Chapter 2
The Landscape of Error in Surgical Pathology

Frederick A. Meier

The landscape of error is a stretch of country that hides dangers. A traveler does well to search out geographic features of risks that lie in wait, study reports of accidents encountered along the way, and seek out the impressions of other travelers.

Definitions: Geographic Features of Error

Error: In everyday language, error is getting things wrong, usually in relation to aims and purposes [67]. There is a different technical use of the term error in statistics [27]. For statisticians, error means differences in repeated measurements. These measurement differences arise from either random variation or bias. Random variation causes inconsistent differences between measurements; bias produces systematic differences between measuring methods or devices.

Ordinary language and statistical error: Ordinary language and statistical uses of error have this in common: we *make* both errors and measurements. Study of the two kinds of error connects in this way: observers detect differences between random variation and events gone wrong systematically by measuring characteristics of the events that fail to achieve their purposes. Observers may then act rationally from their understanding of nonrandom variation to reduce and sometimes prevent practical errors. This way of connecting systematic event measurement with process improvement follows from the insight into production processes first articulated by the statistician Walter Shewhart, then extended and made famous by, Shewhart's student W.E. Deming [13, 79]. The Shewhart–Deming approach investigates *practical errors*, failures of steps in a process to achieve their objectives, and

F. A. Meier (✉)
Department of Pathology and Laboratory Medicine, Henry Ford Hospital,
2799 West Grand Blvd, Detroit, MI 48202, USA
Email: Frederickmeier4@gmail.com

Department of Pathology, School of Medicine, Wayne State University, Detroit, MI 48202, USA

R. E. Nakhleh (ed.), *Error Reduction and Prevention in Surgical Pathology,*
DOI 10.1007/978-1-4939-2339-7_2

attacks the variations in events that go wrong because of identifiable *root causes*, influences on processes that are motors of nonrandom variation.

Interpretative error and observer variation: Interpretative errors are impressions of how things are that turn out to be wrong. Investigations of statistical error make another distinction that carries over into the study of surgical pathology error: this concept is *intermethod* or *interobserver* variability. Observer variability is important if one is to understand the strengths and limitations of *review*. Review, looking again at diagnoses that have already been made, is the most frequent way to study everyday *interpretative error* in surgical pathology. The important statistical distinction for interpretative errors is between variability that occurs when the same method or observer makes repeated measurements (*intramethod* or *intraobserver* variability) and variability that occurs when two or more methods or observers measure the same phenomenon (*intermethod* or *interobserver* variability). Most of the time, interpretative error in surgical pathology, comes wrapped in *interobserver* variability, while *intraobserver* variability lingers in the background.

Practical errors in the surgical pathology production process: A production process is a series of steps. In the case of surgical pathology, the process turns patient samples into diagnostically, prognostically, and therapeutically relevant information. At each step in the process, marks can be missed. As outlined in Fig. 2.1, the

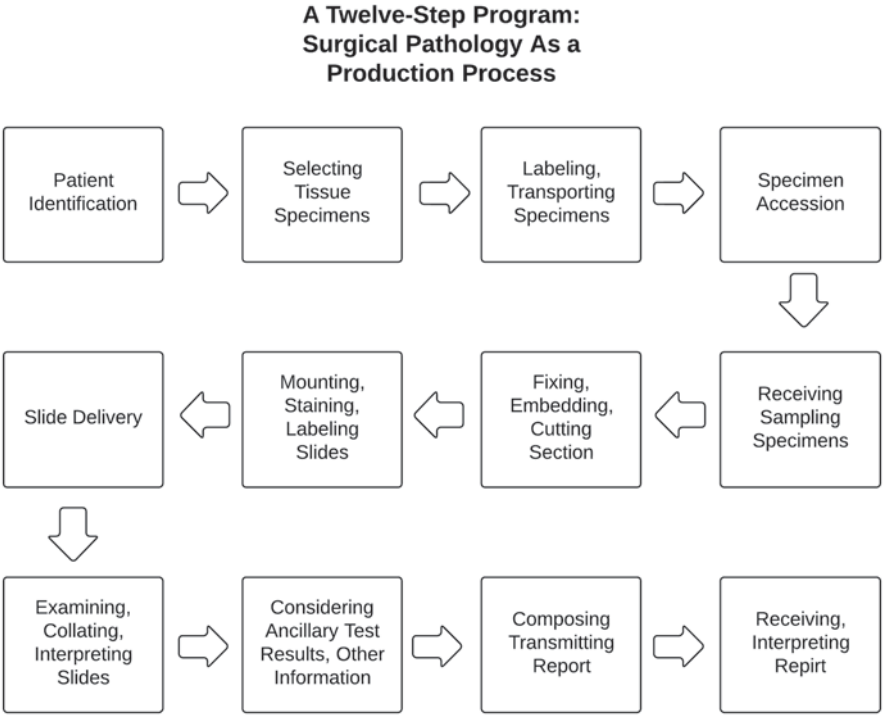

Fig. 2.1 A twelve-step program: surgical pathology as a production process

production process begins with identifying patients, goes on to select specimens, then proceeds to label, transport, and accession them. The process continues with steps of describing received specimens, sampling them, fixing, embedding and cutting them, mounting processed sections of samples on slides, then staining the slides, labeling them, and delivering them to surgical pathologists. These interpreters of slides, in the central step in the process, examine the sections on the slides. At this point, surgical pathologists also obtain information from other sources—especially ancillary test results and reports of clinical circumstances—and may request further in these reports they transmit, ultimately, to readers, who may act or not on the information. Made up of all these steps, the surgical pathology process is a complex task open to practical defects beyond errors in the central step of interpretation.

Amended Reports and practical errors: Amended reports in surgical pathology are like accident reports. As sources for a taxonomy of defects, amended reports particularly help practitioners study practical errors in surgical pathology. They highlight the sorts of defects that lean production policies and procedures help decrease or eliminate in the surgical pathology production processes.

Information theory and error: In terms of Claude Shannon's *mathematical theory of communication* [13], observer variability is variation in signal reception. Shannon's theory, on which computer programming is based, predicts that getting from antecedent potential message to subsequent actual message always entails making errors [19, 77]. Information theory, as worked out by Shannon and his colleagues, provides a framework within which to think about the making of diagnostic message, the central task of surgical pathology.

Interpretative error in the surgical pathology information flow: Error arises in the information flow (Fig. 2.2) either by commission, not getting the information

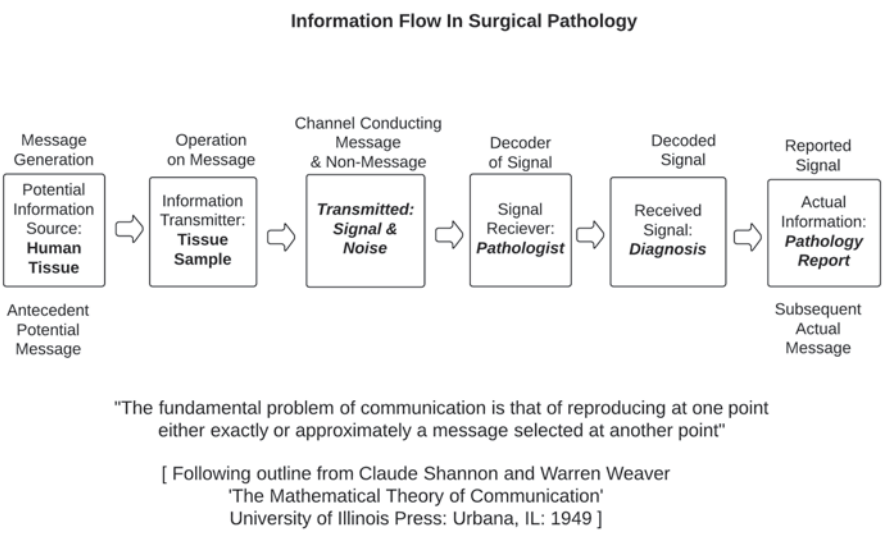

Fig. 2.2 Information flow in surgical pathology

that is signaled from slides right, or by omission, missing the potential information that the slides have to offer. Practical and interpretative errors are distinct sorts of defects. They are studied differently [39]. In this chapter, we focus on amended surgical pathology reports as the most convenient source of information about practical errors and reviews as the most available source for rates of interpretive errors.

Root causes: Root causes are primary defects that occur earliest, farthest upstream, in the practical production processes. There are more steps in the production process (Fig. 2.1) than there are in cognate information flow (Fig. 2.2). Practical errors are, it follows, most often the root causes of errors in surgical pathology; this is particularly true of errors that can be prevented. For this reason, root cause analysis of errors in the surgical production process is the key to developing practical counter measures to improve the process's performance [66].

Cognitive errors: Information theory gives the best account of how errors about facts arise in the interrogation of tissues. As presented in reports, surgical pathologists' mistaken beliefs about matters of fact and classified states are *cognitive errors*. Nicholas Rescher observes: "specifically cognitive error roots in our human need to resolve issues of thought and action in conditions of imperfect information" [65] or, in the foundational insight of the information age, articulated by Claude Shannon, any sort of information is always imperfect [19, 77].

