td>
<td>The incorrect fit between an individual’s qualification, training, or education and a particular task</td>
</tr>
<tr>
<td>HRC</td>
<td>Coordination</td>
<td>A lack of task coordination within a health care team in an organization</td>
</tr>
<tr>
<td>HRV</td>
<td>Verification</td>
<td>The incorrect or incomplete assessment of a situation, including related conditions of the patient/donor and materials to be used before beginning the task</td>
</tr>
<tr>
<td>HRI</td>
<td>Intervention</td>
<td>Failures that result from faulty task planning and execution, which would be selecting the wrong rule or protocol (planning) or executing the protocol incorrectly (execution)</td>
</tr>
<tr>
<td>HRM</td>
<td>Monitoring</td>
<td>Failures that result from monitoring of process or patient status</td>
</tr>
<tr>
<td>HSS</td>
<td>Slip</td>
<td>Failures in the performance of highly developed skills</td>
</tr>
<tr>
<td>HST</td>
<td>Tripping</td>
<td>Failures in whole-body movement; errors are often referred to as “slipping, tripping, or falling”</td>
</tr>
<tr>
<td>PRF</td>
<td>Patient-related factors</td>
<td>Failures related to patient/donor characteristics or actions that are beyond the control of the health care professional team and influence treatment</td>
</tr>
</tbody>
</table>

| Other factors | Unclassifiable | Failures that cannot be classified in any of the current categories |
performance data was that system issues could be studied better [34]. A problem in studying diagnostic testing and screening error is that test failures may not be known for considerable time periods until a repeat test or definitive procedure shows a different disease process than was originally diagnosed or clinical suspicion evokes case re-evaluation.

Raab and colleagues [34] coded errors using the Eindhoven Classification Model and created a table displaying major factors that contributed to error. A few examples of organizational error causes in thyroid gland fine-needle aspiration are as follows [34]:

1. OM (organizational management priorities): Radiology division processes patients too quickly to allow for proper fine-needle aspiration performance. The cytology schedule is too busy to implement cytologists in the performance of immediate interpretations. The hospital does not mandate the sending of patients with palpable lesions to more experienced aspirators.

2. OP (organizational protocols and procedures): There is a lack of standardization for pathology sign-out procedures, diagnostic criteria for category use, and radiology procedures.

3. OC (organizational culture): The system focuses on punishment and not improvement. There is no system for formal root cause analysis.

4. OK (organizational transfer of knowledge): New cytologists or less experienced cytologists are not taught in a rigorous fashion.

Raab and colleagues [34] also evaluated the causes of error in individual cases and constructed causal trees that represented the factors, activities, and decisions that possibly lead to errors. Although specific for individual cases, false-negative and false-positive diagnoses generally were related to multiple causes, including lack of immediate interpretation services, patient-related factors, and overall high workloads. These error causes compounded other knowledge-based error causes that led to the procurement of less-than-optimal samples or misinterpretation of these samples.

Most thyroid gland fine-needle aspiration errors are detected by the cyto logic-histologic correlation process, which traditionally has used a root cause analytic method that classifies error into the two categories of sampling and interpretation. The studies by Raab and colleagues [34] and Nodit and colleagues [35] indicated that this binary classification is highly useful in studying general process failures but is less informative in determining latent error causes in the interplay between sampling and interpretation error. Simply classifying cyto logic-histologic correlation errors as either interpretation or sampling generally does not provide sufficient information that may be used for system QI. The understanding of the root cause of error requires more detailed analysis of factors extending beyond anatomic pathology laboratories.

Raab and colleagues [36] reported that the range of pairwise interobserver kappa values for pathologists who assessed error cause of pulmonary
cytologic-histologic correlation cases was −0.154 to 1.0. These data indicated that the traditional method of root cause analysis for cytologic-histologic correlation was handled differently in different hospitals and that some pathologist pairs exhibited marked disagreement in assessing if the error cause was sampling or interpretation. Raab and colleagues [37] found that this disagreement generally was based on variable assessments of specimen interpretability, defined as the combination of sample quality (eg, representativeness, obscuring factors) and the amount of tumor present. Pathologists disagreed on what constituted a good sample and the amount of tumor necessary to render a malignant diagnosis.

Raab and colleagues proposed using a new QC method for cytologic-histologic correlation, termed the “No Blame Box” (Fig. 1). The amount of tumor is depicted vertically, increasing from “no tumor” at the top of the box to “abundant tumor” at the bottom. A specimen that contains many cancer cells would be graded on the lower portion of the vertical axis and a specimen that contains only rare, questionable cancer cells would be graded higher on the vertical axis. The specimen quality is depicted horizontally, increasing from a poor quality specimen at the left to an excellent quality specimen at the right. Specimen quality relates to pre-analytic and analytic processes. The four squares of the No Blame Box divide specimens into combinations of cancer/no cancer and good quality/poor quality. Completing the No Blame Box generally illustrates that the error cause is multifactorial.

![Fig. 1. “No Blame Box” for root cause analysis.](image-url)
Summary

The study of anatomic pathology quality and patient safety is ongoing, and currently, much effort involves defining and measuring error. Data that link error to patient outcome are critical for developing QI initiatives targeting errors that cause patient harm. Using methods of root cause analysis beyond those traditionally used in cytologic-histologic correlation also assists in developing error reduction and QI plans.

References

Malpractice in Dermatopathology—Principles, Risk Mitigation, and Opportunities for Improved Care for the Histologic Diagnosis of Melanoma and Pigmented Lesions

Whitney A. High, MD*

Department of Dermatology, University of Colorado Health Sciences Center, P.O. Box 6510, Mail Stop F703, Aurora, CO 80045, USA

The misdiagnosis of skin cancer is a substantial source of medical malpractice litigation. A recent analysis of claims submitted to a single large national malpractice carrier revealed that 8.6% of all claims against pathologists and 14.2% of claims against dermatologists involved the terms “skin cancer” and/or “melanoma” [1]. Studies in the pathology literature have characterized further the medicolegal risk associated with the histologic misdiagnosis of melanoma [1–3].

In a 2003 study, Troxel [1] reported on 362 pathology-related claims submitted to The Doctor’s Company for the years of 1995 to 2001; 46 (13%) involved the misdiagnosis of melanoma [1]. Melanoma ranked second only to breast cancer as the most common source of pathology-related claims. Of the 46 claims pertaining to melanoma, 32 were for false-negative (“missed”) diagnoses. In two later papers, using the same claims database but for the years 1998 through 2003, Troxel [2,3] identified 335 pathology-based malpractice cases. A false-negative diagnosis of melanoma was the single most common cause of a malpractice claim, accounting for 42 (13%) of the 335 cases. A recent examination of a report from the Physician’s Insurance Association of America found melanoma represented a common cause of “medical misadventure” among dermatologists, many of whom interpret their own histologic specimens [4].
Multiple lines of evidence converge to suggest that melanoma represents a substantial source of risk within the dermatology and dermatopathology [1–5]. The substantial medicolegal risk associated with melanoma is intellectually satisfying because (1) melanocytic lesions are perhaps the most common neoplasm in humans, (2) melanoma is arguably one of the most lethal of all malignancies, (3) the distinction between benign and malignant melanocytic lesions is often vexing, even for the most skilled dermatopathologists, and (4) melanoma is a disease that afflicts both the young and old, causing more years of lost life than any other malignancy, except leukemia [6,7].

With the increasing incidence and prevalence of melanoma in the United States [8], and the inherent medicolegal risks involved in the histologic interpretation of pigmented lesions, a review of relevant medicolegal principles, suggestions to optimize care, and strategies to mitigate medicolegal risk are needed. This article seeks to provide general pathologists, dermatologists, and dermatopathologists alike with

1. An overview of the basics principles of medical malpractice litigation
2. A review of the essentials of reporting and the importance of expert consultation for melanoma and pigmented lesions
3. Twelve suggestions to improve quality of care and to reduce the medicolegal risk associated with melanoma and pigmented lesions

**Basic tenets and principles of malpractice litigation**

Malpractice is adjudicated using state law. Even for cases filed in federal court, any substantive decisions are based on the law of the state in which the events transpired [9]. Substantial overlap exists among the states; only a generalized overview is provided here, and variations may be encountered.

*Elements of medical malpractice*

Medical malpractice requires the assertion and proof of six requisite elements: (1) duty, (2) standard of care, (3) breach of duty, (4) cause-in-fact, (5) proximate cause, and (6) damages. These elements must be established individually for each case. Failure to prove any element, typically to a standard of “more likely than not” or a greater than 50% likelihood, results in victory for the defense. From the outset, the plaintiff must submit evidence substantiating each element to establish a *prima facie* case. With the establishment of a *prima facie* case, the burden of proof shifts, and the defendant must produce contradictory evidence to support his/her position. To provide a better understanding of the legal process, each of these elements is discussed in greater detail.

*Duty*

Duty arises through the establishment of the physician–patient relationship. In a clinical setting, duty typically is realized through taking a history
and performing physical examination. In the realm of pathology, it occurs when tissue is accessioned and examined, and a report is issued. Duty arising from “curbside consultations” is a quagmire of uncertainty [10]. Differing court decisions on the matter exist [11,12]. For this reason, avoidance of “curbside consultation” is often advocated [13]. When “curbside consultation” does transpire, some record of the event, even if only handwritten notes, is recommended to prevent the consulting physician’s use of the advice from becoming the only documentation of the event [13,14].

In dermatopathology, another less formal means of consultation often used is an intradepartmental second opinion. Experts suggest that intradepartmental second opinions are of great utility in providing optimal care in difficult cases [15,16], but the rendering of an intradepartmental consultation should be well documented, including the names and qualifications of those involved, the date consensus was achieved, and the title of any conference at which the material was reviewed. Failure to do so renders the reasonable and prudent approach of not one but multiple physicians undocumented [14].

Finally, pathologists and dermatopathologists, like their clinical brethren, are vicariously liable for the actions of laboratory employees under their control, so long as the employee is acting within the scope of his/her duties [17]. Similarly, physicians working in education are responsible for the dutiful but negligent actions of students, residents, and fellows following the legal theory of respondeat superior.

Standard of care

The term “standard of care” refers to a minimum level of care owed to a patient by a physician who has incurred a duty to that patient. The term “reasonable and prudent physician” often appears in reference to the standard of care, but in truth standard of care often is a matter assessed by custom. A physician must act with the minimum knowledge and skill common to members within his/her profession in good standing. Importantly, this standard is not simply what a majority of providers would do, because courts routinely recognize the concept of a “respectable minority” [18]. Accordingly, a jury is not bound to accept only the majority approach but may opt instead for a different standard that is promulgated by a reasonable and respectable minority. Clearly, an incident occurring 5 years earlier must be judged using the state of knowledge at the time the care was rendered, not using medical knowledge or additional testing options realized since the diagnosis was rendered. Historically, the standard of care was subject to geographic variation, but with improved communications, overnight couriers, and standardized medical education, the modern trend is towards a uniform national standard [19,20].

The standard of care also may vary with respect to the training of the involved provider. For example, a general pathologist might be held to different standard than a board-certified dermatopathologist, but this standard must be established independently in each case through expert witnesses
and evidence; it is not presumed. Although little direct research into the matter is available, Trotter and Bruecks [21] demonstrated only a 1.4% incidence of major discrepancies between general pathologists and dermatopathologists. Such evidence could be used by a plaintiff to suggest the standard of care between a dermatopathologist and a general pathologist in electing to interpret skin lesions is much the same. The defense, however, should counter that the existence of an entirely separate training process and board-examination, jointly administered by the specialties of dermatology and pathology and recognized by the American Board of Medical Specialties, suggests a more limited standard of care for general pathologists and also suggests the value of consultation in difficult cases.

Regardless of how the standard of care might be litigated, when a physician voluntarily represents himself/herself as an expert in the field (eg, an internist serving as a medical director at a laser center or a general pathologist who interprets skin specimens when dermatopathology expertise is available in the institution or area), the plaintiff will contended that the physician has committed to providing a standard of care commensurate with this representation [22]. Also, if a general pathologist attempts to diagnose a specialized problem that normally would call for a referral to a specialist, he/she may be held to the standard of such a specialist [23].

With the obvious exception of licensing requirements and restrictions on the scope of practice (more common for nonphysician providers), the standard of care is not legislated. Instead, it is established through the testimony of “expert witnesses.” An expert witness must possess expertise equivalent to that of the accused, the subject matter must be of substantial overlap between disciplines, or the expert must have reasonable knowledge of the scope of training of the defendant. For example, an orthopedic surgeon might not be able to establish the standard of care for a podiatrist, unless the orthopedic surgeon had detailed knowledge regarding the training afforded within podiatric medicine or unless the subject was of substantial overlap between disciplines [24]. Published practice guidelines, consensus conferences, or peer-reviewed articles of particular impact or merit may be used to establish the standard of care. Regulations or societal recommendations pertaining to the operation of a pathology laboratory might be germane to the standard of care for lawsuits based on procedural or processing errors.

Rarely, a suit may be brought for which expert testimony regarding the standard of care is unnecessary. Such cases are based on the legal theory of res ipsa loquitur (“the thing speaks for itself”) [25]. For example, if a pathologist allowed a dog into the grossing room, and the dog ate a large tissue sample, res ipsa loquitur may apply—there is no reasonable way this error could occur without negligence, and experts are not required.

**Breach**

Breach is care rendered below the established standard. To discourage baseless suits, many states require an affidavit from a licensed physician at
the outset stating that on review of the facts there is a reasonable likelihood
of negligence existing.

**Cause-in-fact**

Justice requires that any injury claimed must be the result of the alleged
incidence of malpractice. Cause-in-fact often is called “but-for” causation:
“but-for” the provider’s action, the injury would not have transpired. In
dermatopathology, if the rendering of an erroneously benign histologic di-
agnosis led to morbidity and/or mortality, the interpreting physician may
well be the cause-in-fact of the injury. Similarly, if the interpreting physician
issued an erroneously malignant histologic diagnosis, and the patient re-
ceived unnecessary surgery or detrimental medical care subsequent to the di-
agnosis, the party in error may be the cause-in-fact of that damage, even
though the damage was caused by the actual treating physicians who acted
in good faith based on the incorrect diagnosis.

**Proximate cause**

Proximate cause can be a difficult principle to distinguish from cause-
in-fact; but on occasion the two elements are analyzed separately by
a court. Proximate cause is akin to the “foreseeability” existing between
a negligent act and the resultant consequence [26]. Attacks on proximate
cause are made by asserting that, given the quality of the negligent act,
the outcome was so unforeseen that it is unjust to assign liability. Exam-
ple 1 illustrates the principle of foreseeability in establishing proximate
cause.

**Example 1.** A 53-year-old man was diagnosed mistakenly as having meta-
static melanoma and was informed of a poor prognosis. Based on this erro-
neous information, assuming he had “no time left anyway,” he immediately
abandoned his family, travels to Thailand, spent all his money on alcohol,
gambling, and brothels, and contracted HIV through unprotected sex.
The error was discovered 3 days later, and all reasonable attempts to contact
him were used, but the patient sues the pathologist for all his losses and for
acquisition of HIV.

In this situation, the defendant should contest proximate cause. Al-
though the misdiagnosis might be the cause-in-fact of the patient’s complete
decompensation, the ramifications were so wholly unforeseen (in addition
to being immoral and possibly illegal) that it would be unjust to assign
liability for the consequences; proximate cause is interrupted. Courts,
however, do not always require that the exact manner of injury be foreseen;
that some negligence has occurred and that some adverse outcome which
could have been foreseen has transpired may be enough to send a case to
trial [27]. For example, if the patient in Example 1 instead had undergone
additional diagnostic procedures at his own expense, a claim for damages
pertaining to these unnecessary procedures might be judged highly
foreseeable in light of the malpractice committed, and the case would be settled or sent to trial.

**Damages**

For a malpractice action to foment, harm must have transpired. The societal purpose of malpractice litigation is to make the victim whole, to the extent that money can do so, through an award of compensatory damages. Compensatory damages may include special damages (tangible items) or general damages (intangible items). Special damages include medical costs (past, current, or future) and/or lost wages (past, current, or future) caused by the injury. General damages include pain, suffering, and loss of enjoyment or consortium.

Punitive damages, which are designed to punish the offender for “egregious and willful acts” and to deter similar actions in the future, are not covered by insurance policies; hence, they are infrequently sought or awarded in malpractice actions against individuals. Given the widely acknowledged difficulty in evaluating pigmented lesions, it is difficult to foresee any reasonable circumstances that would yield punitive damages against an individual physician for the microscopic misdiagnosis of melanoma or its mimics.

Both the severity of the injury and the age and occupation of the patient affect the total damages sought. Furthermore, calculation methods may be contested fiercely, even when fault is agreed upon. For example, the plaintiff’s speculation regarding potential career advancement will be opposed vigorously by the defense because of the multiplying effect it would have on damages. Because of inherent advantages in the size of the award, plaintiffs often request that general damages be calculated on a per diem basis (eg, $1000/day for pain and suffering for X days), whereas defendants instead seek a lump-sum payment.

Without damages, there can be no malpractice. For example, when a skin cancer is missed during an initial examination but is discovered just 1 week later by a different physician, how should damages be valued? Did the cancer progress significantly during those 7 days? Has the effectiveness of any potential treatment diminished over such a short time span? Damages in this scenario are so difficult to ascertain that they threaten the survival of the claim irrespective of the egregiousness of the error.

**Timeliness of the complaint**

Malpractice suits must be filed in a timely manner. It would not be practical for a physician to be sued for an incident occurring 30 years previously: records may be destroyed, witnesses may have died, and memories of the events probably are unreliable. Therefore, each state has a statute of limitation applicable to medical malpractice. These statutes bar prosecution of a case not commenced within a proscribed period of time.

In most states, the allocated time period ranges from 1 to 5 years, but exceptions exist. For example, the time period may not commence at the
moment of the negligence but instead may begin when a "reasonable person would have been made aware" of the injury.

Example 2. A patient sought care for a mole. The dermatologist performed a biopsy, and the lesion was interpreted as benign. Two years later, the patient sought care for the same lesion. The original tissue was reviewed a second time and was interpreted as a melanoma. The report was amended. The state had a 2-year statute of limitation.

The court had several ways to dispose of this case. It could have ruled the statute of limitation had commenced (1) at the time of the initial biopsy (as the defendant contended), (2) when the patient sought additional care for the recurrent lesion, or (3) when the report was amended (as contended by the plaintiff). The court ruled that the statute of limitation began to toll when the "injury occurred," and that injury occurred at the point when the patient's melanoma "moved from the epidermis to the dermis" [28]. The court reasoned further that, because the defendant produced no evidence to the contrary, this event was presumed to have occurred within the last 2 years; thus, the case was timely filed.

This decision highlights several important points:

1. The timing of a statute of limitation may be contested by the parties and does not begin to toll automatically when the report is issued.
2. Any affirmative defense pertaining to a statute of limitation must be raised by the defendant, and evidence must be supplied.
3. The court may decide medical matters in a way that is counterintuitive to physicians but is designed to achieve an equitable end.

For minors, the statute of limitation may be suspended until the patient reaches a specified age (often 18 years of age, but varying from state to state). This exception greatly prolongs the susceptibility of providers to suits arising from the care of children [29]. In dermatopathology, this issue may be of particular relevance to experts who render diagnoses of Spitz nevi or atypical spitzoid proliferations in children.