Information theory maps surgical pathology error: The characteristic features of interpretative error in surgical pathology unfold in the terms of information theory. As outlined in Fig. 2.2, surgical pathologists search tissue samples for answers to questions: in the most frequently considered example they question whether or not a malignancy is present, what sort of neoplasm it may be, which features predict its behavior, and whether characteristics are present that indicate a particular therapy. Pathologists' reports convey information about primary matters of fact: a tissue sample does or does not contain lung cancer; primary matters of classification: a lung cancer is or is not adenocarcinoma; they also inform about secondary matters of fact: an adenocarcinoma does or does not appear within vessels or lymph nodes; and secondary matters of classification: a particular sample of adenocarcinoma of the lung has or lacks specific molecular signatures that indicate susceptibility or resistance to specific chemotherapeutic agents.

The information stream: Shannon discovered that, in the flow of information, a message is selected at an anterior (upstream) point then reproduced at a posterior (downstream) point. This sequence always runs from information sources to messages. In Fig. 2.2, we match the Shannon sequence to surgical pathology terms. From an information source (human tissue) of antecedent, potential information, a transmitter (the tissue sample) selects antecedent message, but the transmitter emits both a signal (anterior, potential information) and noise (mixed-in nonsignal that yields nonmessage). From this mix of signal and noise, receivers (surgical pathologists) select received signals (diagnoses, in Shannon's terms, subsequent message), which they then pass on as posterior, actual messages (reports).

Interpretive errors and uncertainty: This is a reality beneath interpretative error: any communication system that fits Shannon's pattern entails uncertainty. Every second, posterior, actual message (every reported diagnostic claim) has some

chance of being wrong (for pathologists, either missed diagnoses, wrong diagnoses, or misclassified diagnosis). Quantification of this chance of being wrong calculates greater or lesser likelihood of interpretive error. This is the underlying variation that review of diagnoses aims to define.

Surgical pathology is also an interpretative framework: At this point, it is worth observing that, besides being a production process, and a pattern of information flow, surgical pathology is also a conceptual structure. This framework is a group of classifications or taxonomies. The taxonomies aim to transmit the diagnostic, prognostic, and therapeutically relevant information that the production process creates. An act of interpretation places a received signal in a category within a classification. The characteristics of various classifications set limits to the reproducibility of the information. The variable applications of taxonomies also limit the validity, reproducibility, and detail of surgical pathology reports [39]. Taxonomic variability, like intraobserver variability, always lurks in the background, when we think about surgical pathology error.

Validity, reproducibility, and detail: In studies of interpretative diagnostic variability, three properties of measurement—validity, reproducibility, and detail—come into play again and again. *Validity* is the extent to which measurements correspond to real states of how things are. Increasing validity depends on decreasing systematic differences between observed appearances and real states of being. *Reproducibility* depends on how often repeated measurements return the same result, whether or not that result reflects the real state of things. Random variation sets limits to reproducibility. *Detail* depends on the amount of information that measurements provide. The degree of detail determines how much an observer knows about what he has measured after he has measured it. Keeping these three attributes in mind aids orderly study of error in surgical pathology. Importantly, interpretative discrepancies produced by review of surgical pathology diagnoses combine differences in validity, with variability introduced by differences in reproducibility, and variation in matters of detail. In review discrepancies, these three contributing features are usually inseparable.

Surgical pathology is, in addition, a dynamic scientific discipline: The developing scientific discipline is the larger context that surrounds study of both process and interpretative, including classification error. As a discipline, surgical pathology has assimilated increasingly elaborate techniques that assist in acquiring and processing information. These ancillary techniques find information both on the slide (as most prominently from immunoperoxidase stains) and from handling the sample in different milieux (as most prominently in molecular tests). The information gleaned from samples by converging morphological, quasi-morphological, and molecular techniques yields the explanatory criteria on which the informative classifications base themselves. The changing state of the discipline limits validity and detail in the information generated by the technical process. In particular, sources of information besides histopathological morphology, especially immunohistochemical profiles and molecular motifs, increasingly influence classification. In this wider context, complexity leads to error. As we will emphasize below, incorporating the

insights of the sociologist Charles Perrow, increasing practical complexity of process compounds increased complexity of interpretation [48, 50].

Oversimplification: Surgical pathologists always generalize from particular findings on slides to general diagnoses of disease states. As actual message, emerging from the information stream, pathology reports inevitably oversimplify. Another of Claude Shannon's seminal insights is that informativeness of a message increases in proportion to its vulnerability to disproof. This is the juncture where detail joins validity and reproducibility in the trio of important attributes of surgical pathology information. As they compose reports, pathologists arrange information content. They may reduce complex data presentations to simple ones; they may proliferate qualifications; or they may take away informative detail. In these three ways, they limit, obscure, or decrease the amount of information transferred to clinicians. With these strategies, pathologists try to prevent error by hedging; they trade off informative message for evidential security. This tactic fails when it drains reports of detail, exactness, and precision [68].

Errors of commission and omission: Shannon's communication theory makes sense of an ancient distinction between errors of commission (getting diagnoses wrong) and errors of omission (missing diagnoses). Errors of commission are misleading messages; these diagnostic failures (wrong diagnoses) appear among positive reports. Errors of omission fail to receive anterior diagnostic message. Errors of omission hide among negative reports. To recognize the commission:omission dichotomy, interpretative error detection must combine two different review approaches: (i) review (often redundantly called double review) of positive reports at risk and (ii) review of negative reports in high-risk categories of specimens [69].

Review in search of error and hindsight bias: Review of diagnoses is to interpretative error, what searching for root causes of defects is to practical error. Review checks the information transfer step in which the pathologist moves from receiving the signal or the slide to composing a report. Important conditions of review are when, where, how, and by whom review is done. *Hindsight bias* is made up of the systematic differences between looking forward at a new set of facts and looking back at an old set. Six systematical differences between the initial diagnostic event and the review event define various mixes of hindsight bias. The first of these distinctions is between *internal* and *external review*. Internal review is carried out within the practice in which the diagnoses under scrutiny were originally rendered. Pathologists in other practices perform external review. The second distinction is between *pre-sign-out review* and *post-sign-out review*. Pre-sign-out review takes place before a report is issued. Post-sign-out review happens after reports are released. A third difference is between *conference review* and *non-conference review*. Conference reviews are those that surround multispecialty gatherings at which cross-specialty agreement on diagnosis, prognosis, and therapy are sought. A fourth distinction appears between *expert* and *non-expert review*. Expert review is by a pathologist with increased knowledge and experience with the sort of diagnoses under review. A fifth pertinent difference is between *blinded* and *non-blinded reviews*. Blinded reviews are those reviews by pathologists with no more information than the primary pathologist possessed about a case; indeed a blinded reviewer

sometimes is given less case-specific information. The last of these variations in review schemes, but probably not the least important, is that between *focused* and *unfocused reviews*. Focused review trains the reviewer's gaze on specific sorts of specimens or diagnoses. Unfocused reviews either take all comers or check a defined fraction of cases without requiring that they be of specific specimens or types of diagnoses. The variable influences of these half dozen factors together make comparison of review discrepancy rates difficult.

Information sources about surgical pathology error: At least two kinds of studies yield useful information about error in surgical pathology: classification of errors turned up by amended reports and sorting of discrepancy rates by review of surgical pathology diagnoses. The effort to understand how the two sources provide information plunges us into occasionally detailed discussions of each of them. The detail is meant to illustrate the two approaches, not to provide an exhaustive evaluation of each approach.

Amended Reports as a Source for a Taxonomy of Surgical Pathology Defects

Amendments: Report amendments record both practical defects and interpretative errors. Because practical errors are more frequent than interpretive errors, root causes of amended reports map more often to the twelve-step production process (Fig. 2.1) than to the six-step information flow (Fig. 2.2). Mapped to either sequence, amended reports offer opportunities to study systematically both surgical pathology errors and the counter measures aimed to decrease them [1, 35, 36, 86].

Amendments vs. addenda: To achieve semantic consistency, the alterations of surgical pathology reports after they have been issued must be separated into dichotomous groups. One group is composed of *amendments*: all changes that were not purely additions of information. The other group is made up of *addenda*: altered reports that include only alterations that purely add information. Adherence to this dichotomy has proven necessary both to detect reports with errors in them and to separate error from other sorts of report variation [34–36].

Taxonomic consistency: Across many institutions, classifiers of altered reports have been able to agree on four defect categories and to sort consistently into these categories [34, 86]. The categories are: *misidentifications, specimen defects, misinterpretations,* and residual *report defects*. Report defects are residual because they classify the amendments that are left over after misidentifications, specimen defects, and misinterpretations have been classified.