The business of medical malpractice

A law firm is a business, and expenses are incurred for front office staff, paralegals, society memberships, licensing fees, and continuing legal education. Most malpractice cases are accepted on a contingency basis. The attorney accepts a percentage of any ultimate collections, but in turn the firm fronts all the money for any expenses to pursue the case. If no verdict or settlement is collected, the expenses become losses.

These economic realities impact the types of cases that a law firm accepts. From the perspective of a plaintiff's attorney, an ideal case demonstrates unequivocal substandard care that directly resulted in easily quantifiable
damages. This economic scrutiny was well articulated in a 2004 report of medical malpractice in California, conducted by the Rand Corporation, a nonprofit research organization [30]:

Attorneys have always needed to be very careful about selecting malpractice cases because they are expensive to prepare and in California [and elsewhere], plaintiffs lose eight out of every ten cases taken to trial.

Indeed, two separate studies demonstrated that 67% to 82% of medical malpractice cases were disposed of without any type of indemnity payment [31,32].

Plaintiff’s attorneys are not particularly interested in “close calls”—cases exemplified by subtle differences of opinion that may be difficult for a jury to comprehend. Only about 1 in every 30 calls to an attorney’s office results in a case the firm is willing to pursue [23]. Although persuasive articles have been written about the detriments of a judicial system based on contingency fees, and indeed some benefit has been suggested for contingency-based systems with caps on the percentage of collections, the inherent economic incentive or disincentive that results from a firm covering expenses in this manner does act as a partial screen against cases of lesser merit [33].

**Damage caps**

Damages caps are statutory limits on damages. Many damage caps seek to limit only general damages for noneconomic harms, such as pain or emotional distress. More rarely, states may impose total damage caps, particularly with regard to public entities. The entities covered may vary with the language of the law. For example, physicians employed by the University of Colorado and caring for patients under those auspices have a total damage cap of $150,000 per person per incident [34]. Damage caps also affect the decisions that attorneys make about accepting cases.

**Reporting and consultation for melanoma and pigmented lesions**

**Reporting**

It is important to realize that, although the standard of care must be established in each individual lawsuit, consensus statements have established a fundamental minimum standard for the reporting of melanoma. In 1992 the National Institutes of Health Consensus Development Conference Statement on the Diagnosis and Treatment of Early Melanoma recommended the following content for a histology report on melanoma [35]:

Essential to the report
- Diagnosis
- Breslow level/thickness (in millimeters)
- Margin status
Suggested additional information

- Subtype (superficial spreading melanoma, nodular melanoma, acral lentiginous melanoma, melanoma in situ)
- Level (Clark I–V)
- Ulceration (present/absent)
- Regression (present/absent)
- Precursor lesion (present/absent; type)
- Satellitosis (present/absent)
- Angiolymphatic invasion (present/absent)
- Mitotic activity
- Host response (lymphocytic infiltrate)
- Radial versus vertical growth phase

Indeed, the items essential to the report are so very basic that it would be difficult to defend the omission of such information even as the actions of a “respectable minority.” In adapting the latest research to this recommendation, the importance of reporting ulceration is now more likely to be supported by the literature [36,37]. The University of Colorado has a standard reporting schema with bulleted points for each item so that the essential and recommended reporting characteristics are always commented on without omission.

Consultation

Some dermatologists, particularly those in the western United States, interpret histology slides themselves. In general, dermatology trainees receive more dermatopathology training than do general pathology trainees, and their board certification qualifies them to read tissue [38]. Dermatologists who do not read their own slides express a strong preference for a board-certified dermatopathologist, whereas general pathologists are more likely to receive specimens from non-dermatologists who perform skin biopsies [39].

Dermatopathology is a subspecialty recognized by the American Board of Medical Specialties [40]. In court, it will be asserted that dermatopathologists are most qualified to interpret difficult pigmented lesions [41–43]. Jean Bolognia, Professor of Dermatology at Yale, has likened the relationship between dermatologists, general pathologists, and dermatopathologists to overlapping bell curves [41]. Certainly it is possible that a non-dermatopathologist might be so far to the right of the bell curve measuring diagnostic prowess within his/her field that he/she surpasses a individual of lesser acumen falling to the left of a separate curve measuring dermatopathologists. In general and under most circumstances, however, the average dermatopathologist would be superior, at least in theory and very likely in the eyes of a jury.

When a generalist assumes care for a condition that normally requires referral, the generalist is held to the standard of care of a specialist [23]. Melanocytic neoplasms recognized as problematic and worthy of consultation are listed in Box 1. Histologic signs that might prompt
caution and consultation are listed in Box 2. Neither list can include all situations that will be encountered in practice.

Spitz nevi deserve special comment. In his 2003 examination of melanoma litigation, Troxel [1] noted a significant number of lawsuits against pathologists that involved melanoma mistaken for a Spitz nevus. Specifically, he advised:

If a pathologist is not seeing Spitz nevi on a regular basis and the patient is more than 20 years of age, the case should be sent to an expert [2]. If Spitz nevi are being seen on a regular basis and the patient is more than 20 years of age, unless all of the diagnostic criteria are present send the case to an expert [1].

There is substantial literature showing that Spitz nevi and spitzoid proliferations generate the least consensus of nearly all pigmented lesions, even among expert dermatopathologists [44].

Some institutions require intramural consultation with two signatures on the final histopathology report. There should be a low threshold for seeking consultative opinion on such cases [45].

<table>
<thead>
<tr>
<th>Box 1. Melanocytic lesions of higher risk for misdiagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spitz nevus/spitzoid proliferations</td>
</tr>
<tr>
<td>Nevoid melanoma</td>
</tr>
<tr>
<td>Lentigo maligna/lentiginous melanoma</td>
</tr>
<tr>
<td>Desmoplastic melanoma</td>
</tr>
<tr>
<td>Regressed melanocytic lesions</td>
</tr>
<tr>
<td>Proliferating nodules in congenital nevi</td>
</tr>
<tr>
<td>Clonal nevi/combined nevi/deep penetrating nevi</td>
</tr>
<tr>
<td>Collision tumors (melanoma and another neoplastic process)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Box 2. Warning signs of melanoma and/or controversial pigmented lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad junctional proliferations</td>
</tr>
<tr>
<td>Heavily sun-damaged skin</td>
</tr>
<tr>
<td>Failure to mature</td>
</tr>
<tr>
<td>Failure to disperse at the base</td>
</tr>
<tr>
<td>Deep pigment</td>
</tr>
<tr>
<td>Deep mitoses</td>
</tr>
<tr>
<td>Junctional confluence</td>
</tr>
<tr>
<td>Adnexal extension</td>
</tr>
<tr>
<td>Dermal nest larger than junctional nests</td>
</tr>
</tbody>
</table>
Opportunities to improve the quality of care for melanoma and pigmented lesions

In 1990, investigators conducted a retrospective review of more than 30,000 hospital records in New York, searching for evidence of malpractice. They discovered evidence of medical error, as judged independently by two physicians, in 1 of every 25 cases, but a lawsuit was filed in less than 4% of cases in which error transpired [46,47]. In a second study, released in 2006, Harvard researchers reviewed 1452 malpractice claims from throughout the country [48]. Although highly critical of the costs (defense fees and plaintiff’s contingency fees averaged $52,521/claim) and the time to resolution (average, 5 years) in medical malpractice cases, they found claims without error or injury represented a small percentage of the total cost. In fact, elimination of these “frivolous” claims would reduce the total cost of malpractice nationwide by only 13% to 16%. Even if malpractice lawsuits are rarer than news headlines suggest, and frivolous suits even more exceptional, the act of being sued is traumatic for the physician involved (and bringing suit is traumatic for the patient). Prevention is clearly the best strategy.

To this end, 10 suggestions to improve the quality of care of pigmented lesions are offered. These suggestions are culled from multiple sources including the indexed literature [1,6,14,45], the author’s experience, and the shared experiences of colleagues and mentors.

Suggestion #1: beware of low-power imitators

Low-power (low-magnification) examination of pigmented lesions can be an excellent discriminator of benignity, but it certainly is not an irrefutable indicator in any case. Experts in dermatopathology have expressed concern that increasing productivity standards, with an emphasis on high slide throughput and rapid turn-around, lead to medical error [1]. Nodular lesions with the low-power architecture of a benign nevus but the cytologic features and behavior of melanoma may be missed under these financial pressures.

Although the terminology of “nevoid melanoma” is debated, the existence of melanoma with many of the low-power features of a benign nevus is undeniable. First proposed in 1985 by Schmoeckel and colleagues [49], nevoid melanoma presents as a dome-shaped or verrucous lesion with little, if any, junctional activity or pagetoid extent. Lateral symmetry may be deceiving for benignity, but nevoid melanoma typically is not symmetric at the base [50]. Sometimes a suggestion of subtle maturation with descent into the dermis may be present [51]. The most helpful features for recognizing this form of melanoma are hypercellularity, cytologic atypia, and the presence of dermal mitoses (Fig. 1A, B), but such features are difficult to detect without a careful examination at higher magnification [1,50]. Immunohistochemistry, including mindbomb homolog 1 (MIB-1) or proliferating cell
nuclear antigen (PCNA) staining (proliferation markers), and homatropine methylbromide (HMB)-45 staining, often with a loss of the normal declining gradation with descent, may be helpful (Fig. 1C, D), but HMB-45–negative cases have been described [51].

Suggestion #2: beware of a partial biopsy

An excisional biopsy often is preferable for pigmented lesions, because the portion of the lesion sampled may not be representative of the entire lesion. Put another way, a partial sampling may lead to a partial diagnosis. Many experts believe the increased use of shave biopsies is driven by cost considerations [52], but any cost savings realized by partial sampling is trivial in comparison to the cost and detriment associated with a malpractice claim [1].
Nevertheless, it is important to understand that circumstances may render it detrimental or even impossible to obtain an excisional specimen. Lentigo maligna, a form of melanoma in situ most commonly seen on faces of elderly patients, can reach a diameter of several centimeters or more (Fig. 2), making an excisional biopsy difficult and excessively morbid if the lesion ultimately is deemed benign. In such situations, a broad shave or, alternatively, multiple punches or multiple incisional ellipses may be useful, with emphasis paid to the thickest and/or darkest areas of the lesion [53–55]. Both the size of the lesion and degree of clinical suspicion should be communicated faithfully to the physician examining the biopsy (see Suggestion #8).

Finally, a point of controversy in dermatology involves saucerization of lesions. Unlike a superficial shave biopsy, saucerization is a deep shave, going beneath the entire lesion, to accomplish a sampling similar to an excision but with greater expediency for the physician and patient alike. Experts in skin cancer have expressed favorable opinions regarding saucerization [56]. Evidence suggests that, when saucerization is performed by a skilled and experienced physician, the results are comparable to those rendered by excision [57]. Superficial shave specimens generally are inadequate for sampling pigmented lesions that are reasonably suspected to be melanoma, with the exception of lentigo maligna.

It is incumbent on the physician examining the specimen to inform the referring practitioner when a sample is inadequate for analysis or for

Fig. 2. A large clinically atypical macular lesion on the left cheek of a 79-year-old woman that was clinically consistent with lentigo maligna. A single 4-mm punch biopsy from the edge, submitted without any clinical information, revealed only a “junctional nevus with architectural disorder and moderate cytologic.” Clinicopathologic correlation and a larger shave biopsy solidified the malignant diagnosis. Communicating the size and clinical suspicion of the lesion and performing a larger biopsy might have minimized the initial confusion in this case.
exclusion of a diagnosis. It is best to inform the referring physician not only that the sample is inadequate but in exactly what way it is inadequate and how a repeat biopsy could be structured to provide improved diagnostic information.

Suggestion #3: beware of inflamed lesions

Benign lichenoid keratoses (BLKs) are benign neoplasms most often seen on the sun-exposed skin of older patients. It is well established that a melanoma in situ, or even a thin melanoma, may manifest a lichenoid inflammatory infiltrate mimicking BLK and even demonstrating necrotic keratinocytes (Fig. 3A, B). Dalton and colleagues [58] reported a lichenoid inflammatory infiltrate mimicking a BLK in 23 of 342 melanomas (6.7%)

Fig. 3. (A) Low-power examination demonstrated a lichenoid inflammatory infiltrate in the dermis in a specimen submitted without any clinical information (hemolysin-eosin, original magnification ×100). (B) Closer examination revealed a nested proliferation along the dermo-epidermal junction and focal areas of lentiginous and pagetoid proliferation of melanocytes. A call to the referring doctor revealed that the clinical suspicion was lentigo maligna (hemolysin-eosin, original magnification ×400). (C) Immunohistochemical staining with melanoma antigen reacting to T cell (MART-1) demonstrated a conconfluent proliferation of melanocytes along the dermoepidermal junction (MART-1, original magnification ×100). (D) Immunohistochemical staining also revealed pagetoid cells within the epidermis assisting in separating this lentigo maligna with lichenoid inflammation from an inflammatory lesion with “pseudonesting” (MART-1, original magnification ×400).
diagnosed at Wilford Hall and Brooke Army Medical Centers from 1999 to 2003. Fortunately, in only 6 of the 23 cases did the lichenoid infiltrate obscure a large portion (>30%) of the melanoma. Three of the 23 cases of melanoma were clinically amelanotic. This report highlights the importance of communication between the clinician and pathologist or dermatopathologist. BLKs often are confused clinically with non-melanoma skin cancer. A clear indication that the clinician observed a pigmented lesion may alert the dermatopathologist to the possibility of melanoma, increasing the likelihood that a correct diagnosis will be rendered. Serial sectioning and immunohistochemical staining for melanoma antigen reacting to T cell (MART-1) or another melanocytic maker may be useful in some cases (Fig. 3C, D) [59].

**Suggestion #4: beware of Spitz nevi**

Spitz nevi were first described by Sophie Spitz [60] in 1948 and were referred to as “juvenile melanoma.” Most Spitz nevi are characterized by spindled and epithelioid cells, with superior clefting at the dermoepidermal junction, fascicular recapitulation in the dermis, and Kamino bodies at the interface between the nests and the epidermis (Fig. 4A, B). Overlapping features with melanoma may include limited pagetoid extension and a few superficial mitotic figures. Clinically, one of the best initial discriminators is the age of the patient: Spitz nevi are common in children, and melanoma is not. Conversely, in adults melanoma is more predominant; in this population, new Spitz nevi occur only rarely. A Spitz nevus with “atypical features” (a controversy unto itself, because even the leading national institutions express disagreement as to the need for sentinel node biopsy

![Photomicrograph of a classic Spitz nevus](image1)

**Fig. 4.** (A) Photomicrograph of a classic Spitz nevus in a 12-year-old boy demonstrating spindled and epithelioid nests along the dermoepidermal junction, superior clefting of nests, and limited pagetoid extent in the epidermis (hemolysin-eosin, original magnification ×100). (B) Higher magnification along the dermoepidermal junction demonstrated well-formed Kamino bodies (hyalinized, eosinophilic conglomerations) at the interface of the nests with the epidermis. Although not pathognomonic for a Spitz nevus, the presence of well-developed Kamino bodies in melanoma is unusual (hemolysin-eosin, original magnification ×400).
for this type of lesion) always should be sent for expert review [61,62]. Finally, it is the recommendation of many experts and is the practice of the University of Colorado Health Sciences Center to recommend complete excision of all Spitz nevi, irrespective of a patient’s age [63].

**Suggestion #5: beware of regressed lesions**

Partial regression occurs in 10% to 35% of cases of melanoma [64]. Complete regression also has been recognized, with at least 38 well-documented cases reported in the literature [65]. Partial regression may consist only of dermal fibrosis and a chronic inflammatory infiltrate. Complete regression may demonstrate attenuation of the epidermis, decreased epidermal melanocytes, papillary dermal fibrosis, a chronic inflammatory infiltrate, and telangiectasias. “Nodular melanosis” is a term describing extensive deposition of dermal melanophages; but this term is not exclusive to regressed melanoma [66].

The importance of identifying histologic regression is twofold. First, partial regression may alter the treatment approach to an otherwise “thin” melanoma. Some investigators have found partial regression to be an adverse prognostic indicator [67–73], probably because of the pre-existing deeper extent of the melanoma, others, however, have failed to discover this relationship [74–81]. Although immediate consensus is unlikely, the prudent pathologist/dermatopathologist may decrease medicolegal risk by noting the regression and informing the clinician and/or patient of the controversy. Counseling with regard to management then may be adjusted. When regression is present, some institutions have proposed sentinel lymph node examinations on what normally would be considered “thin” melanomas [68,82,83], but other recent investigations have noted no association between regression and lymph node status [84,85].

Second, when evidence of complete regression of a pigmented lesion is observed, a prudent pathologist/dermatopathologist should recommend a thorough history and full-body skin examination, a clinical examination for lymphadenopathy, and possibly consideration of an eye examination [65]. Additional studies, such as a chest radiograph, imaging of the central nervous system, or lymph node sampling, might be considered based on the results of the examination. As treatment for metastatic melanoma improves, damages awarded for unconsidered complete regression will increase, because of the commensurate lost opportunity to treat.

**Suggestion #6: beware of desmoplastic melanoma and other spindle-cell imitators**

Desmoplastic melanoma is a spindle-cell malignancy most common in the elderly [50]. Unfortunately, these elderly patients also are predisposed to other spindle-cell neoplasms, such as spindle-cell squamous carcinoma
and atypical fibroxanthoma. Compounding the diagnostic difficulty, only about 50% of desmoplastic melanoma is pigmented [86].

When matched for depth, desmoplastic melanoma has been demonstrated to have a lesser metastatic potential but a higher chance for neurotropic invasion and local recurrence. It has been reported that 50% to 80% of desmoplastic melanoma demonstrates an overlying lentigo maligna [50], highlighting the need for accurate communication between the clinician and the pathologist/dermatopathologist. Experts have recommended a diligent search for a junctional component whenever an atypical spindle-cell proliferation is noted in the epidermis [87]. Desmoplastic melanoma often is bottom heavy, with cellular atypia more pronounced in the deeper extents of the tumor [50]. If the possibility of desmoplastic melanoma is being entertained for a lesion that has been superficially sampled, the prudent pathologist/dermatopathologist should request a second, more generous biopsy.

Finally, immunohistochemical studies may be necessary to confirm the diagnosis, but it is important to recall that desmoplastic melanoma often may be HMB-45 negative; S-100 staining is preferred for detection. Although there is increasing evidence that procollagen I staining is useful to confirm a diagnosis of atypical fibroxanthoma, a rare procollagen I–staining melanoma may be encountered, highlighting the need to use a panel of stains [88].

Suggestion #7: beware of confusion with atypical/dysplastic nevi (Clark’s nevus)

There is significant controversy surrounding the reporting of dysplastic nevi [89]. Grading of dysplastic nevi has become the expectation of patients and clinicians alike [90]. Even for those who do not formally grade lesions, re-excision of Clark’s nevi that have “indeterminate biologic potential” or “features of incipient melanoma” is recommended. Regardless of the terminology employed, certain abnormal nevi (atypical/dysplastic nevi or Clark’s nevi) demonstrate histopathologic features that overlap with melanoma, often to such a degree that distinction is difficult. In this situation, the nature of the biopsy (partial versus complete) and/or the status of surgical margins may prompt further advice to the referring clinician.