Misidentifications fail to designate accurately *patients, tissues, laterality*, or other *anatomic localization. Specimen defects* include submitted specimens that are *lost*, those of *inadequate sample volume* or *size*, those with *absent* or *discrepant measurements*, and those with *inadequately representative sampling*, as well as, importantly, and less intuitively, those with *absent* or *inappropriate ancillary studies.*

Misinterpretations fail to state diagnostic information accurately. They have an internal structure more complex than misidentifications and specimen defects. This complexity has led to misinterpretations being divided into three *subtypes*. The first subtype includes errors of commission; these are false-positive diagnoses, or *overcalls*. This sort of amendment registers the retraction of wrong information. The second subtype is made up of errors of omission; these are false negatives or *undercalls*. This second sort of amendment registers either failures to recognize accurate information or initial loss of information that later was found to reside in the sampled tissues. The third subtype is *confusion* or *conflation* of relevant, similar, but distinct diagnostic categories. The findings in the third subtype are not over- or underdetermined, rather, they are misnamed diagnostic designations. Such a finding in question was initially registered correctly as positive; it is not overcalled nor is it undercalled, but the initial designation that it received is later realized to be not as accurate as a revised formulation. The amendment registers this revision. The three misinterpretation subtypes, in turn, relate to two *levels* of diagnostic message: *primary level* amendments register failures to distinguish positive from negative, malignant from benign; and *secondary level* amendments mark failures to characterize subordinate diagnostic features appropriately. The subordinate secondary diagnostic features affect clinical context, prognosis, or susceptibility to specific therapies. Most often these secondary characteristics are grade, stage, state of surgical margins, or lymph node status in specimens resected for malignancy.

Report defects: After misidentifications, sample defects, and misinterpretations have been excluded, the residual category in the taxonomy is *report defects*. Report defects also present themselves in three subtypes: (i) missing or erroneous *non-diagnostic information*—absent or wrong information about practitioners, procedures, billing codes, etc., (ii) *dictation or transcription errors*—typographical errors in the strict, proof-reader's sense, and (iii) failures or *aberrations in electronic report formats* or transmissions—the miscues colloquially called computer glitches. These report errors are all defects in product, but they have in common that they do not directly affect diagnostic information. Misidentifications, misinterpretations, and specimen defects, in contrast, all directly interfere with the diagnostic message itself. Report defects, however, are not unimportant. Although they fail to muddle message directly, as Shannon realized, they harm the information flow by reducing information redundancy [19]. Redundancy is the informative context in which the text of any message always arrives.

Root causes of amendment types: In the twelve-step production process (Fig. 2.1), the root causes of misidentifications and sample defects appear mostly in the early steps of the surgical pathology process, during specimen collection and sample processing, but, in a minority of instances, they pop up later. The root causes of misinterpretation focus in the middle of the process, when the case is on the pathologist's desk. Root causes of residual report defects inject themselves into the process at multiple points, but they also tend to cluster at its beginning, before the case reaches the pathologist, and at its end, after the pathologist has settled on diagnostic interpretations.

Application of the Amended Reports Taxonomy: Uniform application of this taxonomy allows consistent monitoring of amended reports among institutions and also within an institution over time. Important to process improvement, when amended rates are followed longitudinally over time, they also evaluate the success or failure of interventions aimed to reduce errors that amendments identify [1, 34–36].

Three characteristics of defect discovery: The amendment taxonomy revealed a trio of characteristics surrounding the discovery of defects. First, the more observers monitoring amendments, using the dichotomous definition, the more amendments are identified, usually at the expense of addenda. Second, clinicians discovered most misidentifications; pathologists found most misinterpretations; but discovery of specimen defects were scattered among different observers and discoverers of report defects usually remained anonymous. Third, clinician calls were the most frequent mechanism for detecting misidentifications, and, initially, conference review was the most fruitful mechanism for detecting misinterpretations. Conference review discovered, in various settings, between a little more than 40% to a little more than 80% of all misinterpretations that produced amendments [34]. The many fewer sample defects were found by multiple detection mechanisms. The residual report defects were found about as often by pathologist review, from clinician calls, at conference review, and surfacing in unknown ways.

Effects of lean interventions: In a large surgical practice that accessioned 45–50,000 specimens each year, real time editing of altered reports, undertaken together with changes in process aimed at reducing and preventing the underlying defects, had the following consequences over a 5-year period. Initially, active monitoring caused amendment rates to rise, from approximately 5-amendments/1000 reports to 10/1000 as altered reports were consistently defined as amendments or addenda. Next, as monitoring continued and counter measures were applied, amendment rates fell back to the 5-amendments/1000 reports level. Lean interventions in surgical pathology report production then caused misidentifications to fall from 16 to 9% of all amended reports. Despite similar interventions, however, the fraction of amendments caused by specimen defects remained at about the same low magnitude (<11%) and continued to be highly variable from year to year. In contrast, the fraction of misinterpretations fell dramatically, from 18 to 3% of all amendments. This fall was associated with introduction of pre-sign-out review of all breast and prostate cases, then, in addition, cases of some gastrointestinal tract lesions. Here, we see a synergistic connection between approaches to surgical pathology error monitoring: a decline in amendments documented the beneficial effect of a particular pattern of diagnostic review. Finally, and reciprocally, as misidentifications and misinterpretations fell, the residual category's report defects increased its fractional contribution, from 64 to 83% of all amendments.

Lessons from root cause analysis: When case-by-case root cause analysis of amendments assessed success or failure of interventions, three findings emerged: (i) efforts to reduce misidentifications at the specimen collection level (where most of these errors occurred) had a measurable, but modest beneficial effect, (ii) extensive standardization of specimen accession and gross examination reduced specimen defects surrounding ancillary testing, but not specimen defects overall, and (iii) most

impressively, introduction of internal pre-sign-out review of all breast and prostate and some gastrointestinal cases was specifically associated with a drastic reduction in misinterpretations [36, 34].

Amendments vs. addenda: The problem with amendment monitoring caused by misclassification of amendments as addenda continued over time. During active monitoring, 10% of so-called addenda have consistently turned out to be amendments. The adoption of misclassification amendments as an index of ongoing professional performance evaluation (OPPE) has now worsened this tendency to misclassify amendments as addenda [33, 35, 36].

Q-PROBES study of amendments using validated taxonomy: In 2011, as part of a College of American Pathologists Q-PROBES study, 73 participating institutions analyzed almost 1700 amendments over a 12-week period [19]. The Q-PROBE study's salient results are presented here to complete our account of how amendments characterize errors.

The taxonomy-classified amendments effectively across 73 institutions: Using the taxonomy, Q-PROBES subscribers classified 1665 of 1688 amendments (98.6%). In contrast to our large institutional experience, however, the fractions of misidentifications (13.3%), specimen defects (13.7%), and misinterpretations (14.6%) were about equal [1].

Amendment rates: Median defect rates among Q-PROBES participants hovered around 5-amendments/1000 published reports: the aggregate defect rate was 4.7-amendments/1000 cases and a median participating institution's defects rate was 5.7/1000. This median amendment rate is similar to the 5/1000 experienced in our single institution monitoring. However, among the 73 Q-PROBES study participants, the range around this median was wide; it extended from 0.9/1000 to 13.5/1000 amendments/reports issued [1].

Misidentifications and sample defects: In the Q-PROBES study, among 225 misidentifications, 31.5% were of patients, 20.0% of tissue type, 23.0% of laterality, and 25.5% of anatomic localization [1]. Among 231 sample defects, more than three-quarters (77.4%) involved ancillary testing and the rest mostly involved gross and microscopic sampling [1]. The growing association of sample-related defects with misdirected or failed ancillary testing is a phenomenon also observed in our single institution's longitudinal monitoring.

Misinterpretations: Analysis of 247 primary and secondary misinterpretation amendments found only 5.7% false positives and only 11.8% false negatives. These fractions are dramatically different from our single institution longitudinal experiences. The difference stemmed from very different rates of diagnostic relabeling. In the Q-PROBES cohort, 44.1% of misinterpretation amendments were attributed to confusion or conflation of similar but distinct diagnoses (misnaming). The Q-PROBES subscribers also produced a different pattern of interpretative errors from that found in the single institution experience. Misinterpretation amendments among the Q-Probes study participants were revised mainly for secondary features in amended reports of malignancy. These amendments usually changed grade or margin status [1].

Residual report defects: Among the Q-PROBES study participants, as in our long-term experience at one institution, the most common causes for amended reports were residual report defects: typographical errors, missing nonidentifying, noninterpretative report attributes, or wrong nondiagnostic report information [1].

Anatomic sites of origin of specimens that produce amended reports: In the Q-PROBES study, the most common tissues of origin for defective reports were the most common sites sampled: the skin, breast, and gastrointestinal tract. Submissions from these sites were about equal defect contributors (18.2, 17.7, and 18.1 %) [1].

Benchmark amendment rates from Q-PROBES study of amendments: The Q-PROBES study of amended reports yielded two benchmarks: First, with a 5/1000-defect rate, the current surgical pathology production process is a 'three sigma' production system for surgical pathology reports. Second, median rates of misidentifications and misinterpretations are fairly consistent. These two rates both run below 1/1000 and are about equal: 0.6 amendments for misidentifications/1000 reports and 0.8 amendments for misinterpretations/1000 reports [1].