Recently, several reports have highlighted the difficulty of distinguishing some junctional nevi and atypical junctional nevi with lentigo maligna. Originally described by Kossard [91], and with modification by King and colleagues [92], this unusual form of lentigo maligna occurring in the elderly is distinguished from junctional lentiginous nevi based on the age of the patient, the central anatomic location of the lesions, and the presence of confluent growth of melanocytes along the dermoeipidermal junction (often highlighted with MART-1 staining). Interestingly, solar elastosis, often an indicator of solar damage and advanced age, was not prominent. More recently, a second group of clinical researchers corroborated the description of
this “lentiginous melanoma,” finding similar absence of atrophy and sun damage, similar central locations, but a larger age range spanning from 24 to 66 years [93]. The viability of this entity as a distinct subtype of melanoma probably will be studied further, and may be challenged, as additional reports accumulate.

Suggestion #8: beware of miscommunication regarding clinicopathologic correlation

Occasionally, one still encounters a referring clinician who declines to share a clinical impression because of a concern that this information will prejudice the outcome. This philosophy is dangerous and misguided. Even the simple indication that a pigmented lesion was encountered would be preferable to sending in all specimens as “rule out cancer” or “238.2” (the International Classification of Diseases-9 code for neoplasm of uncertain behavior). For example, noting that the lesion was pigmented may prevent a melanoma with intense lichenoid inflammation from being mistaken for benign lichenoid keratosis.

Furthermore, one must reasonably conclude that any financial savings realized by not taking time to jot down a meaningful clinical impression (literally seconds) are rapidly outstripped by the injurious and emotional impact of a malpractice claim. Even the mismarking of check-boxes regarding the nature of the specimen (shave, punch, excision), although seemingly trivial, may lead to concerns by the dermatopathologist that the specimen somehow has been misidentified, either in the physician’s office or in the pathology laboratory. This concern is greatly heightened when a clinical suspicion is inaccurate or omitted.

In sum, reliable communication between the clinician and the dermatopathologist enhances the accuracy of the end result, thereby improving the quality of care. Neglect of this communication can result in unnecessary medical error and heightened risk.

Suggestion #9: beware of “overcall”

In response to medicolegal pressures, it is understandable that pathologists and dermatopathologists might adopt an attitude of hypervigilance, often manifesting as a lower threshold of suspicion toward malignancy. Repeated studies have demonstrated, however, that the “overcall” of malignancy, when the lesion is benign, may result in damages to the patient and resultant malpractice litigation [1–3]. A recent review of the medicolegal aspects of neoplastic dermatology noted that melanocyte activation in the setting of psoralen and ultraviolet light therapy, nevi on “special sites” (the breasts, flexural areas, ears, genitalia, and acral skin), spindle cell nevus (of Reed), combined nevus, and deep penetrating nevus (Fig. 5) all have the potential for “overdiagnosis” when encountered by an inexperienced examiner [45].
In one example, a patient had received an erroneous diagnosis of melanoma when in fact the lesion was a deep penetrating nevus. Subsequent to this misdiagnosis, he received biochemotherapy consisting of interferon and interleukin-2, and the fluid shifts and the general toxicity of this regimen led to a myocardial infarction. The patient filed a malpractice claim against the pathologist for the misdiagnosis of melanoma that had precipitated the unnecessary biochemotherapy. Hence, simply becoming more likely to suspect malignancy will not render one impervious to claims of malpractice; more importantly, it actually may harm patients. Similarly, despite documenting the largest single series of deep penetrating nevi and clonal nevi in the literature [94], all of which showed benign behavior, the author often recommends complete excision (if clinically feasible) when margins are involved, chiefly because a recurrent lesion of this type may lead to diagnostic confusion, particularly if the future recurrence is sent to a laboratory that does not have access to the original material. This situation ultimately may lead to emotional distress or further physical harm to the patient.

Even immunohistochemistry studies, although seemingly protective, may lead to the “overcall” of melanoma. For example, it is increasingly recognized that pseudonests, or presumed amalgams of necrotic keratinocytes, benign melanocytes, and inflammatory cells, can simulate the junctional

Fig. 5. Photomicrograph of a deep penetrating nevus removed from the face of a 23-year-old woman demonstrating a deeply penetrating, bulbous, epithelioid melanocytic proliferation with associated dermal melanophages intimately associated with adnexal structures. The lesion was re-excised because of the potential of recurrence, and 4 years later there has been no recurrence or metastasis (hemolysin-eosin, original magnification ×25).
activity of melanoma, leading to a false diagnosis of malignancy. Pseudoneoplasts even may mark with MART-1 on occasion [95,96], emphasizing the importance of unfettered communication between the clinician and pathologist/dermatopathologist.

**Suggestion #10: beware of hubris**

_It was pride that changed angels into devils; it is humility that makes men as angels._

—Saint Augustine (354–430 AD)

Humility is not only good for the soul: it also is a good strategy for minimizing medicolegal risk. When a lesion is truly difficult, and reasonable physicians could come to differing conclusions, all parties should acknowledge this fact. Although there certainly are many verbose ways to do so, the laboratory at the University of Colorado Health Sciences Center uses a simple direct statement in the Comment section: “This is a difficult case.”

The purpose of such a statement is multifaceted: (1) it informs the clinician that differing opinions regarding this lesion may exist; (2) it alerts patients to any legitimate uncertainty with regard to the ultimate biologic behavior, so the patient may consider this uncertainty when soliciting additional opinions or consenting to treatment; (3) it signals to other dermatopathologists that conflicting signs were observed, and that differences of opinion may exist and are to be respected.

**References**


Briggs JC, Ibrahim NB, Hastings AG, et al. Experience of thin cutaneous melanomas (less than 0.76 mm and less than 0.85 mm thick) in a large plastic surgery unit: a 5 to 17 year follow-up. Br J Plast Surg 1984;37:501–6.


Quality Improvement in Laboratory Medicine: Extra-Analytical Issues

Giuseppe Lippi, MD\textsuperscript{a,\*}, Roberto Fostini, MD\textsuperscript{b}, Gian Cesare Guidi, MD\textsuperscript{a}

\textsuperscript{a}Sezione di Chimica e Microscopia Clinica, Dipartimento di Scienze Morfologico-Biomediche, Università degli Studi di Verona, Ospedale Policlinico G.B. Rossi, Piazzale Scuro, 10, 37134, Verona, Italy
\textsuperscript{b}Dirigenza Medica, Azienda Ospedaliera di Verona, Ospedale Civile Maggiore, P.le Stefani 1, Verona, Italy

Key points

- Blood collection remains an error-prone phase of the testing cycle.
- Critical areas include the appropriateness of the test request, patient and sample identification, criteria for acceptance and rejection of specimens, and the communication and interpretation of results.

In the age of evidence-based medicine, results of laboratory testing are integral to clinical decision making, to assist diagnosis, to guide or monitor therapy, and to predict health outcomes. Because of the increasing demand for laboratory testing and the pressure from cost containment and reimbursement policies worldwide, the primary goal is to achieve a high degree of throughput and efficiency without compromising quality. The balance between efficiency and quality is emerging as a strategic goal (Fig. 1). Lundberg’s brain-to-brain loop gives a comprehensive representation of the total testing process, divided into preanalytic, analytic, and postanalytic phases (Fig. 2) [1]. Because the majority of errors occur within the extra-analytic areas of testing, quality improvement programs focused purely on the analytic phase miss opportunities for improvement. Opportunities for improvement include reliable procedures for patient identification, quality criteria for specimen acceptance, and clinical consultations to ensure the appropriate ordering and interpretation of tests. The role of laboratory professionals has evolved to encompass consultation rather than merely the

\* Corresponding author.
E-mail address: giuseppe.lippi@univr.it (G. Lippi).

0272-2712/08/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cll.2007.12.007
performance of tests [2–4]. Efficiency in the total testing process develops through a complex strategy that incorporates the appropriate use of laboratory resources, reasonable management of throughput and complexity, reduction of turnaround time, rational interpretation of test results, and
clinical auditing. Among the extra-analytic processes, blood collection is perhaps the most critical [5]. In the United States it is estimated that more than 1 billion venipunctures are performed annually, and errors occurring within this process may cause serious harm to patients, either directly or indirectly [6]. This article focuses on the collection of blood specimens as a paradigm for quality improvement in the extra-analytic phases of testing.

Appropriateness of the test request

According to Lundberg’s brain-to-brain loop model, the total testing process begins with the formulation of a clinical hypothesis and selection of the most appropriate tests [1]. Inappropriate, redundant, or repetitive testing produces considerable organizational and economic problems for both the laboratory and the health care system. Some data suggest that up to 50% of the analyses performed daily in clinical laboratories may be inappropriate [7]. Clinicians order excessive tests for defensive reasons, for ease of access, or simply because of fear of uncertainty [8]. Repetition of laboratory tests also is driven by the lack of adequate and universally agreed upon standards for transferability of laboratory data [9].

Efforts to develop consensus, traditional education methods, use audits, and presentation of laboratory charges are most effective when combinations of strategies targeted at many behavioral factors are implemented (eg, dissemination of guidelines and best expert clinical opinions, peer management, information systems, modification of laboratory requisition forms, and payment deletions) [10,11]. Artificial intelligence may represent a valid solution, by implementing computerized order entry systems, checks for redundancy in test requests, protocol-driven requests, and decision-making strategies [12–14]. The formulation of clinical queries, rather than simple test requests by physicians, and the addition of interpretative comments to clinical reports are helpful if the process reflects best practice and reliable guidelines [15,16].

Overcoming test redundancy is more complicated. Advances in basic science, informatics, and applied technology have increased the complexity of laboratory investigation. There is evidence that a variety of innovative tests made available by the diagnostic industry have not been subjected to rigorous clinical and analytic evaluation [17–22]. Wide interlaboratory differences in data reporting compound the problem. The lack of standardization among laboratories adversely impacts the ability to establish common reference intervals and critical limits [9,23–25]. Implementation of common reference intervals and standards for data transfer will reduce unnecessary testing. A further approach is the longitudinal comparison of data, according to the concept of reference change value, originally proposed by Harris and Yasaka [26–28], which would make it possible to evaluate the significance of the change in two or more sequential measurements made at different institutions.
Patient and sample identification

The International Organization for Standardization 15.189:2003 clause 5.4, “Pre-examination procedures,” includes requisites for request forms, primary sample collection manuals, traceability of primary samples to an identified individual, monitoring of samples in transport, recording of receipt of samples, processing of urgent samples, and policies for rejecting samples [29]. This paradigm is acknowledged in the 2007 Laboratory Services National Patient Safety Goals from the Joint Commission on Accreditation of Health care Organization [30,31]. The document focuses on some critical extra-analytic activities, such as the accuracy of patient identification (Goal 1) and improving the effectiveness of communication among caregivers (Goal 2).

Ensuring positive identification of the patient is a challenging endeavor in all health care settings. Patient misidentification errors are potentially associated with the worst clinical outcome because of the possibility for misdiagnosis and mishandled therapy [32,33]. Identification errors are especially common with inpatient samples [33]. On average, 1 of every 18 identification errors results in an adverse event affecting the patient’s health result [34]. A simple method to decrease misidentification errors is to use at least two patient identifiers before collecting blood samples (eg, name and birth date). In the United States most hospitalized patients wear wristbands to aid in their identification. Barcode technology has been applied extensively in several areas of patient care, including the identification of patient specimens, medications, and blood products. Wristbands with unique barcoded patient identifiers have great potential for reducing patient misidentification [35–37]. Unfortunately, wristband errors also occur [38]. The Q-Tracks Study issued by the College of American Pathologists (CAP) identified six major types of wristband errors: absent or wrong wristband, wearing of more than one wristband, and partially missing, erroneous, or illegible information on the wristband. The CAP clearly indicates that continuous wristband monitoring decreases identification errors. The most reliable suggestion for improvement is that phlebotomists should refuse to collect blood from the patient when a wristband error is detected [39,40]. Additional suggestions include

1. Formulating easy-to-understand policies for collecting samples
2. Checking requisitions against results
3. Using information technology (ie, expert systems, autovalidation, delta checking, and similar algorithms) and using advanced technologic approaches for collecting patient information (eg, newer systems of bar-coded wristbands and radiofrequency chips)
4. Using of personal biometric information (eg, fingerprinting, iris scanning)
Criteria for specimen acceptance

The most common reasons for unsuitable blood specimens are hemolysis and clotting, which traditionally can be traced to incorrect procedures in sample collection [32,33,41]. Insufficient volume and clotted specimens are the most common causes for rejection of inpatient samples, whereas the prevalence of inappropriate containers is particularly high for outpatient specimens [33].

Hemolysis is a rather frequent occurrence and accounts for nearly 60% of rejected specimens [33,42,43]. A visually hemolytic specimen generally has extracellular hemoglobin concentrations greater than 0.2 g/L (3.1 μmol/L) [44]. Hemolysis produces false elevation of some parameters (amino transferases, creatinine, creatine kinase, iron, lactate dehydrogenase, lipase, magnesium, phosphorus, potassium, and urea), whereas others may be spuriously decreased (albumin, alkaline phosphatase, chloride, gamma-glutamyl transpeptidase, glucose, and sodium) [45]. Hemolysis also may interfere with some test methodologies, especially spectrophotometric assays, producing substantial analytic interference [45,46]. The most critical issue the laboratory faces with hemolytic specimens is distinguishing in vitro from in vivo hemolysis, a strategy that should be guided by clinical rather than by analytic considerations. In vitro hemolysis requires a standardized executive flowchart, encompassing hemolysis detection and quantification, analysis of the interference on test results, and recollection of new samples. In vivo hemolysis may be associated with clinically threatening pathologies that require immediate action [45].

Prolonged venous stasis, as commonly generated by tourniquet placement, may produce spurious results in laboratory testing. Ideally, the tourniquet should be applied only if necessary and should be removed quickly when the needle is in the vein. Venous stasis produces hemoconcentration and increased concentration of cells and large molecules, such as proteins and protein-bound substances [47–49]. Educating phlebotomists about this unwanted variability is critical [47–49].

Blood collection tubes need to be filled up to the correct volume to ensure a sufficient sample for testing and a proper balance between the blood and the additive in the tube (blood-to-additive ratio). The optimal sample volume needed for laboratory tests has been defined as twice the analytic volume of serum or plasma required for laboratory tests plus the dead volume of sample cup, replicates, and secondary tubes [50]. Underfilled tubes of blood may influence the reliability of results significantly, especially when the anticoagulant-to-blood ratio is critical, as in coagulation testing. Current Clinical and Laboratory Standards Institute/National Committee (CLSI/NCCLS) guidelines recommend that coagulation samples should be discarded if the evacuated tube contains less than 90% of the expected fill volume, regardless of the type of analyses performed [51]. This precaution is probably the most appropriate advice, because phlebotomists are not always aware of the variability introduced by a suboptimal
blood-to-anticoagulant ratio. Underfilling the ethylenediaminetetraacetic acid (EDTA) blood collection tube for hematologic testing also can lead to spurious test results, including low blood cell counts and hematocrit, morphologic changes to erythrocytes, and staining alteration. On the other hand, overfilling the blood collection tube prevents the tube from being mixed properly, leading to platelet clumping and clotting. Consequently, the CLSI/NCCLS recommends the hematologic tube be filled to ±10% of the stated draw volume [52,53]. The procedures that are used to handle the samples before testing also may affect the reliability of results; they must be shown to preserve sample integrity for the designated length of time under ideal storage conditions. Strict criteria for sample storage, handling, and transportation, established on the basis of documented stability tests, should be detailed, implemented in the routine practice, and monitored thorough the entire preanalytic processing by using ad hoc performance indicators.

**Communicating and interpreting results**

An audit of performance is essential to allow the laboratory to monitor its own activity and to assure good practice. The goal of medical interpretation of test results is optimization of the diagnostic decision process to improve the quality of medical care [54–56]. There still is an unacceptable gap in the understanding of the role of communication services in health care delivery. The International Organization for Standardization 15,189 standard, the very first standard developed particularly for the accreditation of medical laboratories, emphasizes the importance of appropriate interpretation and advisory services, although it does not specify requirements for assessing appropriateness, requesting tests, and interpreting results. Physicians’ satisfaction and clinical audits are valuable surrogate markers that can be used to assess the effectiveness of clinical laboratories [29]. Therefore, improvement in appropriateness may be achieved through promoting knowledge, by obtaining feedback from physicians in multidisciplinary meetings or interpretive rounds, and by reassessing the role of laboratory specialists in providing clinical advice regarding the use of laboratory resources and interpretation of test results.

Hyperkalemia may result from errors in sample handling (in vitro hemolysis, excessive shaking, delay in separation of serum from blood cells, inadequate clotting, low temperature of transportation, excessive centrifugation speed) or from acute and severe pathologies, such as hemolytic anemia, acidosis, and acute renal failure [57]. It would be rather difficult for the laboratory staff and the clinicians to identify promptly the source for this unexpected result, and misinterpretation would have a considerable impact on the well being of the patient [58]. Laboratory professionals and other health care employees must acknowledge the importance of their partnership and improved communication.
Summary

Box 1 summarizes the development of a multifaceted strategy to enhance quality throughout the total testing process.

The volume and complexity of testing are increasing constantly, together with the demand for a greater degree of quality and efficiency. Error management involves the identification of error-prone steps and the development of defensive strategies. Most medical errors are attributable to the system but are compounded by human error [59]. Effective quality control programs should encourage root-cause analysis of the processes involved and minimize fear of retribution for acknowledging errors [59].

Although technologic advances downstream of sample collection and handling have allowed several analytic steps to be redesigned, providing a greater degree of efficiency and quality, there is firm evidence that the risk to patients caused by laboratory errors throughout the total testing process is still high [60]. Experience teaches that extra-analytic quality is widely assumed, rather than being assured or guaranteed [61]. Laboratory automation can rationalize the entire workflow, reduce the number of manual errors, and enable a greater degree of safety for the operators and patients [62]. Reinforcement of laboratory–clinic communication also is essential to optimize the use of laboratory resources and to enhance the appropriateness and improve the interpretation of laboratory data [63].

Acknowledgments

The authors thank Dr. Elston for the excellent re-editing of this article.
References

Quality and Safety in Outpatient Laboratory Testing

Nancy C. Elder, MD, MSPHa,*
John Hickner, MD, MScb, Deborah Graham, MSPHc

aDepartment of Family Medicine, University of Cincinnati, PO Box 670582, Cincinnati, OH 45267-0582, USA
bDepartment of Family Medicine, The University of Chicago, Pritzker School of Medicine, 5841 S. Maryland Avenue, MC7110, Room M-160, Chicago, IL 60637, USA
cNational Research Network, American Academy of Family Physicians, 11400 Tomahawk Creek Parkway, Leawood, KS 66211-2672, USA

Clinicians in practice depend on laboratory testing to assist in diagnosing, treating, and monitoring their patients. Family physicians order laboratory tests on approximately 29% of the patients they see; internists order tests on 38% [1]. Although an increasing number of tests are performed in physicians’ offices by their own staff, physicians still depend on outside laboratories for most of their testing needs. When it comes to the total testing process, most laboratory personnel have limited knowledge of the steps taken by physicians’ offices, and similarly, most physicians and staff in physicians’ offices have limited knowledge about the laboratory. We believe that improving testing quality begins with eliminating some of this ignorance.

Key points for improving quality are as follows:

- Most laboratory errors originate in the pre- and postanalytic phases
- The greatest opportunity for improvement in quality exists in pre- and postanalytic processes
- Critical areas include specimen collection, labeling, transport, data entry, and communication of results
- Gains in quality can be achieved by collaboration between clinicians and laboratorians

Most “laboratory errors” originate outside of the laboratory. They occur not in the analytic phase but in the pre- and postanalytic phases. Almost two

* Corresponding author.
E-mail address: eldernc@fammed.uc.edu (N.C. Elder).