Defects in the surgical pathology production process as normal accidents [48]: Findings about surgical pathology errors uncovered by root cause analysis of amendments agree with studies of other production processes in different settings [48, 50]. From studies in a variety of complex production processes, Charles Perrow defined untoward events, like those which amendments document as *normal accidents*. He argued that these events occur in conditions of *complexity* created by interconnecting subsystems. In surgical pathology, the interconnecting subsystems are the preanalytic, analytic, and postanalytic phases of the report production process. A second error-inducing characteristic, *tight coupling*, then mediates the connection between subsystem derangement and damage to the final product. A third characteristic is *concentration*. In surgical pathology laboratories, high volumes of specimens are concentrated by converging from multiple collection sites to enter the production process. Once concentrated in the process, these specimens are also subjected to complex ancillary tests that detect specific antigens by immunohistochemistry panels and detect molecular motifs using nucleic acid amplification. Computer-enabled communication tightly couples pathologists with pathologist assistants, histologists, and clinicians. Perrow argues persuasively that such concentration, complexity, and tight coupling together inevitably amplify practical error [47, 49].

Eight contributors to normal accidents: All eight features that make systems prone to normal accidents are present in the surgical pathology production process [47]. In the following list we cite, next to each error-promoting feature, examples of its appearance in the surgical pathology setting:

1. *Proximity of components*: proximity appears among specimen jars awaiting samples in endoscopy suites and in shopping bags full of many different patients' skin biopsies arriving at accessioning stations
2. *Common-mode connections*: large specimen gross examination stations are common mode connections when pathologist's assistants examine in succession

multiple partial mastectomy and lymph node dissection specimens or multiple colon resections during the same accessioning shift
3. *Interconnected subsystems*: subsystems interconnect when prostate biopsies obtained in an ambulatory surgery setting arrive simultaneously at the same accession desk with the products of a radical neck dissection from a frozen section room
4. *Feed-back loops*: different feed-back loops cycle simultaneously as telephone calls go back and forth between pathologist's reviewing slides and pathologist's assistants returning to fixed specimens to harvest more tissue samples, while, at the same time, pathologists send computer messages to histologists to request additional levels and special stains
5. *Limited substitutions*: the constraints due to the different tissue processor cycles limit substitutions of cassette batches depending on run times
6. *Multiple interacting controls*: multiple interacting controls appear at accession in identification of specimens, in the histology laboratory, with the sorting of blocks, and on pathologist's desks at the arrival of slides
7. *Indirect information transfer*: indirect information transfer occurs when clinical features about cases are reported only in shouts over the shoulder of an operating room technician hurrying down a hallway, critical choices in specimen sampling are made only in whispers among residents at specimen processing stations, or vital new clinical information arrives only in muttered remarks from a clinical fellow who has come to look at slides
8. *Limited understanding of the requirements of the process*: clinical staff collecting specimens have limited understanding of what requirements for histologic diagnosis are; pathologists as they interpret slides have limited understanding of what information clinicians imagine reports will contain

Ambivalent effect of electronic information transfer in complex processes: Computerization brings both positive innovations and dangers to the complex process that fits Perrow's description. The positive changes have reduced unwanted variation, standardized data input, and reduced dependence on the variable information transfer media. Computerization has also helped by programming formats like synoptic report checklists and has facilitated automation of routine tasks, like bar-coded logging-in specimens, collated with bar-coded requisition documents. However, negative changes brought by programmed processes of electronic information transfer require invariant sequences, stipulate one way to perform a component task, allow only limited buffers, and force only designed substitutions [47]. As computer-facilitated standardization has been achieved, former safeguards, redundancies, buffers, and alarms in previous surgical pathology systems have been eliminated. With newer complex systems come tighter couplings. High volume, complex, tightly coupled systems open themselves to untoward events in which two or more, often individually small failures interact in error-causing ways that process designers and operators have not anticipated. Such event sequences, Perrow and another student of system error, James Reason, find, precipitate disproportionately bad outcomes that Perrow has designated *catastrophes* [50, 59].

Lessons of lean principles and practices: In these distracting circumstances, sustained practical error reduction, incorporating lean industrial engineering principles and practices, has become a valuable response [3, 7, 11, 82, 94–96]. The lean approach, (systematic practical error detection, then error reduction, prevention, and amelioration through countermeasures), addresses all four defect types recorded by amended reports. For the three practical sorts of defects, the analysis makes connections like those presented in the next three paragraphs.

1. *Misidentification* is the practical error with the most devastating potential [82, 96]. To attack it, colleagues who labor upstream in the process must accept forcing functions, labeling standards, and new labeling procedures; the beneficial effects of this apparently extra upstream effort often exert themselves only downstream where those making the changes cannot see their laudable effects. Nevertheless, a trio of worthwhile points has emerged from lean interventions that improve patient and specimen identification. Detecting and preventing misidentification entails: (i) training in labeling standards that extends outside surgical pathology premises, to dermatologists' offices, endoscopy suites, and operating rooms, (ii) recognizing that batched printing of labels is a recurrent misidentification threat; flow design must avert it as much as possible, (iii) designing identification checks into multiple steps in the process, especially at two important checkpoints—(a) arrival of requisitions and specimen containers at accession and (b) reconciliation of requisitions with reports just before reports are released [82, 96].
2. *Specimen defects*: Practices following lean principles have also attempted to reduce specimen defects. Root causes of specimen defects increasingly reveal that ambiguities and delays in potentially decisive ancillary test results, especially those from molecular tests, are a growing cause of specimen defects [82].
3. *Result reporting*: In result reporting, the increasing importance of ancillary testing in surgical pathology often now forces a Hobson's choice. The unattractive decision falls between either issuing an incomplete report liable to later amendment or delaying the report until potentially modifying ancillary information can be combined fully into an integrated report [16].

Report errors and the benefits of redundancy: Another lean lesson also involves errors documented in residual report defects. As Shannon deduced about communication in general, redundancy has more substantial benefits than may be intuitively obvious [19, 77]. In the surgical pathology report production process, completeness of report information, other than patient and specimen identifications and diagnoses themselves, turns out to be helpful in averting error. For example, the presence of inconsistent clinical information on a requisition may be the only sign that a specimen jar has been mislabeled. Electronic medical records (EMRs) also supply useful redundancy. They provide access to clinical information that can help root out misidentifications and call into question dubious diagnoses. As counter measures, structured searches of EMRs confirm or expand the clinical context in which a submitted specimen has arrived. These routine searches can be of great assistance, both

in reducing practice report defects and in leveraging redundant report information to decrease misinterpretations.

Case Review to Detect and Reduce Interpretative Error

Active vs. Passive Monitoring: Reports of reviews are the main source of studies about surgical pathology interpretative error [39, 73, 81]. Review looks again at cases that have already generated diagnostic message, so review entails *active monitoring*; it searches for discrepancies, where classifying amended reports and pursuing their root causes which we have just discussed, is, in contrast, *passive monitoring*.

Review and information flow: In relation to information flow (Fig. 2.2), review exposes the same signal to different receivers each of whom has his or her own noise thresholds and variable sensitivities to signal reception. These different receivers generate discrepancies, the products of review. Review, however, finds other sources of variation besides error in discrepancies.

Effect of interobserver variability on review: In active monitoring, interobserver variability always comes into play because implicit diagnostic thresholds and the application of explicit classification criteria are products of experience. Experience among pathologists inevitably differs. Importantly, primary diagnosticians and secondary reviewers also tend to function at different diagnostic thresholds.

Internal vs. external review: There is a relevant contrast, mentioned earlier in this chapter, *internal review* [41, 44, 46, 60, 63, 89, 92] among members of the same practice group and *external review* [30, 83] that involves members of different practices. In the first context, the internal reviewer is checking to see whether a local colleague is right. In the second context, the external reviewer is checking to see whether a distant noncolleague is wrong.

Expert vs. non-expert review: Another contrast appears when either internal or external reviewers are or are not subspecialist experts. In internal expert review, subject matter specialists within a department may set different diagnostic thresholds and use different explicit or implicit classification criteria than do general pathologists in the same practice. However, the internal expert's view of things usually affects his or her nonexpert colleagues' thresholds and criteria by feedback over time and through accumulation of shared cases (see calibration effects paragraph later in the text) [55]. Primary pathologists and external expert reviewers may diagnose against not only different horizons of experience but also different clinical objectives. In a common setting of external review, reviewers at an oncology hospital locate diagnoses on different horizons of experience than do less specialized referring pathologists. The oncology hospital pathologists also prepare their reports for specialist oncologists whose needs (and sometimes prejudices) are opaque to the primary diagnosticians [26].

Blind vs. informed review: A different sort of variable that affects the difference between thresholds is whether the secondary examination of the case is *blind*

review, whether the secondary case examiner does not or does know the primary pathologist's initial diagnosis, and whether the second examiner does or does not know more or different clinical information than did the primary pathologist [62, 64].