0272-2712/08/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
labmed.theclinics.com
thirds of the errors occur in the preanalytic phase, including transport problems, which result in delayed or lost specimens, incorrect labeling, issues related to phlebotomy, and data entry errors [2,3]. Unfortunately, most studies that address preanalytic quality focus on hospital laboratories that care for hospitalized patients. Quality concerns regarding pre- and postanalytic steps as they relate to outpatient physician’s offices have received little attention. This lack of attention represents a significant gap in quality data, because most medical care in the United States takes place not in the hospital but in the outpatient office setting. In a given month, for every eight patients who are hospitalized in the United States, 217 visit a doctor’s office. More than one half of those visits are to primary care clinicians [4]. A recent Institute of Medicine report, “Patient Safety: Achieving a New Standard for Care,” stressed the importance of focusing on outpatient care: “The consequences of medical errors in the (outpatient) setting—and the opportunities to improve—may dwarf those in hospitals” [5].

The goal of this article is to review the pre- and postanalytic testing process steps as they occur in primary care offices, with a special emphasis on the handoffs between physician’s offices and laboratories. We review the literature and report on our own research on the testing process in primary care offices. By expanding the understanding of the testing process to include steps that occur outside the laboratory walls, we hope to encourage individuals involved in laboratory quality to consider all steps of the testing process in quality initiatives.

The testing process in primary care

The testing process begins and ends with the patient. We have developed a model of the testing process as it occurs in primary care offices, and we believe that the model also likely applies to specialty practices (Fig. 1). Clinicians assess their patients’ complaints and conditions by obtaining a history, conducting a directed physical examination, and deciding whether laboratory or imaging tests are indicated. The test order is implemented in processes that often involve multiple staff in the clinicians’ offices and laboratories. While the test is being performed at the laboratory, results are tracked in the office until they are returned first to the office and then to the ordering clinician. The clinician reviews and responds to the results, documents this response in the medical record, and files the results for future use either electronically or in a paper chart. Patients are notified of the results and their meaning, appropriate clinical action is taken, and patients are monitored through any needed follow-up. This idealized testing process, however, often falls short in the realities of medical practice. Although a portion of primary care is provided within “closed” systems, in which primary care, specialty care, laboratory testing, and hospital care occur within integrated information and technology systems, most primary care is still provided by smaller, independent groups or individual providers. Most
primary care offices accept payment from multiple insurers. In 2001, 75% of primary care offices reported having three or more managed care contracts, and one third of practices had more than ten [6]. Each insurer may require that a practice order test from a specific laboratory or hospital [7], which makes communication between the physician and laboratory more complex. Although technology could assist with this complexity, most primary care offices do not have electronic medical records; offices that do are not typically digitally connected to their reference laboratories [8–10].

Errors and problems in the testing process in primary care offices

For 8 months in 2004–2005, family physicians and their staff from eight practices anonymously reported errors related to the testing process to a research database of the American Academy of Family Physician’s National Research Network (AAFP NRN). Participants submitted 590 error reports [11]. Eighteen focus groups were formed in these practices to discuss testing process problems and concerns [12]. The following discussion draws on this study to demonstrate the errors and problems that occur in all steps of the testing process in primary care offices.

Ordering and implementing

“It was a supervisor in the lab who told me wrong. It comes back to the day shift telling us one thing—and the afternoon doing another.”

–Comment from a participant in the AAFP NRN study.

A study performed a decade ago in the Ambulatory Sentinel Practice Network noted that 56% of “laboratory errors” reported by family physicians occurred in the preanalytic phase [13]. Common sources of error included mislabeled specimens, improper specimen collection or handling, clerical problems, and ordering mistakes. Our recent research confirms that ordering and implementing steps account for 31% of testing process errors reported in family physicians’ offices [11]. In a small number of technologically advanced offices, clinicians are able to enter laboratory orders directly into a digitally connected laboratory order system [14]. In most
offices, however, the order is transmitted verbally or in written form several times before the order reaches the technician who performs the test. In hospitals, many tasks to implement a laboratory order are performed by laboratory personnel, including phlebotomy, specimen labeling, specimen preparation, and transportation. In office practice, however, these tasks are typically performed by office staff with varying backgrounds and training.

**Tracking and return of results**

“If the patient doesn’t call and say, ‘I haven’t heard about my test results,’ we really don’t know that they’re not back.’”

—Comment from a participant in the AAFP NRN study.

Many physicians’ offices do not have adequate systems to monitor the flow of the laboratory tests they order. A 1996 survey of internists and family physicians found that only 17% of physicians reported having an effective test tracking system [15]. A 2000 study of Oklahoma family physicians noted that 78% of physicians had a tracking system but only 57% thought it was effective [16]. Difficulties with tracking include the complexity of how tests are ordered and implemented (they can be implemented at the time of the order, ordered now for implementation at a future time, or exist as a standing order), the number of laboratories used, the various ways in which test results are returned (digital, fax, mail, phone), and the competing demands on office practices that make test result tracking just one of many tasks performed to provide quality care to patients. Offices that do have test tracking systems find that these competing demands often lead to poor compliance with tracking protocols [12]. Physicians offices report that some test results never return from the laboratory. In a study of Colorado office practices, physicians reported that laboratory and radiographic reports make up half of the missing clinical information in their records [17].

**Clinician response to results and documentation**

“I think one of the biggest problems is the timeliness of the doctors...some are really bad, their charts just sit for weeks and weeks and weeks with the lab results, even though the nurses look at them first and know that they’re pretty much normal, but normal isn’t always normal.”

—Comment from a participant in the AAFP NRN study.

In the AAFP NRN study, physicians and their staffs reported that 7% of all testing process errors related to incorrect, delayed, or inappropriate clinician responses [11]. Errors related to charting and documentation are also common and may account for as many as 15% of reported errors. These errors may relate to missing an important abnormal result or misinterpreting results in their clinical context or, more commonly (as noted in the previous quote), delay in responding to results. Results sometimes can
be filed mistakenly in a patient’s record with no review by a clinician. Again, laboratory test results are just one piece of clinical information with which clinicians and their staffs deal on a daily basis. Without organized systems and processes to manage these clinical data, important results may be overlooked.

**Patient notification and follow-up**

“In our practice, there are about 12 different systems for notifying patients. Every doctor has their own way of doing that.”

–Comment from a participant in the AAFP NRN study.

Clinicians and their staffs acknowledge that notifying patients of their results is inconsistent at best. In a 1996 survey of internists and family physicians, only 36% responded that they routinely notified all patients of abnormal results [15]. A more recent survey that documented the notification of abnormal pap smears and cholesterol levels found such documentation in patient charts only 47% and 67% of the time, respectively [18]. Although most patients want to be notified of their test results, even when normal [19], there is still a “no news is good news” attitude in some physicians’ offices [12].

Following up with patients to ensure that the appropriate action has been taken based on the test result is even more difficult. Although a certain amount of responsibility for follow-up does rest with patients, clinicians have ethical and legal obligations to attempt to make sure patients get the follow-up care they need. Finding workable systems to ensure follow-up has been difficult [11,12,15,16]. As one participant in the AAFP NRN study put it, “Once the letter is mailed [to the patient with the results] it’s kind of up to that patient.”

**Handoffs between laboratory and physician’s office**

Participants in our AAFP NRN focus group study gave us insights into why problems exist during the transitions between their office and laboratory. These problems occur with the ordering and implementation of the test and with the return of the test results. Three main contributing factors to these transitional problems were communication, technology interfaces, and conflicting policies and procedures.

**Communication**

That poor communication is a common factor in office-to-laboratory-to-office problems is not surprising. A study of errors in family physicians’ offices in Colorado found that communication problems were contributing factors in 71% of all reported medical errors [20]. Participants in our study cited problems with informal communication with laboratory personnel,
such as when conflicting advice is received, and with formal communication. For example, a physician noted that “It amazes me how often you’ll write an order and say ‘Fax report to me as soon as possible’ or you’ll give them a pager number and say ‘Page me’ and that never gets done.” Anyone who ever played the childhood game “telephone,” in which a message is whispered person to person around a circle with the final message sounding nothing like the original one, can understand how the sheer number of people involved in these transitional testing process steps (eg, clinician, medical assistant, office clerical staff, laboratory clerical staff, medical technologist, laboratory manager) can lead to messages and information getting lost, delayed, or altered.

**Technology interfaces**

Although technology has the potential to greatly improve the transfer of data between the office and the laboratory [21], it also can be a source of error. Participant offices in our study included urban academic practices, suburban group practices, and rural community practices. None of our practices was in health systems that were fully digitally integrated with their laboratories or hospitals, and all had technology problems. As one physician noted, “You may or may not be surprised, it would literally take an act of God to get those computers to talk to each other… Even though it’s all put in electronic format, and all potentially available instantaneously, it still comes in paper.” When computer systems do interface with each other, it is not always successful. “We have a major problem with identifying who really ordered the test and that’s really something beyond our control because we have to use a certain hospital lab which is really not designed to be an outpatient clinical lab but an inpatient hospital lab and they just have an archaic computer system and we tried really hard to change that and as of yet it really hasn’t changed in 10 years.”

**Conflicting or unknown policies and procedures**

When physician office staff are involved in implementing test orders, such as completing requisitions, drawing blood specimens, preparing specimens, and arranging for transportation, they may not always be aware of procedural changes at the laboratory. For example, a study participant describes an episode, “You’d order something the right way, you know that it’s supposed to be a serum tube and the lab says ‘Oh no, we need a lavender.’ Well it didn’t print up that you needed a lavender, and I’ve never had to give you a lavender for that before. You’ve got to call the patient back in and redraw it.” Just as laboratory testing is only a small part of patient care for practices, the office-based client may be only a small part of the laboratory’s workload, and up-to-date notification of procedural changes to the physician’s office is unlikely to be a laboratory priority.
Ensuring safety and quality in the office–laboratory transitional steps

Physicians’ offices and laboratories have their work cut out for them to improve the quality of the testing process. Luckily, many individuals and organizations are working to find the best ways to do just that [21–24]. Some general patient safety principles, applied to the pre- and postanalytic transitional steps, might improve the safety and quality of the testing process, including the following steps:

- **Standardization**: This includes standardizing processes within a practice, within a laboratory and between offices and laboratories. Sometimes standardization can be community wide. For example, groups of physicians, insurers and laboratories may work together to standardize nomenclature for commonly performed tests and panels of tests, and develop requisition forms for each laboratory that use the same nomenclature and listing order. Physician’s offices that use several laboratories could improve the quality of their test ordering when they don’t have to navigate through many different forms and lists or deal with varying terminology. An office and a laboratory may standardize how test results are returned to the office, as well. If a laboratory both faxes and uses a computerized print out to the same office, the staff that collect and process the results will have more opportunities for error.

- **Communication**: From the office and laboratory sides, improved communication in the form of meetings, phone calls, and written and electronic memos can assist in preventing pre- and postanalytic problems. Putting human faces and names on contacts between laboratory and office can make asking questions, getting clarifications, and instituting new policies easier than when dealing with a generic “someone in the lab.” Tracking concerns and problems from the laboratory and the office and arranging regular meetings to discuss common issues and solutions may prevent “unlabeled specimen,” “incorrect tube,” and “quantity not sufficient” calls to the office and “we never received the results” calls to the laboratory.

- **Planned redundancy and back-ups**: Although redundancy and efficiency hardly seem compatible terms, planned redundancies—the built in double checks and back-ups—usually save time by preventing errors and decreasing the amount of time spent scrambling to recover information. Planned redundancies include tracking systems for laboratory orders that use a log book and a copy of a requisition or a computerized lab log with a back-up computerized billing log, along with a system of oversight for reviewing log books. Laboratories could institute similar system of double checks for sending test results to physicians’ offices.

- **Safety culture**: A safety culture has been defined by the Institute for Health Care Improvement as “an atmosphere of mutual trust in which all staff members can talk freely about safety problems and how to solve
them, without fear of blame or punishment” [25]. Safety cultures are important in the laboratory and the physician office setting. People who work in health care, although occasionally responsible for making errors, also help prevent and ameliorate errors [26]. A safety culture seeks out individuals and the safety tasks they provide to incorporate such actions into the day-to-day functioning of medical practice.

Summary

Laboratory testing is an integral component of high-quality primary health care. For example, monitoring glycosolated hemoglobin levels in patients with diabetes, screening for lipid disorders in patients with heart disease, and performing Papanicolaou smears in adult women are just some examples of recommended primary care quality measures [23]. Primary care clinicians need a reliable laboratory to direct patient care. Threats to safety and quality pervade the testing process, however, with many of the risks occurring during the pre- and postanalytic phases during which specimens and information transfer between the laboratory and the physician’s office. We believe that improvement begins with awareness that a problem exists and belief that the problem can be fixed. Laboratories and primary care physicians are working to improve quality in their respective fields, but greater gains can be achieved if quality efforts are integrated to include the entire testing process that begins and ends with the patient. We hope that this brief review of the testing process from the physician’s office perspective can serve as an impetus for further innovations in testing quality.

References


The Dangers of False-Positive and False-Negative Test Results: False-Positive Results as a Function of Pretest Probability

Brian R. Jackson, MD, MSa,b,*

aARUP Laboratories, 500 Chipeta Way, Salt Lake City, UT 84108, USA
bUniversity of Utah, Salt Lake City, UT, USA

Recent decades have seen a dramatic expansion in the technical capabilities of clinical laboratories. Whereas 50 years ago a few dozen discrete tests were available, currently there are thousands. This progress in technical capability has not been matched by progress in using these tests appropriately, however [1,2]. With the contemporary focus on quality and efficiency of care, the time is ripe to correct this imbalance.

Overordering, underordering, and misordering of laboratory tests are all common [3]. There seem to be many factors at work: difficulty in staying up-to-date with the thousands of tests currently available in clinical practice, lack of familiarity with the principles and tools of evidence-based medicine, and the perception that ordering more tests is always better than ordering fewer tests [4]. When a physician is considering a diagnosis for which a test is available, there is often a perceived need to order that test even in the absence of familiarity with that test’s analytic and clinical accuracy [2,5]. (Note that in many cases, reliable data on clinical accuracy may not even exist.) When multiple tests are available, a physician may feel the need to order many or all of them. Once those tests have been performed, physicians often overweight or underweight specific results in deciding how to treat a patient. Finally, an unnecessarily ordered test may lead to an entire cascade of further testing to clarify false-positive or ambiguous results. Two assumptions underlie these practices: (1) information is by
definition helpful in making a diagnosis and (2) physicians can mentally integrate numerous pieces of information into a single, accurate diagnosis.

There are several problems with this reasoning: (1) No test is perfect (i.e., every test sometimes produces false-positive or false-negative results). Many common medical tests have surprisingly high rates of false-positive or false-negative results. (2) Humans, including doctors, do not intuitively integrate multiple pieces of information in a mathematically sound way; faulty data synthesis is a common cause of diagnostic error [6,7]. (3) Because diagnosis drives therapy, false-positive diagnoses can lead to unnecessary and potentially dangerous therapy.

**Imperfection of laboratory tests**

As laboratorians know well, every test in clinical medicine is subject to false-positive and false-negative results (Box 1). For many common analytes, reference intervals are set to the central 95% of values based on a healthy population, which ensures a 5% false-positive (more precisely, false “abnormal”) rate when that test is ordered on a healthy patient (95% specificity). Sensitivity, on the other hand, varies greatly based on the test and the disease under consideration. For other analytes, reference intervals are based on clinical criteria so that specificity and sensitivity vary based on the test and the disease.

This picture is complicated by larger combinations of test results. When multiple tests are ordered simultaneously (assuming the tests to be statistically independent), the likelihood of having at least one misleading result grows. For example, a complete metabolic panel contains 14 discrete tests; so assuming approximate statistical independence, one half of healthy patients have at least one abnormal result (0.9514 ≈ 0.49; 49% have no abnormal results).

The optimal role of any laboratory test for a particular disease depends highly on its sensitivity and specificity. For example, HIV and Lyme disease can be diagnosed serologically via a two-step algorithm, namely enzyme

---

**Box 1. Common avoidable causes of false-positive and false-negative results**

- Use of a test at an inappropriate time (e.g., when the patient is on a medication that interferes with the results)
- Use of an obsolete test
- Use of a test with inherently poor sensitivity or specificity
- Use of a test on a patient population with low or high prevalence of the disease under consideration
- Use of a test that lacks extensive clinical validation
- Use of a test on a patient population that differs from the intended or studied population
immunoassay followed by Western Blot. In the case of HIV, positive results are diagnostic of disease in almost all patients older than 1 year of age, and false-negative results are essentially never seen once the infection is more than 2 weeks old. The diagnosis of HIV infection is almost entirely laboratory based. For Lyme disease, on the other hand, sensitivity of the two-step algorithm has been reported to be as low as 10% [8,9]. Lyme diagnosis is primarily based on history and clinical findings, with laboratory results relegated to a supporting role [8]. A 1997 American College of Physicians position paper recommended not even ordering Lyme serology unless the pretest probability is between 20% and 80% [10]. Below the bottom threshold, other diagnostic considerations should be pursued instead. Above the upper threshold, antibiotic treatment should be initiated regardless of laboratory results.

Integration of evidence

Many people assume that natural human intuitive reasoning follows statistical principles. For example, a recent medical journal commentary entitled, “Why clinicians are natural Bayesians” argued that experienced clinicians are more skeptical of a given test in the setting of low pretest probability as opposed to one of high pretest probability [11]. The title reference was to Bayes’ theorem, a mathematical mechanism for combining prior evidence and new evidence into a single number [12]. Evidence shows, however, that humans (including physicians) rely primarily on rules of thumb (heuristics) rather than any sort of subconscious statistical calculations. The examples in the commentary were actually better represented as a heuristic: “When pretest probability is low, a positive result is more likely to represent a false positive.” A more rigorous approach to Bayes’ theorem would go further and provide a quantitative estimation of just how likely that false-positive result is.

Bayes’ theorem

Bayes’ theorem, also known as Bayes’ rule, comes from a 1763 essay by the Reverend Bayes [12]. It states that to find the conditional probability of some event B given that you know event A to be true, you can multiply the conditional probability of A given B by the unconditional probability of B, then divide by the unconditional probability of A (Fig. 1) [13]. A much simpler formulation for use in diagnostic reasoning is that the posttest odds of disease equals the pretest odds times the likelihood ratio of a test [13]. Note that this assumes the test to be statistically independent of any tests performed previously to arrive at the pretest diagnosis (eg, performing the same test 100 times and getting the same result each time normally should not influence the diagnosis any more than doing the test once). Put even more simply, you start with a number that reflects the level of belief that the patient has a particular diagnosis, given all previously observed signs and symptoms. Upon performing an additional diagnostic test and receiving the results, you then multiply
the first number by a ratio that reflects the test results and the discriminative power of that particular test. The result is a number that reflects the updated probability that the patient has that diagnosis.

One catch is that to use this equation, a physician must express probabilities in terms of odds and likelihood ratios, both of which are unfamiliar and unintuitive to most physicians. Although the word “odds” is colloquially used as a synonym for probability, the technical meaning is actually the ratio of a probability to its complement. For example, if the pretest probability of disease is 20%, the pretest odds equals \(0.2/(1.0 - 0.2) = 0.25\). Converting odds back into probability requires dividing the odds by (1 plus odds). For example, an odds of 3.0 corresponds to a probability of \(3.0/(3.0 + 1.0) = 0.75\) (see Fig. 1).

Likelihood ratios are also potentially confusing. The likelihood ratio that a patient with a given test result has a particular disease equals the likelihood (loosely speaking, the probability) that a patient with that disease would be expected to have that test result divided by the likelihood that a patient without the disease would have that test result. If a test has only two possible results (positive and negative), then the likelihood ratio for a positive test result equals sensitivity divided by 1 minus specificity, and the likelihood ratio for a negative test equals 1 minus sensitivity divided by specificity. Most laboratory tests have more than two possible results, however. Serum glucose, for example, is reported as an integer value that can range from close to 0 to more than 500 mg/dL. For such tests, the likelihood ratio for each possible result can be constructed from frequency distributions of analyte levels in diseased and healthy populations.

\[
\text{Likelihood ratio of a positive test} = LR^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}}
\]

\[
\text{Likelihood ratio of a negative test} = LR^- = \frac{1 - \text{Sensitivity}}{\text{Specificity}}
\]
To simplify application of Bayes’ rule in medical diagnosis, proponents of evidence-based medicine have published nomograms that allow quick estimation of posttest probabilities as long as the likelihood ratio for a test is known (Fig. 2) [14]. The need to know the likelihood ratio remains a significant limitation, however, because most laboratories do not provide likelihood ratios along with their test results.

Intuitive decision making

In the real world, most doctors do not directly use Bayes’ theorem, either by performing the calculations or by using nomograms. Many doctors believe that they use Bayesian reasoning [11] in the sense that they adjust their diagnostic beliefs upward or downward based on each new piece of evidence. Such intuitive probability estimation tends to be inaccurate, however [15–18]. Doctors, particularly novices, tend to underestimate high pretest probabilities and overestimate low pretest probabilities [15]. In adjusting
these probabilities based on test results, doctors tend to overweight abnormal results and underweight normal results [19]. Doctors also tend to be overly influenced by negative results of tests with poor sensitivity and positive results of tests with poor specificity [16].

Most human decision making is based on heuristics (mental shortcuts or rules of thumb) rather than statistics [16,20]. Heuristics are adaptive in the sense that they allow humans to make the numerous quick decisions necessary to survive in a busy world; however, they are also subject to several well-described biases. An example that is particularly applicable to diagnostic testing is referred to as the anchoring bias [20,21].

**Anchoring**

Bayes’ theorem can be used to demonstrate that given a series of independent diagnostic tests, posttest probability depends only on pretest probability and each of the likelihood ratios associated with those test results—not on the order in which the tests are performed. Psychologically, however, the anchoring phenomenon causes the first test result in a series to be overweighted with respect to subsequent results. More specifically, people tend to use the first piece of evidence presented to them as a starting point for estimation and then adjust it up or down based on subsequent pieces of evidence. The adjustments are often insufficient, however [22]. For example, in one psychology study, subjects were asked whether the percentage of African countries with membership in the United Nations was higher or lower than a stated number, and subjects were then asked to estimate the percentage. For subjects given 10% as a starting point, the median estimate was 25%, but for subjects given 65% as a starting point, the median estimate was 45% [20].

**Potential harms of diagnostic testing**

All of this information would be primarily of academic interest to physicians but for the fact that diagnosis drives therapy. When a laboratory result misleads a physician into making an inaccurate diagnosis, the patient is at risk for receiving unnecessary therapy or having delays in receiving necessary therapy. Even when a diagnosis is corrected before initiation of therapy, the patient may undergo unnecessary and potentially invasive follow-up testing or suffer psychologic harms (Box 2).

A dramatic and tragic example of misdiagnosis caused by laboratory testing involved a healthy 36-year-old woman with false-positive human chorionic gonadotropin (hCG) results [23]. On routine preoperative testing for sinus surgery, an elevated serum hCG led her physicians to believe that she had choriocarcinoma. Repeat hCG levels were similarly elevated. Results of a series of invasive and noninvasive diagnostic tests were all negative, including dilation and curettage, laparoscopy, and CT and MRI imaging of the chest, abdomen, and pelvis. Despite these negative results,
she underwent methotrexate chemotherapy. When her hCG levels failed to return to normal, she underwent a repeat pelvic MRI scan, which showed a possible endometrial lesion. Based on this finding, a hysterectomy was performed. Histology showed only a focus of endometrial hyperplasia, however. The patient then underwent combination chemotherapy that resulted in severe pancreatitis and coma. At that point it was discovered that the patient’s urine hCG was not elevated, and the elevated serum hCG results had been false positive because of heterophile antibodies. The patient eventually recovered but was unable to bear children [23].

In this example, the treating physicians seemed to have anchored their diagnostic reasoning to the initial serum hCG results, although extensive subsequent diagnostic evaluation produced negative results. Importantly, this set them up to overinterpret a later ambiguous diagnostic finding, which resulted in unnecessary hysterectomy. With a more quantitative assessment of the evidence, the physicians might have combined the low pretest probability (gestational trophoblastic disease has an incidence of approximately 5 per 100,000 woman-years) [24] with the multiple negative imaging studies to conclude that the laboratory results were highly likely to be inaccurate.

A less dramatic, but much more common, patient harm occurs when a positive screening test result is found on confirmatory testing to be a false-positive result. Patients may not be physically harmed (whether biopsies and other invasive tests should be categorized as a harm is debatable), but they are negatively impacted. In one study, women with false-positive mammogram results were found to suffer significantly more cancer-related anxiety than a control group with negative mammogram results [25]. There are also significant economic costs attributable to false-positive test results, mostly because of the cost of follow-up testing [26].

**Evidence-based approaches to specific diagnostic scenarios**

**Emerging diagnostic tests**

The regulatory and economic environment in the United States and many other countries guarantees that many laboratory tests will be commercially

---

**Box 2. Potential consequences of false-positive and false-negative results**

- No impact in some cases
- Cascade of increasingly expensive or invasive follow-up testing
- Lengthened hospital stay
- Additional office visits
- Inappropriate therapy
- Psychological trauma caused by false belief of having a disease
available before they have been clearly demonstrated to be of clinical benefit [27]. Initially, an analyte is found to be correlated to a particular disease state, and information is published in the medical literature and presented at medical conferences. With testing available through research laboratories, enthusiasm for the test begins to build among the medical specialists associated with that disease. In turn, this enthusiasm drives market demand for the test, which leads more clinical laboratories to offer the test. Ideally, research studies to evaluate clinical use will be undertaken in parallel with the growing availability of a test, but in practice it varies from test to test. Some tests become popular in the absence of good evidence; others lag in popularity despite the presence of good evidence. If the test is first developed in a commercial setting rather than an academic setting, there may be aggressive marketing of the test to clinicians, which drives widespread use well in advance of good clinical studies. For diseases that lack clear gold standard diagnostic criteria, there may be additional psychologic pressure to order newly available tests.

There are two main problems with the use of emerging diagnostic tests in clinical care. One is that early reports of diagnostic accuracy often turn out to be overly optimistic [28]. The patients in early studies often have more severe or more classic findings than the overall diseased population, which results in selection bias [28]. A few examples of this phenomenon include carcinoembryonic antigen for colon cancer [29], the dexamethasone suppression test for depression [30], and iodine 125-labelled fibrinogen scans for deep venous thrombi [31]. The second problem is that in many cases, new tests turn out to be highly correlated with existing tests, so they do not add as much information as initially believed [32,33].

Mathematically, new and clinically unproven tests can be represented with likelihood ratios that have wide confidence intervals (see following case report). Note that it is not sufficient to derive such confidence intervals from the original research literature because of potential selection bias. Likelihood ratios based on early studies should be considered best-case scenarios.

**Case illustration: application of new diagnostic technology**

Dr. Jones has a young adult patient with chronic diarrhea consistent with inflammatory bowel disease. Based on clinical history, radiology, endoscopy with biopsy, and serologic testing for p-ANCA and anti-*Saccaromyces cerevisiae* antibodies, Dr. Jones estimates that there is an 80% probability that the patient has Crohn’s disease and 20% probability that the patient has ulcerative colitis. However, Dr. Jones would like to further confirm the diagnosis of Crohn’s disease. Dr. Jones also has heard about a new test, outer membrane protein-C (OmpC) IgG/IgA, which is purported to help distinguish between Crohn’s disease and ulcerative colitis. Dr. Jones consults PubMed and locates a paper by Zholudev and colleagues [34] on the value of OmpC testing. In that article, anti-OmpC antibodies were found...
in 25% of patients who had Crohn’s disease versus 11% of patients who had ulcerative colitis. Based on this (admittedly small) study, the likelihood ratio of a positive result for differentiating Crohn’s disease from ulcerative colitis would be 0.25/0.11 = 2.3, and the likelihood ratio of a negative result would be 0.75/0.89 = 0.84. Because this is a relatively early study, Dr. Jones assumes that these likelihood ratios are optimistic and (arbitrarily) estimates true likelihood ratios as being anywhere from 1.5 to 2.3 for a positive test and 0.84 to 0.95 for a negative test.

Under the optimistic assumptions, after converting the pretest probability into pretest odds, multiplying by the likelihood ratio, and then converting back to a probability, the posttest probabilities of CD would be 90% and 77% for a positive and negative OmpC result, respectively. Under the pessimistic assumptions, the posttest probabilities would be 85% and 79% for a positive and negative OmpC result, respectively. Either way, the test would not substantially alter the diagnostic probabilities.

Now consider a bit further the implications of a negative OmpC result. Mathematically this result should be completely disregarded by Dr. Jones (77% and 79% are practically indistinguishable from 80%, particularly because 80% was a rough estimate to begin with). On the other hand, because it contradicts the doctor’s working diagnosis, it creates the possibility that Dr. Jones will feel obligated to order one more test to “clarify” the diagnosis. It might, for example, steer Dr. Jones toward performing an additional endoscopy with its associated cost, discomfort, and risk. The patient in this case would have been better off not having the OmpC test in the first place, because its likelihood ratio was a priori too weak to tilt the diagnostic balance away from Crohn’s disease and toward ulcerative colitis.

(Note that this example is for illustration only and ignores the more recent literature on OmpC, although at the time of this writing the clinical evidence supporting use of OmpC was still sparse [35].)

**Replacement of older tests by newer tests**

Another common clinical scenario is when a new diagnostic test is developed that on clinical validation is demonstrated to be superior to an older test. In such cases there tends to be a lengthy overlap between widespread adoption of the new test and discontinuance of the old test [36]. During this overlap, the tests are used in tandem. A possible rationalization is that this gives physicians time to develop trust in the new test or perhaps to recalibrate their diagnostic intuition to reflect the performance characteristics of the new test. Once a new test has been clearly shown in large clinical studies to be superior to the older test, however, routine use of the older test becomes highly questionable.

An example is the gradual replacement of creatine kinase-MB fraction (CK-MB) by cardiac troponins for diagnosis of acute myocardial infarction. Although troponin I and troponin T have replaced CK-MB as the preferred...
test because of improved sensitivity and specificity [37], at many institutions the standard algorithm for ruling out myocardial infarction is serial ordering of a panel that contains troponin and CK-MB (and sometimes myoglobin). The harm in this is not so much the direct cost of the extra CK-MB tests but rather the downstream costs of potential discordant results. Suppose, for example, that a patient enters the emergency room with chest pain, does not have diagnostic changes on ECG, and has repeatedly negative results for troponin levels. Such a patient would ordinarily be discharged at that point. If the CK-MB measurements were slightly elevated, however, even in the absence of a typical time course, this could be disconcerting to the physician. Combined with the fact that missed diagnoses of myocardial infarction are a major source of malpractice suits [38], a physician might be tempted to delay discharge pending a stress test or one more cardiac marker panel.

Cancer screening

Cancer screening is a particularly interesting special case of the false-positive problem in laboratory testing. In most clinical scenarios, physicians do not consider ordering a laboratory test to diagnose a particular disease until the pretest probability based on symptoms and other prior evidence exceeds some minimum threshold. In primary screening, however, patients are, by definition, asymptomatic and pretest probabilities of disease are low. Studies of radiologic screening for two cancers with the highest incidence—mammography for breast cancer in elderly women and CT for lung cancer in smokers—have shown the prevalence of screen-detectable disease to be approximately 1%. (False-positive rates in both these studies were approximately 10% [39,40].) Most other cancers have much lower prevalence and much lower pretest probability. Ovarian cancer, for which a large number of screening tests have been proposed, has a prevalence of undetected disease of approximately 0.04% in postmenopausal women [41]. A screening test for ovarian cancer would have to have a remarkably high specificity rate of 99.6% just to achieve the 10% positive-predictive value of the previous two examples [42].

Not all cancers are equal

Discussing a cancer screening test in terms of sensitivity and specificity (or likelihood ratios) actually glosses over a key point: not all cancers, even of the same histologic type, are created equal. Because screening tests are only performed periodically (as opposed to monthly or daily), some cancers become clinically detectable during the interval between tests despite a negative prior test. The more aggressive a cancer, the more likely it is to fit this pattern and be missed by a screening regimen. In other words, screening tests are more sensitive in their ability to detect low-grade cancers than high-grade ones.

Some of the low-grade cancers may represent what has been called pseudodisease [43] (ie, “disease” that may be histologically real when detected by
a screening test but would never have caused the patient any problems). This category includes early-stage cancers that do not progress, such as many cases of breast ductal carcinoma in situ [44], and cancers that do progress but the patient dies of something else before the cancer becomes symptomatic, such as in many cases of prostate cancer. Detection of pseudodisease may represent true-positive results from the histologic perspective but false-positive results from the patient perspective because the patient does not benefit from the knowledge.

An illustration of pseudodisease comes from neuroblastoma screening. In theory, urine vanillylmandelic acid and homovanillic acid can be used as noninvasive and relatively inexpensive screens for neuroblastoma in infants. Mass screening was instituted in the 1980s in Japan [45], and large clinical trials were subsequently performed in Canada and Germany [46,47]. Both clinical trials failed to show reduced mortality in screened children (both detected a small but statistically insignificant increase in mortality) and concluded that screening was of no benefit. How could this conclusion be correct, considering that neuroblastoma often can be treated successfully and that treatment success is higher with early-stage disease? The answer is in part because many cases of neuroblastoma spontaneously regress in the absence of treatment and in part because the therapy itself is highly toxic. For every child in these clinical trials saved by early detection and treatment, there was apparently another child who died as a result of therapy despite a tumor that would have regressed on its own [43].

It is not surprising that of the large number of laboratory tests that have been proposed for cancer screening, only a small handful are currently recommended in national screening guidelines [48,49].

**Use of multiple biomarkers for a disease**

Another special case involves the situation in which a large number of markers are available for a particular disease. Coronary artery atherosclerotic disease (CAD) is an excellent example: proposed markers include total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, high-sensitivity C-reactive protein, homocysteine, e-selectin, fibrinogen, d-dimer, PAI-I, ICAM-1, interleukin-6, herpes simplex virus-1 antibody, cytomegalovirus antibody, folate, and others [32,33]. Each of these markers has been shown in one study or another to be an independent predictor of CAD-related events. As of this writing, however, neither the US Preventive Services Task Force nor the National Cholesterol Education Program recommends laboratory markers beyond total cholesterol, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol [49,50]. Why would it not make sense to use a much broader panel to assess CAD risk?

Suppose a clinician were to order ten of the previously mentioned markers on a 50-year-old woman who does not have diabetes and does
not smoke but whose father died at age 65 of a heart attack and whose mother died at age 75 of cancer. Suppose that five of the serum markers were within the normal range, two had results that suggested lower-than-normal risk, and the remaining three had results that suggested higher-than-normal risk. How should the clinician integrate these results to determine how aggressively to treat the patient? If each of these markers is equally important, then “averaging” their interpretations would lead the clinician to conclude that the patient was of roughly average risk of CAD-related events. Are the markers really of equal importance? Are they really all independent of each other, or could a more limited set of markers give the same information as the expanded set?

Two recent studies attempted to answer these questions. As part of the ongoing Framingham Heart Study, Wang and colleagues [32] measured a set of ten biomarkers in addition to the “standard” risk factors of smoking, diabetes, hypertension, body mass index, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, age, and sex. Multivariate regression was used to quantitate the information added by the ten novel biomarkers, including C-reactive protein, B-type natriuretic peptide, and homocysteine. Death and major cardiovascular events were used as outcomes in the models. Although several of the markers were statistically significant independent predictors of death or major events, the receiver operation characteristic curves for the multimarker models were almost identical to those of the standard risk factor models. Using the C-statistic, age and sex alone predicted approximately 75% of death risk; the other traditional markers increased that to 80%. Adding the ten-biomarker panel increased it further, but only to 82%. Prediction power for major cardiovascular events was similar [32].

The Atherosclerosis Risk in Communities Study considered a larger but overlapping set of 19 novel risk markers and computed the area under the receiver operation characteristic curve to quantify the value that they added to a basic risk factor model. The basic Atherosclerosis Risk in Communities model was similar to that of the Framingham study. Most of the novel markers did not enhance the area under the curve to a statistically significant extent, and the remaining markers did so only slightly [33].

The problems with novel cardiac biomarkers are twofold. First, the information value that they add to traditional risk factors is, in absolute terms, limited. Second, accurate integration of results into a global risk assessment requires fairly sophisticated multivariate modeling such as used by the two research groups mentioned previously. A physician who uses the Framingham model [51] to assess risk based only on traditional risk factors almost certainly obtains a more accurate risk estimate than the physician who orders a much wider array of biomarkers but attempts to integrate the results through intuition.

**Closing the quality chasm related to spurious laboratory results**

A comprehensive discussion of improving physicians’ use of laboratory tests is beyond the scope of this article, but the practices in Box 3 can
help reduce the impact of false-positive and false-negative results. Note that they are best thought of as complementary rather than standalone practices.

Summary

Advances in clinical laboratory science have led to impressive tools for identifying and categorizing disease, which in turn drive improvements in therapy. Like all tools, however, laboratory tests can harm instead of help when they are not appropriately used. Given the limitations inherent in all laboratory tests together with the limitations in physician cognition, physicians should exert care in choosing tests and interpreting test results. This care includes awareness of tests’ clinical performance characteristics, judicious use of tests that have been well validated in clinical studies, avoidance of tests that lack good clinical validation, and probabilistic thinking about the role of a given test in each individual scenario. These decisions will likely lead to less overall testing but more accurate and efficient diagnosis and management of patients.

References


Quality Improvement Opportunities in Blood Banking and Transfusion Medicine

Jay P. Brooks, MD, MBA a,b,c,*

a Department of Pathology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd., Room 451, Oklahoma City, OK 73104, USA
b Transfusion Medicine, OU Medical Center, 1200 Everett Drive, Oklahoma City, OK 73104, USA
c Department of Veterans Affairs Medical Center, 921 NE 13th Street, Oklahoma City, OK 73104, USA

Making blood transfusion safer has been the top priority for blood banking and transfusion medicine. Blood collection centers have developed donor selection strategies and implemented infectious disease testing that have made the blood in the bag safer than ever. Pathogen-inactivation methods probably will improve future safety further, and efforts to reduce transfusion-related acute lung injury (TRALI) are currently at the fore. Hospital transfusion medicine services, charged with getting the right blood to the right patient, are beginning to use electronic identification technology and create positions such as transfusion safety officer to reduce mistransfusion.