Effect of calibration: Active monitoring within a practice group may also produce *calibration* effects. Calibration appears when pathologists compare many cases over time and converge on similar thresholds and criteria. In practices with a dominant expert, calibration often converges on the dominant expert's thresholds and criteria in "the big dog effect" [55].

Interventions that lessen interobserver variation: Experience argues that calibration through consensus conferences and calibration slide sets are counter measures, which reduce the interobserver variation that difficult cases bring to internal review. These two strategies provide structured opportunities for practice colleagues to articulate agreement on diagnostic criteria and develop a shared vocabulary in which to discuss problematic cases. Consensus and calibration mechanisms also provide critical occasions for practice colleagues to address together the influence of modulating factors on diagnostic differences. In multispeciality conference settings, they can take into account the influence of clinical features and ancillary test results, and bring above the horizon variations in taxonomies.

Different signal vs. noise thresholds: Different thresholds produce obstacles to equating discrepancies with errors in primary diagnosis that is analogous to statisticians' type I error. When reviewers mistake noise for signal because of threshold differences, signal initially received appropriately is missed on review or miscategorized by the reviewer, due to variable reception of message in relation to noise.

Repeated failure to detect signal: Another difficulty in establishing 1-to-1 correspondences between review discrepancies and diagnostic errors occurs when both initial observations and reviews of the same signal miss diagnostic information that is really there. Reviewers' subsequent acceptance of initial missed diagnoses is analogous to statisticians' type II error. In this failure, review continues to leave initial false negatives, failures to register signal, in place.

Taxonomic variation: Differences in application of taxonomies present one more obstacle to the equation of review discrepancies with errors. Different diagnoses may or may not reflect the same constellation of signal findings. As observed earlier, diagnostic taxonomies have explicit and implicit features that nonexpert and expert users deploy differently. This is a specific instance of a general phenomenon. Each pathologist throws various taxonomic nets of diagnostic designations over histopathological realities. Different taxonomic nets may fit a reality better or worse, but different nets may also just fit the same reality differently, just as a triangular, coarse mesh may as accurately describe the same surface, as does a square, fine mesh. It is very hard to compare and contrast the relative fit to the reality of different taxonomic nets; however, these differences seem to lead to discrepancies in how different observers register the same realities [90].

Disparate information sources: A final barrier to equating discrepancies with errors arises when disparate pairs of diagnoses are reviewed. Examples of odd couples under review are frozen section:permanent section comparisons and

cytological:histological correlations. In both these pairs, the initial diagnosis in the dyad was from a different sample—or a differently processed sample—compared with a subsequent more information-rich specimen presentation. In these settings, method differences and observer differences get mulched together as diagnostic discrepancies [53, 54, 17].

Different Sorts of Review Compared and Contrasted

Internal vs. external reviews: Timeliness give pre-sign-out internal review major advantages: It often prompts prospective resolution of diagnostic discrepancies (for example between cytological and histological diagnoses from the same tissue source) that can trouble both pathologists and clinicians in retrospect. It gives internal experts opportunity to calibrate a practice's diagnostic thresholds and standardize application of taxonomies. Most importantly, it obviates need for report amendments when discrepancies are discovered. For large practices, internal review is usually faster and less expensive than external review, but the time commitment involved in internal review may be impractical for small practices. In small practice settings, the more probative weight of external expert review also carries added value with skeptical clinicians. For medium size and larger practices, external review is more expensive; it also may provide revision of diagnosis only after an embarrassingly long time. Clinical decision making is then either delayed or second-guessed. To avoid delays and potentially contradictory revisions, middle size practices tend to rely on conference-based review. This format combines opportunity for local expert review with additional clinical context.

Inevitably retrospective reviews: The benefits of internal, upstream, over external, downstream review, suggests that reviews should be carried out, in most settings, either before cases with identified risks are signed out or before their clinical implications can be acted on. Some correlations, however, remain inevitably retrospective. The correlation of uterine cervical cytology or cervical biopsy diagnoses with diagnoses from excision specimens is an example of this sort of unavoidable retrospection. Another, necessarily retrospective sort of review is the practice, already discussed earlier, of reviewing diagnoses of malignancy after patient referrals to centers for cancer treatment. Both of these review mechanisms remain ingrained in good practice [17, 20, 32, 91, 76].

Unfocused vs. focused reviews: As will be cited again, complete and set percentage reviews tend to produce lower frequencies of discrepancies than do focused reviews [57]. They do, however, remove selection bias. Reviews of cases focused on specific organs detect both false-positive and false-negative interpretations as well as misclassifications. In contrast, reviews focused on specific diagnoses catch only false positives and misclassifications. They provide, however, initial confirmation of the most significant diagnostic product of a pathology service: positive diagnoses are usually the most clinically relevant products of the surgical pathology produc-

tion process, so, if only one sort of case can be subjected to review, new positive diagnoses should be it [38, 42].

Subjects of focused reviews: Focused reviews most often train attention on specimens that are both relatively often submitted and relatively challenging to classify. Preneoplastic or borderline neoplastic breast [24, 45], melanocytic skin lesions [23, 80, 84, 88], and female genital tract lesions [32, 93] are frequent foci of review before sign-out. Another criterion for focused review is a high likelihood of interobserver variation. In these situations, the wisdom of sorting out local interobserver variation prospectively rather than retrospectively recommends internal, pre-sign-out review. Gleason grading of prostate adenocarcinoma [6, 15, 28, 51, 87], grading and staging of uterus and ovary malignancies [14, 93], and classification of thyroid lesions [4, 12, 25] commonly satisfy this criterion. More recently subclassification of adenocarcinoma of the lung has joined this group of classification challenges [21, 43, 58, 74].

Taxonomies that interact with ancillary studies, like those for adenocarinoma of the breast, lung, and kidney, are instances in which ancillary information's integration can be decisive. Classifications of leukemias and lymphomas and bone and soft tissue sarcomas are further instances in which complexity increases the degree of difficulty encountered on the way to review consensus [29].

Percentage vs. focused review: Stephen S. Raab led a study that compared two contrasting review approaches—percentage review and focused review [57]. Raab and his colleagues compared random review of 5% of specified sorts of cases with focused review of suspected troublesome specimen types of primary diagnoses. The study found a much higher discrepancy rate from focused review: 13.2% from focused review vs. 2.7% from percentage review. Raab and colleagues also looked at potential downstream implications of the uncovered diagnostic discrepancies. Instances they classified as major errors were found a power of ten more often in the focused review approach than they were in the random review scheme: 3.2 vs. 0.36% of cases. In Raab's study, higher yield makes focus review appear a wiser use of review time and expertise [52].

Focus of review and amendment rates: Andrew A. Renshaw and colleagues have used amendment rates to project the relative utility of different review strategies, by comparing the fractions of different case types with discrepancies and amendment rates in these case types [60–64]. In one study, they demonstrated that breast lesions, cytological:histological correlations of genital tract lesions, and thyroid diagnoses were particularly likely to lead to amendments. The relationship between the two fractions of discrepancies and amendments was: 27% of discrepancy-producing cases produced 88% of amended reports [61]. Renshaw also found a less dramatic but similar disproportion when he examined the case discrepancy: amendment relation for initially nondiagnostic or atypical/suspicious diagnoses. Cases with borderline diagnoses made up 4% of discrepancy-producing cases but 14% of amendments [61]. Two take-home lessons appear here: First, the mix of cases that a practice examines influences discrepancy frequency patterns. Second, borderline lesions (intraductal and lobular breast proliferations, intraglandular prostrate prolif-

erations, equivocal gynecological cytology classifications, and ambivalent thyroid cytology findings) increased the likelihood of discrepancies [61].

Burdens of review: Reviews cost time and effort. The essential burden of documenting individual reviews and collating the information aggregated from reviews is a major investment in data analysis. As a rough estimate of the number of cases that fall under the gaze of review, a recent survey by Nakhleh et al. found that, in typical settings, review protocols cover approximately 8% of a practice's cases [41]. Raab's seminal study (which should be duplicated in multiple, different settings) suggests, that focusing review on sorts of cases with known high rates of missed decisions or revised diagnoses is the preferred approach to case selection [57]. Another pivotal decision, which bears further investigation, is whether review should be blinded or not.

Review as a quality measure: We now reach the central paradox of review. Despite all the influences and interferences that make a one-to-one correspondence between review discrepancies and errors impossible, case review remains the main source of information about interpretative errors. Pathologists' knowledge and experience, their ability to correlate clinical failures with histopathological findings, their skill in combining morphological with nonmorphological (or quasi-morphological) ancillary findings, and their mastery of coherent taxonomy are four basic professional aptitudes into which review delves, however inadequately. The results of review offer both providers and users of surgical pathology reports imprecise but implication-rich indicators of diagnostic integrity. This indicator function is currently review's main contribution to the evaluation of quality in surgical pathology. Because so many variables affect review, comparisons among discrepancy rates, from one set of reviews to another, remain, however, unavoidably approximate.