With the advent of blood banks and subsequent advances in the collection and preservation of blood, the transfusion of blood allowed the performance of blood-losing surgery and the support of rigorous cancer treatments, as well as the support of patients who have hematologic illnesses. As the life-saving functions of transfusion were realized, so were the associated hazards inherent in this most widely performed type of tissue transplantation. This article reviews the quality initiatives that have been used to make blood transfusion safer and offers suggestions for further quality improvement (Box 1).

Blood transfusion safety encompasses having a safe product in the blood bag and transfusing this scarce resource into the right patient in appropriate quantities. The blood in the bag is purer than it has ever been because of
remarkable improvements in both donor selection and the testing of donated blood. Current efforts to reduce exposure of recipients to plasma from female donors will reduce one of the most serious residual transfusion risks, TRALI. On the hospital blood transfusion side, the main efforts have been in the area of patient, patient sample, and blood bag identification. Electronic methods of patient and blood identification—barcoding and radio frequency (RFID) identification devices—have the potential to reduce significantly the incidence of misidentification and ABO hemolytic transfusion reactions.

**Blood collection quality initiatives**

*Hepatitis*

In the 1960s, it became apparent that many patients who had received blood transfusions developed hepatitis. Close examination revealed that the incidence of posttransfusion hepatitis was fivefold higher in patients who had received blood from paid donors than in patients who had received blood from volunteer donors [1]. A prospective study in patients undergoing open-heart surgery in 1970 confirmed the much higher rate of posttransfusion hepatitis [2]. These seminal observations propelled blood collection centers in developed countries to move from the use of remunerated donors, and an all-volunteer blood supply has become a goal worldwide [3].

The cause of most cases of posttransfusion hepatitis was elucidated as the hepatitis B virus (HBV). A test for the surface antigen was developed and

---

**Box 1. Significant quality initiatives**

*Initiatives by blood collection centers*
- Volunteer blood supply
- Universal leukoreduction of cellular products
- Extensive infectious disease testing
- TRALI reduction initiatives
- Pathogen inactivation (future)

*Initiatives by hospital transfusion services*
- ABO/Rh and compatibility testing
- Electronic methods of patient/blood component identification
  - Barcoding
  - Radiofrequency Identification
- Creating the position of transfusion safety officer
- Irradiation of cellular blood products

---

*Key opportunities for improvement*
implemented as a screening test for donated blood. Combined with the elimination of paid donors, this advance further reduced the incidence of posttransfusion hepatitis to 7% of the previous rate [4]. With current serologic tests the risk of HBV infection is approximately 1:200,000 units of blood transfused (Table 1) [5]. HBV nucleic acid amplification testing (NAT) is being performed on a research basis in some blood centers, and it is likely that it eventually will become a Food and Drug Administration (FDA) licensed and required test in the future.

Although these steps dramatically decreased the incidence of posttransfusion hepatitis, a residual level of hepatitis remained. The incidence of this form of hepatitis, termed “non-A, non-B hepatitis” (NANB), was reduced by the introduction of testing for elevated levels of the liver enzyme alanine amino transferase (ALT) and antibody to hepatitis B core antigen. These surrogate markers reduced the incidence of this residual hepatitis by 30% to 50% [6]. Although ALT was effective in reducing the amount of NANB hepatitis, it was not specific, and many noninfectious causes, such as even moderate alcohol ingestion, could elevate ALT levels, resulting in deferrals of healthy donors. The causative agent of most cases of NANB hepatitis was later identified and designated “hepatitis C virus” (HCV). A test for antibody to this virus has further reduced this infection posttransfusion [7]. It also eliminated the need for the nonspecific ALT test. The addition of NAT to serologic testing has reduced the incidence of posttransfusion HCV infection to 1:2,000,000 units transfused [5].

Despite excellent laboratory tests, there is still a residual risk, due to the window period detection problem and testing errors. This is why a detailed donor questionnaire and honest donors are essential to making the blood supply as safe as possible.

Posttransfusion AIDS

There is no greater story of both tragedy and success in the history of transfusion medicine than that of transfusion-transmitted AIDS and the efforts to eradicate it. In the early 1980s, it was realized that patients who had received blood transfusions had been infected with the agent later identified as the HIV. In some areas of the country, it was estimated that 1 out of 100 blood products was infected [8]. Through 1999, 8910 patients were reported to have been infected with the virus through blood transfusion.
A combination of deferrals by blood collection centers and self-deferral by donors at risk significantly decreased HIV transmission by blood transfusion. Implementing a test for antibody to the virus further reduced transmission of this illness through transfusion.

Refinements to the antibody test have made it more sensitive, but the problem of the window period between the time of infection and the development of serologically detectable antibody response, during which infections eluding the tests can be transmitted, remained. Molecular testing offered further advances in closing the window period. An NAT for viral RNA can detect infections as soon as 11 days after infection. It is estimated that the residual risk of infection from tested blood from repeat donors is approximately 1 in 2,000,000; cases of posttransfusion HIV infection now are so rare that when they do occur they are national news. Antibodies to HIV1 and HIV2 are detected by laboratory tests that detect seroconversion to either strain of this virus. HIV type O is a subtype seen primarily in Africa that that cannot be detected by any currently available laboratory test. The donor questionnaire addresses this issue, however, and persons who were born in or who have had sexual relations with someone born in that region are indefinitely deferred from donating.

Every laboratory test has a failure rate, but the tests for HCV and HIV use separate antibody tests and NAT tests that provide some redundancy. Although viral variants, atypical seroconverters, and testing errors can cause false-negative tests for these viruses, the main remaining problem is the window period. Although there are refinements that could be made to NAT testing (eg, changing from testing of pools of 16 or 8 donors to individual donor testing), the cost would be high and the benefits miniscule. Eliminating the window period probably is not possible, so questioning donors for risk factors and eliminating those at risk from the donor pool remains an important quality initiative.

**West Nile virus**

The story of transfusion-associated West Nile virus (WNV) transmission shows how blood collection centers and regulators learned from past experience and promptly sprang into action to prevent infection by this agent that can cause neuroinvasive illness in immunosuppressed individuals, which includes many transfusion recipients. Transfusion-associated infection with WNV was recognized in 2002, and national blood donor screening was implemented in June 2003. Unlike HIV and hepatitis viruses, which result in chronic infections, WNV has a short asymptomatic period and a short infectivity period. Infection with WNV elicits an IgM response by day 8. WNV can be transmitted by blood only early in the infection; by the time IgM is serologically detectable, the disease probably cannot be transmitted. An NAT for WNV RNA that can detect infection while the disease is transmissible has been developed.
The risk of infection with WNV is related to the risk of exposure to its vector, the mosquito. It is a seasonal problem, and the length of the season varies with location. There is debate as to whether donated blood should be tested year round or if limiting testing to specific times, depending on the geographic region, is adequate. There is also disagreement as to the use of mini-pool versus single-donor NAT detection. Currently the most widely used approach is year-round mini-pool testing with individual tests implemented when epidemiologic evidence of a higher likelihood of infection in an area is seen.

**Chagas’ disease**

*Trypanosoma cruzi* is the protozoan parasite responsible for Chagas’ disease. Found primarily in Latin America, *T. cruzi* infections have been on the rise in the United States and Canada because of increased immigration. Transmitted by the feces of reduviid bugs, the lifetime risk of severe heart or intestinal problems in infected individuals is about 30% [14]. The FDA recently has licensed an antibody test kit for the organism, and the American Red Cross and Blood Systems have implemented its use. It remains to be seen if use of this test will become the standard of practice adopted by all blood centers.

**Universal leukoreduction**

One of the more controversial steps taken by blood collection centers has been the implementation of prestorage leukoreduction of all cellular blood components. Many arguments for and against this practice have been published [15–20]. A variety of claims have been made advocating this practice; the following advantages are generally well accepted:

1. Reduction in the transmission of viruses that reside inside leukocytes, such as cytomegalovirus
2. Decrease in febrile nonhemolytic transfusion reactions
3. Decreased HLA isoimmunization

Other, less well-established benefits include a decrease in transfusion associated immunomodulation, which has been cited as a cause of increased numbers of infections following transfusion, and a decrease in cancer recurrence rates in transfused patients.

Although there is little question that prestorage leukoreduction provides some benefit to certain groups of patients (eg, the immunosuppressed), there is disagreement regarding the benefits for all patients, leading many to argue that universal leukoreduction is very cost ineffective. Nonetheless, universal prestorage leukoreduction is performed in Canada and in the majority of blood collection centers in the United States.

**The donor questionnaire**

Although laboratory testing for infectious diseases is a mainstay of blood screening, the donor questionnaire remains an invaluable part of the
selection process. For diseases that do not have currently available tests for detection, such as variant Creutzfeldt-Jacob disease (vCJD) donor questioning for risk factors is the only method, albeit an imprecise one, to reduce the risk of transmission of infectious agents. Even for transmissible illnesses that have sensitive tests for detection, the donor questionnaire is still important. As Solodan [21] states, “As long as there are marker-negative windows during infection, and the testing process is fallible and does not address all infections that may harm recipients, both the incidence and the prevalence of blood-borne infections in blood donors remain important.”

Caused by infectious protein particles termed “prions,” vCJD is a fatal neurodegenerative disease with a long incubation period that has been implicated in posttransfusion illnesses in Great Britain. vCJD is a transmissible spongiform encephalopathy, and the causative agent is thought to be the same as in bovine spongiform encephalopathy (“mad cow” disease). There is no currently available laboratory test for screening blood donors, and the FDA has enumerated donor deferral criteria for persons who have lived in or visited the United Kingdom or Europe for extended periods of time to reduce the possibility of transmission-related exposure [22].

In its most recent guidance document, the FDA estimates that about 5% of donors will be deferred, with the deferral rate reaching up to 10% in large coastal cities [23]. Particularly disturbing is that many of those eliminated have been donors who developed the habit of donating during military service and became frequent repeat donors. Research on tests that may be able to detect abnormal prions and devices that would filter out infective prions is promising, and laboratory tests to detect infection also are being developed [24,25]. If either strategy promises virtual elimination of the risk, perhaps current donor restrictions will be relaxed in the future. The FDA, however, is very conservative in making decisions that might put the blood supply at increased risk, and the evidence of efficacy would have to be very convincing.

**Bacterial contamination of platelets**

In terms of the transmission of infectious diseases, the most dangerous component issued from blood banks is platelets. Stored at room temperature up to 5 to 7 days, platelets have a serious risk of bacterial growth and subsequent sepsis in transfused patients. The risk of serious illness and death is particularly high because many platelet recipients are cancer patients who are significantly immunosuppressed.

It is estimated that more than 10 million units of platelets were transfused in 2001 [26], with 1 in 3000 platelet units being bacterially contaminated. Bacterial contamination of platelets can result in life-threatening bacteremia in 1 out of 100,000 transfusions, with fatalities occurring in 1 of 500,000 transfusions [27].

The American Association of Blood Banks (AABB) has instituted requirements for accredited institutions to reduce the incidence of bacterially contaminated platelets. AABB Standard 5.1.5.1 requires the use of
bacterial detection procedures for every platelet unit [28]. Culture-based bacterial detection is used for apheresis platelets, whereas whole blood derived platelets can be tested for pH and/or glucose before release from transfusion services. Clearly, the culture-based methods are superior to pH or glucose tests but currently are best suited for apheresis platelets. Studies have shown that culturing platelets has resulted in the prevention of the transfusion of bacterially contaminated platelet units [29,30].

Transfusion-related acute lung injury

In 1985, Popovsky and Moore [31] published findings from the Mayo Clinic describing lung injury following the transfusion of plasma containing blood components, but immediate attention to this entity was diverted for a number of years by the focus on posttransfusion AIDS. TRALI is characterized by acute hypoxemia accompanied by bilateral pulmonary infiltrates occurring within 6 hours of receiving plasma containing blood components in the absence of evidence of circulatory overload. Although the exact mechanism of TRALI is not known, one popular hypothesis involves the sequestration of neutrophils by the pulmonary endothelium caused by the patient’s underlying medical condition; it is suggested that these neutrophils subsequently are activated by the infusion of substances in the donor plasma, such as antibodies against recipient leukocyte antigens (HLA or human neutrophil antigens) or other biologically active substances. The activation of these neutrophils results in damage to the pulmonary endothelium, capillary leak, and acute lung injury. TRALI is similar to the adult respiratory distress syndrome but has a better prognosis, usually responding to ventilatory support and other supportive measures [32–35].

In 2003, TRALI overtook acute hemolytic transfusion reactions as the most common cause of posttransfusion death reported to the FDA [36]. The exact incidence of TRALI is not known, but in the United States it has been estimated to be between 1 in 5000 and 1 in 1323 plasma-containing transfusions [31,37,38]. Leukocyte antibodies are seen frequently in the plasma of women who have been pregnant multiple times. Great Britain’s National Blood Service has taken the step of eliminating female donors from donating plasma, although they are allowed to donate red blood cells and platelets. The AABB is pushing to reduce the incidence of TRALI by moving toward limiting transfusion of plasma and platelets donated by women [39] and has promulgated the following guidelines:

1. Blood collecting facilities should implement interventions to minimize the preparation of high plasma volume components from donors known to be leukocyte-alloimmunized or at increased risk of leukocyte alloimmunization.
2. Blood transfusion facilities should work toward implementing appropriate evidence-based hemotherapy practices to minimize unnecessary transfusion.
3. Blood collection and transfusion facilities should monitor the incidence of reported TRALI and TRALI-related mortality.

With the exception of the requirement that hospital transfusion services implement good utilization review practices, the AABB stops short of requiring specific strategies to meet its goals. It does, however, enumerate several suggestions for how blood collection facilities could reduce the risk of TRALI:

1. Use plasma exclusively from male donors for high plasma volume components (plasma, apheresis platelets, whole blood).
2. Use plasma from female donors who have never been pregnant or who test negative for leukocyte antibodies.
3. Obtain a lifetime transfusion history from all donors and exclude donors who have a history of transfusion from donating high plasma volume components or test these donors for leukocyte antibodies.

The AABB is requiring full implementation of these measures pertaining to plasma and whole blood by their member organizations by November 2007. Full implementation regarding whole blood and platelet components is required no later than November 2008 [39].

Pathogen inactivation and the future of infectious disease testing

In addition to its lifesaving properties, blood is, unfortunately, an excellent culture medium. Most agents capable of existing in blood are capable of being transmitted by blood transfusion. As new blood-borne pathogens emerge, they outstrip the ability of blood collection centers to devise effective new screening tests. Additionally, the cost of adding test after test threatens to push the price of blood perilously high. Every laboratory test has a false-positive rate; therefore, as laboratory screening tests proliferate, the likelihood of an uninfected unit of blood being discarded needlessly increases dramatically. The donor questionnaire has an even higher false-positive rate, and, when effective laboratory tests do not exist, expanding donor deferrals reduces the donor base and excludes more healthy donors.

Pathogen inactivation is a proactive process aimed at a broad spectrum of infectious agents. Most pathogen inactivation techniques involve targeting the nucleic acid of pathogens to render them incapable of reproducing and are useful against many viruses, bacteria, and protozoa. Agents that have been studied include amotosalen, Inactine (Amersham Pharmacia Biotech, Uppsala, Sweden), methylene blue, and solvent detergent [40,41]. The mechanism involves adding a compound to blood components that, when activated, binds to nucleic acids and inhibits replication. Most of the added compound is ultimately removed by absorption or washing before transfusion. Although certain of the compounds have been shown to be nontoxic in the short term, concerns over long-term toxicity remain.
The solvent detergent process has been applied to fresh frozen plasma in the United States, Europe, and elsewhere. It disrupts the membranes of several enveloped viruses but fails to inactivate non-enveloped viruses such as hepatitis A. because of its high cost and concern about its potential for inducing thrombotic events, solvent detergent—treated fresh frozen plasma is no longer used in the United States but continues to be used in other countries [40].

Pathogen inactivation has the advantage of being effective against a broad spectrum of infectious agents, but not all infectious agents are readily inactivated by the proposed processes. Some combination of laboratory infectious marker screening combined with broad pathogen inactivation probably will be the paradigm of the future.

Transfusion-associated graft-versus-host disease

Transfusion-associated graft-versus-host disease (TA-GVHD) is a usually fatal transfusion complication caused by transfusion of T lymphocytes present in cellular blood components (whole blood, red cells, platelets, granulocytes) that proliferate in the recipient and attack the recipient’s organs [42]. Recipients who are at risk for TA-GVHD are those who are profoundly immunosuppressed and whose immune systems are unable to destroy the invading T lymphocytes. Also at risk are immunocompetent recipients whose immune systems are unable to recognize the invading T lymphocytes as foreign. Patients who have congenital immunodeficiencies, Hodgkin’s disease, and those receiving fludarabine therapy are at risk of TA-GVHD because of their profound immunodeficiency. Patients who receive transfusion from donors with similar HLA haplotypes also are at risk, because the recipient’s immune system fails to recognize the donor T lymphocytes as foreign. In countries such as Japan, where there is a high degree of genetic homogeneity, TA-GVHD is a significant problem. Fortunately, gamma irradiation (25 Gy) of cellular components renders T lymphocytes incapable of proliferating and eliminates the possibility of TA-GVHD. In Japan, all cellular blood components are irradiated routinely. Patients who have the previously mentioned risk factors should receive irradiated cellular blood products. Many cancer centers irradiate all cellular blood components because of the consequences of a patient who needs irradiated blood not receiving an irradiated product. Directed donations from first-degree relatives as well as cellular blood components that have been HLA matched or cross-matched are routinely irradiated. Many larger hospital transfusion services have their own irradiators, and blood centers irradiate for smaller facilities.

Forces for quality initiatives in blood collection centers

The FDA is the regulatory agency that sets the minimum standards that blood collection centers in the United States must meet. Blood centers are
free to apply standards that are more rigorous, be they laboratory screening tests, manipulation of blood components, or blood donor health inquiries. For example, although the FDA does not mandate universal prestorage leukoreduction of cellular blood components, a majority of blood centers have opted to leukoreduce their inventories.

The AABB also sets standards for blood collection centers and transfusion services through its voluntary accreditation program. Approximately 99% of blood centers in the United States are accredited by the AABB (personal communication, Holly Rapp, AABB Accreditation Department, Bethesda, MD, March 2007), which gives it considerable authority over their practices. The AABB has required blood centers to take actions to reduce bacterial contamination of apheresis platelets and has mandated actions to reduce the incidence of TRALI. Through its private voluntary accreditation program, the AABB frequently can move more quickly than the FDA, which is constrained by its massive bureaucracy and compliance with governmental regulations.

Actions of large blood collection centers can have the effect of setting a standard of practice. Providing about half of the nation’s blood supply, the American Red Cross wields considerable influence in setting standards. Certainly its decision to leukoreduce its cellular blood products prompted other centers to follow suit. The American Red Cross and another large provider, United Blood Services, have decided to implement Chagas’ testing, and this decision may play a significant role in the decisions of other blood centers to implement the test.

### Quality initiatives by hospital transfusion services

Great strides have been made by blood collection centers in increasing the pharmaceutical purity of the blood in the bag, but it is up to the hospital transfusion medicine service to get the right blood to the right patient in the right amount. The biggest problem transfusion services face is that of mistransfusion: a patient getting the wrong blood product. The product for which the consequences of mistransfusion are most ominous is red blood cells. Transfusion of ABO-incompatible red blood cells can induce an immediate, complement-mediated, intravascular, hemolytic transfusion reaction that can lead to disseminated intravascular coagulation, renal failure, and death. A study of transfusions over a 10-year period in New York State found the incidence of ABO-incompatible red cell transfusions to be 1 in 38,000 red cell transfusions [43]. This risk is quite high compared with the relatively low risks of viral transfusion and represents an area in which resources can be profitably invested, as opposed to pursuing further reductions in viral transmission that result in miniscule returns on investment [44,45].

Although the hospital transfusion service performs tests to ensure compatibility between recipient and donor and a myriad of quality control and quality assurance activities to minimize possible errors, these efforts
are complicated by the fact that most errors that result in incompatible transfusions occur outside of and, to a great extent, beyond the control of the transfusion service [46,47]. Before the receipt of the pretransfusion specimen in the transfusion medicine service, errors can occur in the identification of the recipient at the time of collection of the pretransfusion specimen and in the labeling of the specimen at the time of phlebotomy. After compatibility testing is completed and the blood component is issued, personnel involved in transfusing the patient must assure that the blood is transfused to its intended recipient. Obviously a multitude of opportunities for error exist in this process as well.

**Better methods of patient and blood product identification**

Traditionally hospitals have relied on personnel to be diligent in identifying patients when collecting pretransfusion samples and when transfusing blood. In dealing with problems of identification, transfusion services can learn from other industries that deal with identification of their products.

**Transfusion safety officer**

Because the process of getting the right blood to the right patient spans multiple hospital services, some institutions have created a boundary-spanning position called a “transfusion safety officer” (TSO). TSOs currently are used at facilities in England and France as well as in the United States, where both Dartmouth Hitchcock Medical Center and the Puget Sound Blood Center employ them [48,49]. The background of the individual depends on the institution’s preference, but knowledge of transfusion medicine is a key component, whether the person is a medical technologist or a nurse. The TSO works primarily outside the confines of the transfusion medicine service, interacting directly with the people who perform pretransfusion phlebotomies and those who actually infuse the blood. TSOs also play a critical role in appropriate usage of blood products by overseeing the process of utilization review.

**Barcoding**

Retail establishments ranging from the corner convenience market to the “big box” stores use electronic barcoding to manage inventory and assure that prices are properly charged. Similar barcoding technology is now in place in some hospitals for blood transfusion, but the percentage is low—estimated to be approximately 1.5% [50].

With a barcode system, a patient wears an armband containing a bar-coded bracelet, and the blood sample tube is labeled with an identical barcode, positively associating the patient with the pretransfusion blood sample. The barcoded specimen label containing patient and transfusion information can be scanned into the hospital and laboratory information
system on its arrival into the transfusion medicine service. A unit of compatible blood labeled with an identical barcode is issued. The barcoded blood bag is scanned at the patient’s bedside, and the patient’s barcoded identification bracelet is scanned also. If the identity of patient and blood bag is confirmed, the transfusionist begins the infusion. If the system indicates there is not a match, the transfusion is aborted and an investigation of the discrepancy is begun [49–52].

Radiofrequency identification devices

Although barcoding can greatly improve patient and blood unit identification, the future belongs to RFID devices. RFID is a technology that uses tags that can identify objects by the use of radio waves emitted by the devices and read wirelessly by receivers [53]. This technology has several inherent advantages. Barcoding requires optical reading and line-of-sight recognition with a scanner. The use of radio waves overcomes these limitations. RFID tags can be implanted in wristbands, laboratory test tubes, and blood bags [49]. A number of institutions worldwide are evaluating the potential of RFID technology. At the Massachusetts General Hospital, RFID antennas have been placed on operating room tables; when an incompatible unit of blood comes near the patient, the transfusionist is alerted not to use the blood [54].

RFID tags about the size of an uncooked grain of rice have been widely implanted in domestic pets for identification purposes. RFID tags can be part of a wristband; similar tags can be implanted in humans as well [55]. Such implanted tags have been heralded as helpful in identifying unconscious patients presenting to emergency rooms and in identifying patients who have diminished mental capacity (eg, patients who have Alzheimer’s disease who have wandered from their homes or care facilities) [56]. A pilot study of RFID implants in patients and the use of reading technology in emergency rooms is under way in New Jersey [57].

RFID has advantages beyond the very important function of patient identification. Blood bags containing RFID tags can be monitored continuously by antennas in refrigerator doors to give the transfusion medicine service an up-to-the-second accurate inventory of blood available within the service and present in any remote refrigerators. RFID tags also can be linked to devices that can incorporate constant temperature monitoring to ensure blood components are not exposed to unacceptable temperatures.

Forces for quality initiatives in hospital transfusion medicine services

The FDA historically has been much more interested in regulating the activities of blood collection centers, concentrating on the pharmaceutical purity of the blood in the bag through its program of good manufacturing processes, than in overseeing the practices of hospital transfusion services.
The FDA plays a critical role in inspecting all centers that collect blood, and its findings often are widely publicized and may involve legal actions. The FDA, however, inspects hospital transfusion services only when they perform significant modifications to blood components such as the irradiation of cellular blood products, or when they are government entities. Therefore, the FDA inspects only a minority of transfusion medicine services, and when inspections do occur, they usually are completed within hours to a few days, compared with the several weeks of intense scrutiny to which blood collection centers often are subjected.

The FDA’s interest in hospital transfusion services has expanded recently with its requirements for barcode labeling [58]. Included in the same rule that requires barcode labels on drugs, this requirement affects blood collection centers and all transfusion services that perform any manipulation, such as pooling of cryoprecipitate or platelets, or that aliquot blood for transfusion. This broad range of services targeted by the FDA indicates that it intends this rule to affect the majority of transfusion medicine services in the United States. The FDA rule requires every blood bag to contain machine-readable information, including a unique facility identifier, lot number relating to the donor, product code, and donor ABO and Rh type.

Whenever a modification, such as pooling or preparing an aliquot, is performed, a new barcode label from the hospital transfusion service performing this task must be applied to the component. The FDA describes its reason for requiring this barcoding:

The rationale behind the bar code rule is to allow for positive identification of the patient and the blood component at the patient’s bedside. If machine-readable information is not on the component for positive identification just before administration at bedside, the main objective for requiring the bar code is lost [59].

The FDA does not specify what symbology must be used in the barcode labeling. The International Society of Blood Transfusion (ISBT) has developed barcode labels, termed “ISBT 128,” proposed for universal use within the field of blood banking and transfusion medicine [60]. The AABB is requiring its member institutions, both blood collection centers and transfusion medicine services, to implement the United States Industry Consensus Standards for the Uniform Labeling of Blood and Blood Components using ISBT 128 [61]. The 24th edition of the Standards for Blood Banks and Transfusion Services (BB/TS Standards), which became effective November 1, 2006, required that facilities have a written plan for the implementation of ISBT 128. The 25th edition of the BB/TS Standards, which becomes effective September 1, 2008, requires full implementation of ISBT 128 by accredited facilities.

Because the AABB accredits virtually all blood collection centers in the United States, the ISBT 128 system soon will be in place in collection facilities. Although the AABB accredits only approximately 30% of the hospital transfusion medicine services in the United States (personal communication,
Holly Rapp, AABB Accreditation Department, Bethesda, MD, March 2007), the FDA will require most transfusion services to have barcoded labels because they modify products, and a transfusion medicine center would be unlikely to adopt a barcode symbology other than ISBT 128 by.

Summary

Great strides have been made in improving transfusion safety. On the blood collection side, declines in infectious disease transmission are attributable to better donor selection and infectious disease screening. The future is likely to include processes for broad-spectrum pathogen inactivation.

For transfusion medicine services, integration of new electronic technology has the promise of reducing patient and blood misidentification and further reducing morbidity and mortality from acute hemolytic transfusion reactions. Efforts to reduce TRALI are being put in place that will decrease this cause of transfusion-related illness. Use of TSOs has the potential of integrating complex processes and improving the prudent use of this life-saving resource.

References


[50] Landro L. Blood mix-ups can be deadly, but they also can be prevented. The Wall Street Journal (Eastern edition) February 27, 2003; D4.


Quality Improvements in the Preanalytical Phase: Focus on Urine Specimen Workflow

Ana K. Stankovic, MD, PhD, MSPH\textsuperscript{a,*}, 
Elizabeth DiLauri, MBA\textsuperscript{b}

\textsuperscript{a}Medical and Clinical Affairs, BD Diagnostics, Preanalytical Systems, 
1 Becton Drive, Franklin Lakes, NJ 07417, USA 
\textsuperscript{b}Alternate Specimen Management, BD Diagnostics, Preanalytical Systems, 
1 Becton Drive, Franklin Lakes, NJ 07417, USA

The preanalytical phase of the total testing process is complex. It starts with the perceived need for the test and ends with specimen processing \[1\]. It is not surprising that between 32\% and 75\% of all testing errors occur in the preanalytical phase \[2\]. When one considers the amount of automation and process controls that exist in the analytical phase versus those in the preanalytical phase, it becomes apparent why this is the case. Not all preanalytical errors cause adverse events, because many of these “upstream” errors may be caught during downstream processes or are minor enough that, if undetected, they do not impact patient outcomes \[2\]. They are often associated, however, with rework or further investigations, which result in unnecessary risk to the patient and unjustifiable costs to the health care system.

For a long time, preanalytical improvements have centered on blood specimens, mostly driven by increasing levels of automation and need for standardization. Urine collection and processing often have lagged behind and represent areas with much opportunity for improvement. Although still predominantly a manual process, urine testing is currently facing increasing levels of automation. As a result, in most markets automated urinalysis is growing faster than automated chemistry (22\% growth rate in 2004). The growth of urine testing is mainly driven by automated microscopy in developed countries and traditional urine chemistry in developing markets \[3\].

---

The views expressed in this article are those of the authors and not of Becton, Dickinson and Company.

* Corresponding author.

E-mail address: ana_stankovic@bd.com (A.K. Stankovic).

These advances in urine testing and the fact that urinalysis remains one of the three major in vitro diagnostic screening tests—after serum chemistry profiles and complete blood counts [4]—create the need to take a hard look at the urine testing workflow and underscore the importance of reducing preanalytical variability.

There are significant opportunities for process improvement, efficiencies, and cost reductions in the area of urine testing. If one assumes that worldwide health care spending on urine testing is approximately $566 million annually [3] and assumes an effectiveness level of 95%, the estimated global cost of inefficiencies in urine collection, testing, and analysis is nearly $30 million per year. This figure generates a new degree of urgency to focus preanalytical improvement efforts on processes associated with this area of laboratory testing.

**Preanalytical phase of urine testing**

The preanalytical phase in urine testing can be divided into six major subphases (Fig. 1):

1. Test ordering
2. Sample collection
3. Specimen transport to the laboratory
4. Specimen receipt in the laboratory
5. Preparation of sample for testing
6. Transportation of samples to the section of the laboratory where the testing occurs

Each of these subphases contains between two and five steps, so the average preanalytical urine testing workflow consists of at least 22 steps. Some of these steps, such as sample collection, can be subdivided into activities that further complicate the urine preanalytical testing process.

Fig. 1. Preanalytical urine specimen workflow.
Urine specimen collection process

Urine specimen collection varies considerably, depending on the setting in which the specimen is collected. The variability in how specimens are collected—where, when, and by whom—results in a wide range of different activities within this preanalytical step and increases the likelihood of errors. For a urine specimen requisition in an inpatient setting, the requisition is provided to the nurse or appropriate health care provider on the unit notifying them of the test order. The health care provider gathers the necessary materials for specimen collection and enters the patient room at the appropriate time based on the requisition. Subsequent steps differ depending on the age and status of the patient.

For an adult patient who does not require assistance, the health care provider should provide the patient with instructions on how to perform a proper collection (ie, a midstream, clean catch) and the required materials to capture the specimen upon voiding. The patient performs the required cleansing and voids, as appropriate, into the container provided. For a dipstick test, the health care provider may insert the dipstick into the specimen at the site of collection and read and record the results into the patient’s medical record. At that point, the urine specimen and container may be disposed of in the patient’s bathroom. These steps for urine collection and handling also may be followed in an outpatient clinic setting.

For a patient who is bedridden or cannot urinate independently, a health care provider inserts a Foley catheter into the bladder through the urethra to collect the urine specimen. Alternately, specimens may be collected directly from a Foley catheter into an evacuated tube or transferred from syringe into a tube or cup [5]. For infants and small children, a special urine collection bag is adhered to the skin surrounding the urethral area. Once the collection is completed, the urine is poured or transferred directly into a collection cup or transferred directly into an evacuated tube with a transfer straw [5].

For a urine test that requires processing by the laboratory, the health care provider should transfer the specimen into a safe, clean transport container (eg, a tube) and label the specimen container at the point of collection. Care should be taken to ensure that the specimen is labeled correctly at the point of collection and transferred into a transport container that is also appropriately labeled. Many urine specimen collection cups are not designed to sustain transport within a hospital’s pneumatic tube system. Multiple systems are available to enable a closed transfer of the specimen from the collection cup into a closed, transport-safe container, including transfer straws and cups with lids that contain integrated transfer devices. These systems ensure that the health care provider’s risk in handling a potentially contaminated specimen is reduced and that there are no opportunities for introducing contaminants into the specimen during transport.
Preanalytical variables in urine testing

Common areas of preanalytical variability in urine testing include patient-related factors, specimen collection, specimen identification and labeling, specimen transfer and transport, and specimen processing. Their effect on urine testing is described here in more detail.

**Patient factors**

Patient factors (eg, diet, medications) can impact various urine test results, such as color, specific gravity, pH or clarity, and can cause testing errors [6]. For instance, consumption of beets and rhubarb can result in change of urine color to red. Presence of high concentration of ascorbic acid in urine has been linked with false-negative urine glucose and bilirubin dipstick results. Similarly, tetracycline therapy has been shown to cause false-negative urine glucose dipstick results.

**Specimen collection**

Specimen collection is an important source of preanalytical variability. The collection method affects the quality of urine specimens and, if selected properly, decreases the risk of specimen contamination and health care workers’ exposure. For urine culture and sensitivity testing in particular, use of the midstream, clean catch method is preferred because it reduces the incidence of cellular and microbial contamination [5]. Appropriate collection containers or tubes can decrease the possibility of specimen leakage in transit, particularly in the pneumatic tube systems, and ensure optimal quality and quantity of specimen for analysis. Collection of unpreserved urine specimens is more likely to result in bacterial overgrowth if the specimens are not refrigerated, as refrigeration decreases urine clarity and causes false-positive nitrite and false-negative glucose results [6]. It is also important to avoid introduction of contaminants during specimen collection. In obtaining pediatric urine specimens, for example, collecting urine from a diaper can introduce contamination from the diaper material, which may affect the test result.

**Inaccurate labeling of a urine specimen**

Inaccurate labeling of a specimen with patient identification, date, and time of collection and suboptimal placement of the label can compromise the laboratory’s ability to process a urine specimen and obtain a correct result. If not recognized in time, inaccurate patient information can result in laboratory errors. On the other hand, if recognized before the onset of testing, it slows the analytical process by requiring either specimen recollection or further physician approval before analysis, thereby delaying reporting of results. Improper label placement can lead to misidentification of the specimen later in the handling and analytical process if the specimen container lid with the label is removed or if the label does not adhere to the container during refrigerated conditions [5]. As a result, the primary goals listed in the Joint Commission’s 2008 National Patient Safety Goals for the Laboratory
include the use of at least two patient identifiers when providing care, treatment, or services and establishing processes for maintaining a specimen’s identity throughout the preanalytical, analytical, and postanalytical processes [7]. Use of evacuated tubes that can be labeled at the bedside reduces the risk of identification errors. It is important to realize that misalignment of barcoded labels on tubes also can be a source of identification errors, however. Instrument barcode readers cannot “read” the barcodes on misaligned labels; sometimes it is enough for a label placement to deviate only a little from the ideal position to present a problem. As a result, the misaligned label must be removed and replaced with a properly placed label, which creates the possibility for an identification error to occur.

**Transport of urine specimens**

Improper transport of urine specimens can impact sample quality directly. Increased length of time between specimen collection and analysis, lack of temperature control, and use of nonpreserved specimens that will not be analyzed within 2 hours of collection (eg, samples that come from satellite locations or are collected during a 24-hour collection) contribute to overgrowth of bacteria in the specimen. This can, in turn, impact urinalysis and culture and sensitivity test results. To prevent this from happening, the College of American Pathologists’ accreditation program requires examination of nonpreserved, nonrefrigerated urine within 1 to 2 hours of collection and defines noncompliance with this standard as a phase II deficiency that requires immediate and documented improvement before accreditation is granted [8,9].

If a laboratory cannot analyze a nonrefrigerated, nonpreserved specimen within the recommended time frame, however, other acceptable options can ensure optimal specimen quality. One of them is the use of nonrefrigerated, preserved urine specimens that can be analyzed up to 48 hours after collection for culture and sensitivity and up to 72 hours for urinalysis (manufacturers’ recommendations should be reviewed for product-specific guidance) [10]. Clinical and Laboratory Standards Institute (CLSI) recommends that refrigerated, unpreserved specimens be maintained at 2°C to 8°C. When preparing for analysis, it becomes equally important to allow refrigerated specimens to return to room temperature in order to enable temperature-dependent enzymatic reactions to occur during analysis [6] and to prevent any false signals from crystals formed during refrigeration. When preserving specimens for transport and analysis, use of an evacuated container helps to ensure that the proper preservative:specimen ratio is adhered to, because evacuated tubes are designed to draw a specific sample volume [5].

**Specimen processing**

Variables in specimen processing, such as centrifugation, impact the quality of results. Variation of speed and time can change the cellular elements obtained in the sediment.
Additional factors

Other factors can contribute to preanalytical error. The chronic struggle of laboratories to address ongoing labor shortages, budget reductions, and increasing test volumes within their own walls can impact the quality of the results the laboratory is able to generate and the revenue or reimbursement that the laboratory is able to capture. Specifically, high turnover and lack of training and education on how to properly collect and handle urine specimens can result in higher specimen contamination rates.

Opportunities for preanalytical process improvement

The various methods through which a urine specimen can be obtained and their manual nature can contribute significantly to preanalytical errors. The fact that many of the steps in the preanalytical phase are performed outside the walls of the laboratory and often by personnel not managed by the laboratory creates further challenges for the laboratory in controlling the steps that contribute to preanalytical variability. The laboratory is often accountable for the outcome (clinical and financial) of this phase and for resolving—if not overcoming—the preanalytical challenges. At the same time, clinical laboratories are constantly faced with pressures to continually increase productivity, lower costs, and improve quality. To address these challenges, the laboratories can use several cost-effective tools designed to standardize and optimize urine collection and handling and positively impact the quality of urine specimen results.

Application of Lean management methodologies

To meet the increasing pressures that the clinical laboratories are currently facing, many laboratory administrators and pathologists are faced with the fact that radical increases in quality, productivity, and error reduction cannot be achieved using the traditional management models. “By contrast, the pioneering laboratories in the United States that use Lean and Six Sigma to redesign workflow in their high-volume core chemistry and hematology labs found that these quality management systems—in a 12- to 16-week project—could lead to a 50% reduction in average test turnaround time for a hospital lab, a 40 to 50% improvement in labor productivity and a comparable improvement in quality of results” [11]. When considering the more highly manual processes involved in urine specimen collection and handling, one also could hope to achieve similar benefits from application of Lean and Six Sigma principles to the urine collection, handling, and processing areas.