A hedge around discrepancy studies' comparison: Studies of interpretative error are hard to compare head-to-head because of the variables that we have been calling to mind as well as differences in study design, variable definitions of discrepancy, differences in mixes of tissues of origin, various canons of case selection, and application of inconsistent classifying taxonomies.

Discrepancy rates: In the complicated context of interferences and modulating factors, that we have just considered, the range of published discrepancy rates is wide. They are, however, stratified relevantly by the different subject matters that they survey: different anatomic origins of the reviewed specimens, different breadth of focus on reviewed characteristics, and different numbers of cases in the reviewed series. Within this wide frame of references, published series do produce a "range of ranges" of discrepancies.

"Range of ranges": A series that take in large numbers of various specimen types anchor the low end of the spectrum (or, better, spectra) of discrepancy rates. A recent well-organized internal random review of surgical pathology reports ($N=1523$) found a discrepancy rate of 2.2% [46]. Such relatively low magnitudes of review differences can be expected from wide-angle, all comers, and fractional reviews.

The next segment of the discrepancy rate spectrum takes in malignancies from specific organ systems (e.g., lymphoma or urological malignancies), all specimens

from specified anatomic locations (e.g., gastrointestinal and liver lesions), a specific neoplasm (e.g., breast cancer), and a genre of neoplasms (i.e., pediatric cancers). At the low end of this segment, one finds lymphoma with discrepancy rates of 6–7% ($N=1291$) [2, 29]. In the next higher stretch of the spectrum are urological malignancies (10%; $N=213$) [87] and gastrointestinal and liver lesions (12.4%; $N=194$) [22]. Next, up in this part of the range is breast cancer (16–20%; $N=610$) [26, 31] followed by pediatric neoplasms (25.1%; $N=705$) [75]. At the top of this segment of the "range of ranges" are, from a small but responsibly done study, in a resource-challenged environment, soft tissue tumors (47%; $N=34$) [78].

Cytological:histological correlations: Correlations of diagnoses from different modalities produce an extraordinarily wide range of discrepancy rates. At the low end of this segment, with relatively few discrepancies, is a correlation of cytological with histological samples obtained at the same bronchoscopy procedure but interpreted independently. Cytological histological correlations of specimens from this source produce a discrepancy rate of only 2.3% ($N=231$) [74]. Next up the scale is over-all cytological:histological correlation of cervical histology specimens for which a recent, large well-done study locates the discrepancy rate at 6% ($N=5159$) [8]. This discrepancy rate is similar to that in a smaller but well-designed study of correlations for all female genital tract tumors, where the rate was 6.8% ($N=279$) [14]. Next in line is fine-needle aspiration noncervical cytological:histological correlation. From this source discrepancy rates are higher, 9–12% ($N=898$) [4, 43]. More focused comparisons produce discrepancy rates dramatically higher on the scale: bladder cancer cytological:histological correlations have a 41% discrepancy rate in a carefully done large ($N=508$) study [56], and cytological:histological correlations of negative fine-needle aspirations from breast lesions have a discrepancy rate 46% in a moderately sized study ($N=90$) [5].

Cytological:histological vs. cytological:cytological discrepancies: An interesting observation about the cytological segment of the discrepancy spectrum is that in the same well-sized study much lower cytological:histological discrepancy rates were achieved in an environment where high cytological:cytological review discrepancies were documented. The observers who documented the relatively low, 6%, cytological:histological discrepancy rated cited earlier in the text for cervical specimens reported a very high overall similar 45% cytology:cytology discrepancy rate ($N=13{,}745$) [8]. Their high rate of intercytological discrepancies is also seen in a similarly designed, smaller comparison (e.g., 54%; $N=209$) [10].

Dermatopathological variation: Another wide variation in the range of ranges appears in the main histological:histological review segment. This variation involved discrepancy rates in comparisons of skin biopsies. Similarly-sized studies ($N=589$ [84] and $N=478$) [23] came up with discrepancy rate as different as 6.5 and 35%. In another disparate pair of studies, skin biopsies for pigmented skin lesions found a 14% discrepancy rate ($N=392$) [80], but a similarly-sized ($N=354$) comparison of primary and review diagnoses of skin biopsies found a four times higher discrepancy rate of 56% [18].

Discrepancies in difficult diagnostic situations: Finally, in our selective tour of the "range of ranges" of discrepancy rates, one finds relatively high and wide (20–

60%) discrepancy rates in studies focused on diagnostic situations that are known to be difficult. One example of this is a small study ($N=30$) of liver transplant biopsies that showed a 43% discrepancy rate between a primary pathologist's and an expert's diagnoses [9]. Thyroid cytology is another example. Comparisons that focused on this well-known trouble spot found, in two modestly sized studies of thyroid aspirates ($N=50$ [25] and $N=113$) [12], very high but also very different discrepancy rates of 52 and 34%. A third and fourth example of foci on known difficult diagnoses both come from the female genital tract: a 23% discrepancy in diagnoses of vulvar dysplasia in a small study ($N=60$) [85], and a 26% discrepancy rate in the diagnosis of gestational trophoblast disease in a well-done, large series ($N=1851$) [20].

Discrepancies due to variable application of taxonomies: The most impressive instance of multiple discrepancy studies documenting poor reproducibility in a specific diagnostic situation regards Gleason grading. A large study of discrepancies ($N=2015$)in resected prostate specimens found discrepancies in Gleason grading in 45% of cases [28]. In two moderately sized studies of prostate biopsies ($N=278$ [6] and $N=151$ [51]), (the former, larger study comparing diagnoses from microarrays); both found 42% discrepancy rates.

General patterns in discrepancies or review: The last few paragraphs are just an aerial tour that points out only selected landmarks on the landscape of interpretative error, as it is imperfectly transmitted by discrepancy rates. Pondering review reveals that specific rates are rarely comparable; a general pattern does, however, emerge from wandering across the range of ranges. The widest-angle (all comers or random) reviews produce the lowest discrepancy rates. Reviews of diagnoses from organs or organ systems or genres of linked diagnoses (like pediatric neoplasms) produce higher rates. Reviews focused on specific, difficult diagnostic categorizations produce the highest discrepancy rates. Among histological:histological review, differences in discrepancy rates among studies are particularly wide in dermatopathology. Otherwise, cytological:histological diagnoses agree rather well, at the level of organ-system comparison; this is remarkable, given the noise documented by attempts at correlation is cytological:cytological reviews. Finally, among the most commonly used classifications, Gleason grading produces the most discrepancies on review [6, 15, 28, 87].

An information age: Finally, as engineering success of electronic data transfer embeds electronic data transfer devices in surgical pathology's information flow. These devices facilitate high-volume, complex, tightly coupled systems. Such systems are normally prone to accidents. The devices' interactions with people and tissue samples both cause accidental error and offer counter measures against it. Aware that we live in this ambiguous environment we do well to act accordingly.

By attending to the steps in the report production process, we can minimize the occurrence of practical errors. When we get to the root causes of practical errors that amended reports memorialize, adjust the process, then see whether the adjustments reduce error frequency, the process, as a whole, benefits.

"*Errors are indeed there to be made*" [71]: Just as the practical complexity of the surgical pathology report production system requires vigilance to detect errors and

invention of countermeasures to avoid them, so Error will not disappear from making diagnostic interpretations. "Our only route to cognitive progress proceeds along a pathway paved with error—we are creatures to whom truth becomes available only by risking error. Our knowledge grows only by eliminating error" [72]. On this pathway, review is valuable. Review does not, in discrepancies, detect interpretative error as such, instead it finds interpretative error encased in other sorts of variation. Still, discrepancy detection, recognition, and resolution, especially when linked up with statistical reasoning, provide a substantive countermeasure against interpretive error in our discipline.

Conclusion

Two main sorts of error: The landscape of error has two main geographical features: practical errors called process defects and interpretative errors uncovered by diagnostic discrepancies. Study of amended reports reveals process defects. Review of diagnoses produces the diagnostic discrepancies. Both of these strategies have been of value in characterizing and reducing surgical pathology error.

Sources of the two main sorts of error: The dangers that lurk in this landscape are also of two sorts. The concentration, complexity, and tight couplings of electronic information transfer, as we have stressed, both engenders practical defects and provides countermeasures against them. Variable validity, reproducibility, detail in diagnostic interpretations, extensions from particular findings to general diagnoses, variations in classifications and changing evidence bases, we have also argued, all contribute to diagnostic discrepancies that include but are not entirely due to interpretative errors.

Analysis of amendments to understand process error: The reports of accidents, in our initial metaphor, are amendments of surgical pathology reports. Studies of amended reports classify surgical pathology production process errors as misidentifications, specimen defects, misinterpretations, and report defects. These studies document a 5-amendments per 1000 (three sigma) defect rate for current surgical pathology production systems.

Review to uncover discrepancies: Impressions of other travelers, in the guiding metaphor, are reviews of surgical pathology diagnostic interpretations. In this chapter, we emphasize how characteristics of review events whether they are internal vs. external, unfocused vs. focused review as well as, most importantly, the diagnostic domain in question, all influence discrepancy rates. We have presented evidence that internal reviews have, in general, advantages over external reviews and focused reviews have, in general, advantages over unfocused reviews and that, from one diagnostic domain to another, discrepancy rates are dramatically different.

The bottom line: monitor amendments and discrepancies: Published evidence suggests that surgical pathologists' most systematic and sensible design for living in the landscape of error is to monitor process errors, to find and eliminate their root causes, and to review interpretative discrepancies in schemes that factor in a discrepancy's relative likelihood in different diagnostic situations.

References

1. Alassi O, Varney R, Meier F, Zarbo R. Studying amended reports: testing effects of time to sign-out, resident involvement and specimen type on amendments. Mod Pathol. 2013;26(2s):471A.
2. Al-Maghrabi JA, Sayadi HH. The importance of second opinion in surgical pathology referral material of lymphoma. Saudi Med J. 2012;33(4):399–405.
3. Bixenstine P, Zarbo R, Holzmueller C, et al. Developing and pilot testing practical measures of pre-analytic surgical specimen identification defects. Am J Med Qual. 2013;28:308–14.
4. Bomseisl PE Jr, Alam S, Wakely PE Jr, et al. Interinstitutional consultation in fine-needle aspiration cytopathology: a study of 742 cases. Cancer. 2009;117(4):237–46.
5. Bulgaresi P, Cariaggi MP, Bonardi L, et al. Analysis of morphologic patterns of fine-needle aspiration of the breast to reduce false-negative results in breast cytology. Cancer. 2005;105(3):152–7.
6. Burchardt M, Engers R, Muller M, Burchardt T, et al. Interobserver reproducibility of Gleason grading: evaluation using prostate cancer tissue microarrays. J Cancer Res Clin Oncol. 2008;134(10):1071–8.
7. Cankovic M, Varney R, Whitley L, Brown R, D'Angelo R, Chitale D, Zarbo RJ. The Henry Ford Production System: LEAN process redesign improves service in the molecular diagnostic laboratory. J Mol Diagn. 2009;11:390–9.
8. Castanon A, Ferryman S, Patnick J, Sasieni P. Review of cytology and histopathology as part of the NHS Cervical Screening Programme audit of invasive cervical cancers. Cytopathology. 2012;23(1):13–22.
9. Coffin CS, Burak KW, Hart J, Gao ZH. The impact of pathologist experience on liver transplant biopsy interpretation. Mod. Pathol. 2006;19(6):832–8.
10. Coleman DV, Poznasky JJR. Review of cervical smears from 76 women with invasive cervical cancer: cytological findings and medicolegal implications. Cytopathology. 2006;17(3):127–36.
11. D'Angelo R, Zarbo RJ. The Henry Ford Production System: measures of process defects and waste in surgical pathology as a basis for quality improvement initiatives. Am J Clin Pathol. 2007;128:423–9.
12. Davidov TI, Trooskin SZ, Shanker BA, et al. Routine second-opinion cytopathology review of thyroid fine needle aspiration biopsies reduces diagnostic thyroidectomy. Surgery. 2010;148(6):1294–9.
13. Deming WE. Out of the crisis. Cambridge: Massachusetts Institute of Technology Center for Advanced Engineering Study; 1986.
14. Eskander RN, Baruah J, Nayak R, et al. Outside slide review in gynecologic oncology: impact on patient care and treatment. Int J Gynecol Pathol. 2013;32(3):293–8.
15. Fajardo DA, Miyamoto H, Miller JS, Lee TK, Epstein JI. Identification of Gleason pattern 5 on prostatic needle core biopsy: frequency of underdiagnosis and relation to morphology. Am J Surg Pathol. 2011;35(11):1706–11.
16. Finkelstein A, Levy GH, Cohen P, Domfeh A, Parkash V. Addenda in pathology reports: trends and their implications. Am J Clin Pathol. 2012;137(4):606–11.
17. Frable WJ. Surgical pathology-second reviews, institutional reviews, audits, and correlations: what's out there? Error or diagnostic variation? Arch Pathol Lab Med. 2006;130:620–5.
18. Gaudi S, Zarandona JM, Raab SS, English JC 3rd, Jukic DM. Discrepancies in dermatopathology diagnoses: the role of second review policies and dermatopathology fellowship training. J Am Acad Dermatol. 2013;68(1):119–28.
19. Gleick J. The information: a history, a theory, a flood. New York: Panther Books; 2011. pp. 204–32.
20. Golfier F, Clerc J, Hajri T, et al. Contribution of referent pathologists to the quality of trophoblastic diseases diagnosis. Human Reprod. 2011;26(10):2651–7.

21. Gupta R, McKenna R Jr, Marchevsky AM. Lessons learned from mistakes and deferrals in the frozen section diagnosis of bronchioloalveolar carcinoma and well-differentiated pulmonary adenocarcinoma. Am J Clin Pathol. 2008;130(1):11–20.
22. Hahm GK, Niemann TH, Lucas JG, Frankel WL. The value of second opinion in gastrointestinal and liver pathology. Arch Pathol Lab Med. 2001;125:736–9.
23. Hawryluk EB, Sober AJ, Piris A, et al. Histologically challenging melanocytic tumors referred to a tertiary care pigmented lesion clinic. J Am Acad Dermatol. 2012;64(4):727–35.
24. Jain, RK, Mehta R, Dimitrov R, et al. Atypical ductal hyperplasia: interobserver and intraobserver variability. Mod Pathol. 2011;24:917–23.
25. Jing, X, Knoepp SM, Roh MH, et al. Group consensus review minimize the diagnosis of "follicular lesion of undetermined significance" and improves cytohistologic concordance. Diagn Cytopathol. 2012;40(12):1037–42.
26. Kennecke HF, Speers CH, Ennis CA, Gelmon K, Olivotto IA, Hayes M. Impact of routine pathology review on treatment for node-negative breast cancer. J Clin Oncol. 2012;30(18):2227–31.
27. Kramer MS. Clinical epidemiology and biostatistics: a primer for clinical investigations and decision-makers. Berlin: Springer-Verlag; 1988. pp. 11–24.
28. Kuroiwa K, Shiraishi T, Ogawa O, et al. Discrepancy between local and central pathological review of radical prostatectomy specimens. J Urol. 2010; 183(3):952–7.
29. LaCasce AS, Kho ME, Friedberg JW, et al. Comparison of referring and final pathology for patients with non-Hodgkin's lymphoma in the national comprehensive cancer network. J Clin Oncol. 2008;26(31):5107–12.
30. Manion, E, Cohen MB, Weydert J. Mandatory second opinion in surgical pathology referral material: clinical consequences of major disagreements. Am J Surg Pathol. 2008;32(5):732–7.
31. Marco V, Muntal T, Garcia-Hernandez F, et al. Changes in breast cancer reports after pathology second opinion. Breast J. 2014;20(3):295–301.
32. McCluggage WG. Ten problematical issues identified by pathology review for multidisciplinary gynaecological oncology meetings. J Clin Pathol. 2012;65:293–301.
33. Meier FA, Varney RC. Unpublished observation. 2013.
34. Meier FA, Zarbo RJ, Varney RC, et al. Amended reports: development and validation of a taxonomy of defects. Am J Clin Pathol. 2008;130:238–46.
35. Meier FA, Varney RC, Neuman C, Ormsby A, Alassi O, Zarbo RJ. Misidentifications as root causes for amended reports in surgical pathology. Mod Pathol. 2011;24(1s):434A.
36. Meier FA, Varney RC, Zarbo RJ. Study of amended reports to evaluate and improve surgical pathology processes. Adv Anatomic Pathol. 2011;18:406–13.
37. Nakhleh RE. Patient safety and error reduction in surgical pathology. Arch Pathol Lab Med. 2008;132:181–5.
38. Nahkleh, RE. Quality in surgical pathology communication and reporting. Arch Pathol Lab Med. 2011;135:1394–7.
39. Nakhleh RE. Diagnostic error in surgical pathology. Diagn Histopathol. 2013;19(12):433–7.
40. Nakhleh RE, Zarbo RJ. Amended reports in surgical pathology and implications for diagnostic error detection and avoidance: a college of American Pathologists Q-PROBES study of 1,667,547 accessioned cases in 359 laboratories. Arch Pathol Lab Med. 1998;122:303–9.
41. Nakhleh RE, Bekeris LG, Souers RJ, Meier FA, Tworek JA. Surgical pathology case reviews before sign-out: a college of American Pathologists Q-Probes Study of 45 laboratories. Arch Pathol Lab Med. 2010;134:740–3.
42. Nakhleh RE, Myers JL, Allen TC, et al. Consensus statement on effective communication of urgent diagnoses and significant, unexpected diagnoses in surgical pathology and cytopathology from the College of American Pathologists and Association of Directors of Anatomic and Surgical Pathology. Arch Pathol Lab Med. 2012;136:148–54.
43. Nizzoli R, Tiseo M, Gelsomino F, et al. Accuracy of fine needle aspiration cytology in the pathological typing of non-small cell lung cancer. J Thor Oncol. 2011;6(3):489–93.