Lean management centers on reducing waste and offers a set of tools that do not require significant capital investments to be effective and can have immediate and recognizable impact when applied properly. Six Sigma is a metric and a methodology that focuses on reducing variability (ie, counting and
decreasing the number of defects in a process). A process that performs at a Six Sigma level realizes defects at a rate of 3.4 occurrences per million opportunities. Used in combination with Lean tools, Six Sigma methodology can enable an organization to monitor and improve its quality performance based on the elimination of errors where possible (ie, Lean) and reduce or manage them in parts of a process that cannot be eliminated (ie, Six Sigma). For the purposes of this discussion, we focus specifically on the application of Lean principles to the preanalytical process in urine specimen collection.

Simply put, the fundamental principles of Lean management emphasize reduction of unnecessary and non–value-added activities to reduce total production time and effort. Lean tools focus on identifying steps that are error prone and must be controlled if they cannot be eliminated altogether. Lean management principles challenge doing things the way they always have been done in favor of simplified and standardized methods of performing tasks that support getting things done right the first time with minimal wasted time and effort. One highly effective tool in applying Lean thinking is the process map, which is constructed by recording all steps in a current process or subprocess and illustrating the current state (ie, “as is”). Lean tools, such as process mapping, enable organizations to take a critical inventory of the activities taking place within their laboratories and activities external to the clinical laboratory but relevant to testing processes. Armed with knowledge of what is actually taking place, as represented by a process map, laboratory management is better equipped to identify how the process should look (ie, the “to be” process) in order to obtain better and more productive outcomes.

“To be” processes are streamlined processes that are flexible, reduce waste, optimize the process, improve process control, and improve use of resources. For the clinical laboratory, improved use of limited resources is a continuous quest, and the benefits to the laboratory of application of process mapping and Lean thinking becomes clear. Fig. 2 provides a simplified example of a current state process workflow analysis, focused on the subprocess within the laboratory from urine specimen accessioning to specimen disposal, at a hospital before a Lean workflow implementation. All of the major process steps are captured, regardless of whether they support good practice or whether they represent value-added activities or contribute to delays in providing results. Steps that represent potential for error and steps that do not contribute value to the target end product, or result, but consume time or resources are indicated as such.

Fig. 3 illustrates the impact of applying Lean management principles to the urine specimen collection and handling process shown in Fig. 2, or “to be” state. Non–value-added or error-prone steps are either removed or controlled in the revised process, resulting in a more efficient, timely, and standardized process. Specifically, by changing the method of urine specimen collection, 11 steps were removed from the process. All 3 of the error-prone steps were eliminated, 2 of them in the preanalytical phase, including the
open transfer of specimens into tubes from cups and the relabeling of specimens. This change created a safer environment for the laboratory staff and decreased the risk of testing errors for the patient. Five non–value-added steps were eliminated, 3 from the preanalytical phase, which increased the efficiency and reduced the waste of the laboratory staff and materials. Overall, the eliminated steps represented 64% of all steps in the process.

This example illustrates how the use of a simple tool, which does not require significant capital investment to execute, can lead to visible quality improvements by simplifying the preanalytical process in the laboratory and eliminating opportunities for errors.

Additional Lean tools, such as value stream mapping, also can be applied. This tool determines the amount of time taken for completion of the entire process, as well as each step in a process flow, and categorizes those activities into value-added and non–value-added activities. The total time to perform value-added and non–value-added activities is then quantified in the pre- and post-Lean implementation flow charts. The reduction in non–value-added time can be quantified as potential savings based on the wage data of laboratory and other hospital personnel involved in performing those steps. Based on the resulting total time indicated from collection to analysis, tools and products can be applied to the new process flow, such as the introduction of tubes with urine preservative. For example, if unpreserved, unrefrigerated urine specimens are not analyzed within 2 hours of collection, then the laboratory can further reduce its specimen errors and improve quality of results in the first attempt by refrigerating the specimen.
at 2° to 8°C or preserving specimens until the specimen can be analyzed appropriately [10].

**Additional tools available**

Additional tools are available to aid laboratories in standardizing work across various collection settings and among various individuals involved in performing urine collections.

**Devices**

Devices are available to provide further support to the laboratory in ensuring quality urine specimens. In manual and automated analytical settings, urine containers with preservative enable the laboratory to balance the flow of specimens requiring analysis that come into the laboratory with the available resources to complete the analysis in a timely manner. For organizations that currently face staffing challenges without adequate volume to justify automation, using preservatives with urine specimens enables the analytical workload to be balanced over the peaks and troughs during a longer timeframe—up to 72 hours for urinalysis and up to 48 hours for culture and sensitivity testing—without refrigerating specimens.

**Closed systems for transfer and transport of specimens**

Closed systems reduce the risk of exposure of health care workers to contaminated specimens and the exposure of specimens to contaminants during
transfer and transport. Amber-colored urine containers protect specimens from light sensitivity for certain urine tests (eg, bilirubin).

Instructions for clean, midstream catch

Instructions posted in areas where patients are voiding into urine containers—translated to languages other than English, if necessary—can reduce the risk of contamination of specimens and mitigate insufficient training caused by high turnover in staff.

Impact of the preanalytical process workflow on automation

Opportunities for automation of urine specimen analysis continue to improve, including the ability to perform urinalysis and microscopy of urine sediment on a single platform. Further opportunities include autoverification of results, which can eliminate the need for medical technologists to review thousands of normal test results before releasing them. Automation of urinalysis enables the clinical laboratories to absorb substantial volume increases without adding staff.

One should recognize that implementation of automation without improvement of the preceding preanalytical processes often creates a new set of issues for the laboratory to resolve. Experience shows that automating a bad process only serves to speed up problems and potentially magnifies the problem and its associated cost. The value of investing effort in streamlining or redesigning current preanalytical processes before implementation of automation should not be underestimated. Lean management and Six Sigma are only two examples of quality management methods that can be used to improve processes before investing in automation.

Business case for preanalytical process improvement

Before determining the appropriate steps that will lead to the reduction of preanalytical variability in urine specimen collection and handling, an institution must understand what factors impact the preanalytical phase of urine testing the most, in order to determine the true causes for preanalytical errors and apply the right solutions to get optimal results—the first time.

The business case for reducing preanalytical variables can be as simple as measuring the number of urine tests that do not get reimbursed because of preanalytical quality issues, such as urine culture contamination or urine bacterial overgrowth. At its most conservative, the lowest urine test reimbursement rate could be used, such as the $3.57 Medicare rate for manual urinalysis without microscopy. At the higher end, the $11.28 Medicare reimbursement for urine bacterial culture could be used [12]. These figures represent the billable charges that a laboratory could capture but did not because of lack of a result. This does not begin to factor in the lost productivity of employees who are already challenged in a resource-constrained laboratory.
environment. Whether viewed as a revenue opportunity or a cost-avoidance opportunity, a laboratory can do a simple calculation to qualify the value of controlling preanalytical variability.

Simple principles such as Lean—eliminating unnecessary activities and getting it right the first time—can be applied easily without sophisticated technology when the right focus and tools are applied. Given the pressures prevalent in the modern laboratory environment, laboratory management cannot afford to ignore the options available to control processes, drive efficiencies, and gain improved charge capture for urine specimens.

Primary constraints that laboratories currently face are personnel shortages and lack of financial resources. The business case for quality improvement is a critical component for justifying the resources required to implement change in the preanalytical process. Fortunately for laboratorians, cost-effective external resources are available to conduct the required analyses to identify an organization’s specific areas for improvement and implement the required changes and process controls. Laboratory success stories exist that support the cost justifications to pursue quality improvement, whether through Lean and Six Sigma applications, use of readily available devices, or other similar approaches.

Summary

In the past, laboratories have addressed issues of preanalytical variability in an opportunistic way, addressing discrete parts of the preanalytical process, such as patient identification, specimen rejection, and blood/urine culture contamination. To obtain needed quality improvements and error reduction, it is necessary to look at the preanalytical process as a whole—from test ordering to the moment the specimen is processed by the analyzer and apply process improvement methodologies, such as Lean and Six Sigma. To achieve this, laboratories should map the preanalytical phase in its entirety, identify steps that are potential causes of unnecessary variability that lead can to laboratory errors, and find ways to either remove them or error proof them. At the same time, by using this approach it is possible to reduce unnecessary waste and obtain needed process efficiencies.

References


The Physician Quality Reporting Initiative

Dirk M. Elston, MD

Department of Pathology, Geisinger Medical Center, 100 North Academy Avenue, Danville, PA 17821, USA

Key points about the Physician Quality Reporting Initiative

- The pay-for-performance initiative is being driven by major corporate health care purchasers.
- Medicare is offering incentive pay for those who report quality measures.
- Private payers may link patient copayments to reporting or performance.
- Measures are developed by specialty societies in cooperation with the American Medical Association Physician Consortium for Performance Improvement.
- Measures must be endorsed by a multi-stakeholder group such as the AQA Alliance (formerly the Ambulatory Care Quality Alliance) or the National Quality Forum.
- Measures must be evidence-based and supported by current vetted guidelines of care.
- There should be published evidence of an existing gap in care and variation in practice.
- Process measures, outcome measures, and structural measures are under development by various specialty societies.
- Some measures may address appropriate use, continuity of care, coordination of care between providers, risk assessment, and early intervention.
- Initially, measures will be captured via claims data, including International Classification of Diseases (ICD)-9 and Current Procedural Terminology (CPT) Category II codes.
- The existing categories of CPT Category II codes (history, physical examination, and intervention) may be challenging to adapt for pathology services.

E-mail address: delston@geisinger.edu
In the absence of appropriate CPT codes, the Centers for Medicare and Medicaid Services (CMS) may establish G codes to facilitate data capture.

Structural measures are of great value in laboratory medicine but are currently difficult to capture through coding and documentation in the medical record.

The Tax Relief and Health Care Act of 2006, P.L. 109-432, established the Physician Quality Reporting Initiative (PQRI) Program for 2007 and 2008 [1]. The 6-month reporting period for 2007 will run from July 1 through the end of December. There are current proposals to delay implementation of a permanent Medicare quality-reporting system to allow evaluation of the 6-month transitional reporting period of the program. There also are proposals to repeal or alter the program. The current program is voluntary, but the regulation allows it to be made mandatory at the discretion of the Secretary of Health and Human Services.

**Evolution of the Physician Quality Reporting Initiative**

The Institute of Medicine report on pay-for-performance (“Rewarding Provider Performance: Aligning Incentives in Medicare”) [2] should be required reading for anyone involved in laboratory quality improvement. This report is one of three that build on the original *Quality Chasm* series, which created the national goal of safe, effective, patient-centered, timely, efficient, and equitable health care. Recognizing that existing payment policies reward physicians according to the number and complexity of services provided, the report recommended that payment incentives be created to encourage structural and process measures to promote higher value.

Key recommendations of the report are

1. The Secretary of Health and Human Services should implement a Medicare pay-for-performance program using a phased approach.
2. Initial incentive should come largely from existing funds.
3. Congress should give the Secretary authority to consolidate financial reward pools when measures allow for shared accountability and coordinated care.
4. The program should reward health care that is of high clinical quality, patient-centered, and efficient.
5. The program initially should reward providers who improve performance as well as those who achieve high performance.
6. There should be incentives for reporting, and data should be made public in ways that are meaningful and understandable to consumers.
7. The Secretary should develop a strategy to ensure that virtually all Medicare providers participate. Reporting should be voluntary initially and should evolve to be broad in scope.
8. The program should reward improved coordination of care.
9. The Secretary should explore options to assist providers in the implementation of electronic data collection and reporting systems.

10. The Secretary should implement a system to evaluate the program.

The report also reaffirmed several principles. Health care should be based on continuous healing relationships and customized according to patient needs and values. The patient should be involved in evidence-based decision making. Patients should have access to their own medical information. Patient safety should be a priority, and the health system should not waste financial resources or patient time.

The push for quality measures comes from many directions. Congress and the public are demanding greater transparency and accessibility of quality data. Employers are pushing for greater accountability for their health care spending. Lee Iacocca noted that “It is a well known fact that U.S. automobile industry spends more per car on health care than on steel” [3]. The list of large corporations pushing for quality measures now includes General Electric, BellSouth, Boeing, Bristol-Myers Squibb, General Motors, Honeywell, Johnson & Johnson, Marriott, Motorola, the Pacific Business Group on Health, Proctor and Gamble, UPS, and Xerox.

The AQA Alliance, formerly the Ambulatory Care Quality Alliance, was founded as an alliance between the American Academy of Family Physicians, the American College of Physicians, America’s Health Insurance Plans, and the Agency for Healthcare Research and Quality [4]. Its membership has expanded, so it now represents a broad multi-stakeholder group. Similarly, the National Quality Forum (NQF), a private, not-for-profit membership organization, includes broad representation and was created to develop and implement a national strategy for health care quality measurement and reporting. These organizations can endorse quality measures to be used by the CMS, but the CMS also may use measures in the absence of AQA Alliance and NQF endorsement.

The challenge of establishing measures for laboratory medicine

Several factors have complicated the establishment of measures for laboratory medicine. The initial phase of the plan did not include structural measures. CPT Category II code categories of history, physical examination, and intervention are designed largely for clinical services. They are difficult to adapt to laboratory services. Also, Part B services that are not paid under the Physician Fee Schedule, such as clinical laboratory services, were specifically excluded from bonus payment [5].

Measure sets initially will be narrow and will exclude many of those who work in laboratory medicine. Standardized reporting requirements for breast and colon cancers are of value but fall short of reflecting the breadth and scope of laboratory medicine. The challenge is to create evidence-based measures that reflect the full scope of laboratory practice, are based
on current vetted guidelines of care, and have published evidence of an existing gap in care.

Attribution is another problematic issue for laboratory measures. Normally, each episode of care and measure is attributable to only one physician. For process and outcome measures, attribution to the clinician excludes the pathologist.

The process of attribution must be transparent. Methods of identifying comparator groups should be transparent also. The challenges are not small, but the goal of improved patient care is important.

**Guidelines for measure development**

Measures are brought forward by specialty societies working in cooperation with the American Medical Association (AMA) Physician Consortium for Quality Improvement. Principles for measure development and areas of focus have been enumerated by the Institute of Medicine, the AMA, the AQA Alliance and the NQF [6–8]. Measures must be evidence-based and supported by current vetted guidelines of care. There should be a comprehensive review of relevant literature as well as involvement of a broad range of interested stakeholders. Measures are intended to improve patient care but should not be construed as establishing a legal standard of care.

An inventory of available guidelines serves as a useful starting point for measure development. A useful guideline inventory can be found at [http://www.guideline.gov/](http://www.guideline.gov/).

Measures should address processes over which physicians have control. There should be published evidence of an existing gap in care, variation in practice, and potential savings in health care costs. Savings can be direct or can result from early disease management reducing later morbidity. Quality measures should be evaluated in relation to cost, and cost-of-care measures should be evaluated for their overall impact on quality.

Measures may address appropriate use of services, continuity of care, coordination of care between providers, risk assessment, and early intervention. Validation of measures is critical to ensure that they accomplish what was intended and to address issues of perverse incentive. Outcome measures are particularly prone to perverse incentive, because they may reward physicians for abandoning their sickest and neediest patients. To avoid perverse incentive, measures must take case-mix into account and allow for risk-adjusted measurement.

Process measures often are bundled as a series of steps to be accomplished. Outcome measures can be clinical, but surrogate laboratory markers of disease control often are used. Thus, the laboratory commonly serves as a means of capturing quality data for clinical specialties. Capturing quality data for the laboratory itself remains a challenge. Structural measures are well suited to laboratory medicine and can be integrated into existing performance improvement plans. Structural measures include
electronic order entry, interoperability and seamless interface between electronic systems, barcoding, and electronic patient identification devices. Capture and attribution to a particular patient encounter can be challenging for structural measures. Whenever possible, the burden of complying with the measure should be minimized by the use of technology. The incentive should be sufficient to offset the expense of implementation.

At least in the initial phases, measures are to be captured entirely through claims data, such as ICD-9 and CPT codes. Category II codes are designed specifically to capture quality data. Drug use data will become available through Medicare Part D, but the CMS is not yet ready to capture Part D data for the purposes of PQRI. CMS may establish G codes to facilitate data capture when other means of capture are not available. G codes often serve as stand-in codes until Category II codes are released.

Ultimately, quality measures should address all populations, but the PQRI program is directed at the Medicare population. Initial measure sets will be narrow in scope, but the ultimate goal is to have measures that reflect the scope and breadth of laboratory practice. The conditions should reflect the patient populations served by laboratory. Measures also should reflect tests or procedures that are high cost, high risk, or high volume. They should account for significant Medicare expenditure. Identified disease states appropriate to laboratory medicine include diabetes and coronary artery disease. These diseases represent a significant risk to public health and account for a large portion of yearly Medicare expenditure. Measures that address these conditions would be welcomed.

**Defining a measure**

Each measure needs a defined denominator and numerator to facilitate data capture. Examples of appropriate denominators and numerators for different types of measures are given in Box 1.

Denominator exclusions must be specified so the physician is not held accountable for patients who should not be counted in the denominator. For process measures, valid reasons for excluding a patient from the denominator include medical reasons, patient reasons, and system reasons. Medical reasons are situations in which the measured action is not indicated or is contraindicated. For example, a patient who has undergone a bilateral mastectomy would not be expected to have a screening mammogram. Patient reasons for denominator exclusion include the patient’s right to decline, economic reasons not to perform the action, and religious reasons not to perform the action. There also may be system reasons to exclude a patient from the denominator; for example, another physician may already have taken the action or be providing the service. Duplication would not be appropriate and would serve only to increase the cost of care. A CPT modifier exists for each class of exclusion: modifier 1P is used to
Box 1. Appropriate denominators and numerators for process and outcome measures.

Process measure:
Percentage of patients with coronary artery disease who were prescribed a lipid-lowering therapy (based on current American College of Cardiology/American Heart Association guidelines)

Outcome measure:
Percentage of patients with diabetes whose most recent low-density lipoprotein (LDL)-C less than 100 mg/dL or less than 130 mg/dL

exclude patients from the denominator for medical reasons, 2P for patient reasons, and 3P for system reasons.

Finally, the intended users of the measure and the patient population must be specified. For example, an LDL target measure for hyperlipidemia is intended for a primary care physician. A breast cancer report standardization measure is designed for a pathologist who examines breast specimens. The laboratory may report the LDL level but should not be held accountable for the control target. A surgeon may receive and read the breast cancer pathology report but should not be held accountable for the form and content of the report. Validation of measures must address the success of the measure in light of the intended users, patient population, and attribution to the correct patient encounter and physician.

Current status of the Physician Quality Reporting Initiative

A list final list of the 74 PQRI quality measures for 2007 is posted at www.cms.hhs.gov/PQRI. Reporting and performance are tracked by National Provider Identification Number, but payments are made by Tax ID number. This system creates issues of accountability and compensation for large organizations with a single tax ID number. Organizations will receive data on performance of individual providers within the group. The CMS also has announced that large multispecialty groups participating in the Physician Group Practice Demonstration Project qualify for incentive payment without reporting 2007 PQRI measures.

For 2008, measures should be adopted or endorsed by a consensus organization, such as the AQA Alliance or the NQF. The measures should have been submitted by a physician specialty, using a consensus-based process for development. There is support for including structural measures, such as
the use of electronic health records, electronic patient identification, or electronic prescribing technology. This inclusion is critical, because structural measures generally are the ones most appropriate for laboratory medicine. Readers are advised to review current information posted on the CMS Web site, https://www.cms.hhs.gov/PQRI.

References