44. Novis D. Routine review of surgical pathology cases as a method by which to reduce diagnostic errors in a community hospital. Path Case Rev. 2005;10(2):63–7.
45. O'Malley FP, Mohsin SK, Bose S, et al. Interobserver reproducibility in the diagnosis of flat epithelial atypia of the breast. Mod Pathol. 2006;19:172–9.
46. Owens SR, Dhir R, Yousem SA, et al. The development and testing of a laboratory information system-driven tool for pre-sign-out quality assurance of random surgical pathology reports. Am J Clin Pathol. 2010;133:836–41.
47. Perrow C. Normal accidents: living with high-risk technology. Basic Books: New York; 1984. pp. 62–100.
48. Perrow C. Normal accidents: living with high-risk technology. Basic Books: New York; 1984.
49. Perrow CB. Complexity, catastrophe, and modularity. Sociol Inq. 2008;78:162–73.
50. Perrow C. The next catastrophe: reducing our vulnerabilities to natural, industrial, and terrorist disasters. Princeton: Princeton University Press; 2011.
51. Poulos CK, Daggy JK, Cheng L. Preoperative prediction of Gleason grade in radical prostatectomy specimens: the influence of different Gleason grades from multiple positive biopsy sites. Mod Pathol. 2005;18(2):228–34.
52. Raab SS, Grzybicki DM. Secondary case review methods and anatomic pathology. Am J Clin Pathol. 2010;133:829–31.
53. Raab SS, Grzybicki DN. Quality in cancer diagnosis. CA Cancer J Clin. 2010;60:139–65.
54. Raab SS, Grzybicki DM, Janosky JE, et al. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. Cancer. 2005;104:2205–13.
55. Raab SS, Meier FA, Zarbo RJ, et al. The "big dog" effect: variability assessing the causes of error in diagnoses of patients with lung cancer. J Clin Oncol. 2006;24:2802–14.
56. Raab SS, Grzbicki DM, Vrbin CM, Geisinger KR. Urine cytology discrepancies: frequency, causes, and outcomes. Am J Clin Pathol. 2007;127:946–53.
57. Raab SS, Grzybicki DM, Mahood LK, et al. Effectiveness of random and focused review in detecting surgical pathology error. Am J Clin Pathol. 2008;130:905–12.
58. Raz DJ, Zell JA, Karnezis AN, et al. Misclassification of bronchioloalveolar carcinoma with cytologic diagnosis of lung cancer. J Thor Oncol. 2006;1(9):943–8.
59. Reason J. Human error. Cambridge: Cambridge University Press; 1990. pp. 173–216.
60. Renshaw AA, Gould EW. Measuring the value of review of pathology material by a second pathologist. Am J Clin Pathol. 2006;125:737–9.
61. Renshaw AA, Gould EW. Comparison of disagreement and amendment rates by tissue type and diagnosis. Identifying cases for directed blinded review. Am J Clin Pathol. 2006;126:736–9.
62. Renshaw AA, Gould EW. Measuring errors in surgical pathology in real-life practice. Am J Clin Pathol. 2007;127:144–52.
63. Renshaw AA, Pinnar NE, Jiroutek MR. Quantifying the value of in-house consultation in surgical pathology. Am J Clin Pathol. 2002;117:751–4.
64. Renshaw AA, Cartagena N, Granter SR, Gould EW. Agreement and error rates using blinded review to evaluate surgical pathology of biopsy material. Am J Clin Pathol. 2003;119:797–800.
65. Rescher N. Error: on our predicament when things go wrong (1–8). Pittsburgh: University of Pittsburgh Press; 2007. pp. 3–17.
66. Rescher N. Error: our predicament when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. pp. 8–9.
67. Rescher N. Error: on our predicament, when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. pp. 1–21.
68. Rescher N. Error: our predicament when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. pp. 51–67.
69. Rescher N. Error: our predicament when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. pp. 10–13.

70. Rescher N. Error: our predicament when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. p. 4.
71. Rescher N. Error: on our predicament, when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. p. 43.
72. Rescher N. Error: on our predicament, when things go wrong. Pittsburgh: University of Pittsburgh Press; 2007. pp. 18–9.
73. Roy JE, Hunt JL. Detection and Classification of diagnostic discrepancies (errors) in surgical pathology. Adv Anat Pathol. 2010;17(5):359–65.
74. Sackett M, Salomao DR, Donovan JL, Yi ES, Aubry MC. Diagnostic concordance of histologic lung cancer type between bronchial biopsy and cytology specimens taken during the same bronchoscopic procedure. Arch Pathol Lab Med. 2010;134:1504–12.
75. Santiago TC, Jenkins JJ. Histopathologic diagnosis of pediatric neoplasms. Arch Pathol Lab Med. 2013;137:1648–53.
76. Santoso JT, Coleman RL, Voet RL, Bernstein SG, Lifshitz S, Miller D. Pathology slide review in gynecologic oncology. Obstet Gynecol. 1998;91:730–4.
77. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27:379–423; 623–656.
78. Sharif MA, Hamdani SNR. Second opinion and discrepancy in the diagnosis of soft tissue lesions at surgical pathology. Ind J Pathol Micro. 2010;53(3):466–70.
79. Shewhart WA. Economic control of quality of manufactured product. New York: D. Van Nostrand; 1931.
80. Shoo BA, Sagebiel RW, Kashani-Sabet M. Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center. J Am Acad Dermatol. 2010;62(5):751–6.
81. Smith ML, Raab, SS. Directed peer review in surgical pathology. Adv Anat Pathol. 2012;19(5):331–7.
82. Smith ML, Wilkerson T, Grzybicki DM, Raab SS. The effect of a lean quality improvement implementation program on surgical pathology specimen accessioning and gross preparation error frequencies. Am J Clin Pathol. 2012;138:367–73.
83. Swapp RE, Aubry MC, Salamoa DR, Cheville JC. Outside case review of surgical pathology for referred patients. Arch Pathol Lab Med. 2013;137(2):233–40.
84. Trotter MJ, Bruecks AK. Interpretation of skin biopsies by general pathologists: diagnostic discrepancy rate measured by blinded review. Arch Pathol Lab Med. 2003;127(11):1489–92.
85. van den Eiden LCG, de Hullu JA, Massuger LFAG, et al. Interobserver variability and the effect of education in the histopathological diagnosis of differentiated vulvar intraepithelial neoplasia. Mod Pathol. 2013;26(6):1874–880.
86. Volmar KE, Odowu MO, Hunt JL, Souers RJ, Meier FA, Nakhleh RE. Surgical pathology report defects: a college of American Pathologists Q-PROBES study of 73 institutions. Arch Pathol Lab Med. 2014;138:602–12.
87. Wayment RO, Bourne A, Kay P, Tarter TH. Second opinion pathology in tertiary care of patients with urologic malignancies. Urol Oncol. 2011; 29(2):194–8.
88. Wechsler J, Bastuji-Garin S, Spatz A, et al. Reliability of the histopathologic diagnosis of malignant melanoma in childhood. Arch Dermatol. 2002;138:625–8.
89. Weydert JA, De Young BR, Cohen MB. A preliminary diagnosis service provides prospective blinded dual-review of all general surgical pathology cases in an academic practice. Am J Surg Pathol. 2005;29(6):801–5.
90. Wittgenstein L. Tractatus logico-philosophicus 6(341). London: Routlege & Kegan Paul; 1922. pp. 172–5.
91. Woolgar, JA, Ferlito A, Devaney KO, Rinaldo A, Barnes L. How trustworthy is diagnosis in head and neck surgical pathology? A consideration of diagnostic discrepancies (errors). Eur Arch Otorhinolaryngol. 2011;268:643–51.
92. Woolgar JA, Triantafyllou A, Thompson LDR, et al. Double reporting and second opinion in head and neck pathology. Eur Arch Otorhinolaryngol. 2014;271(5):847–54.
93. Zaino RJ, Kauderer J, Trimble CL, et al. Reproducibility of the diagnosis of atypical endometrial hyperplasia. Cancer. 2006;106(4):804–11.

94. Zarbo RJ, D'Angelo R. Transforming to a quality culture: the Henry Ford production system. Am J Clin Pathol (Pathol Patterns Reviews). 2006;126(S1):21–9.
95. Zarbo RJ, D'Angelo R. The Henry Ford production system: effective reduction of process defects and waste in surgical pathology. Am J Clin Pathol. 2007;128:1015–22.
96. Zarbo RJ, Tuthill JM, D'Angelo R, Varney R, Mahar M, Neuman C, Ormsby A. The Henry Ford production system: reduction of surgical pathology in-process misidentification defects by bar code-specified work process standardization. Am J Clin Pathol. 2009;131:468–77.

http://www.springer.com/978-1-4939-2338-0

Error Reduction and Prevention in Surgical Pathology

Nakhleh, R.E. (Ed.)
2015, VIII, 223 p. 25 illus., 14 illus. in color., Hardcover
ISBN: 978-1-4939-2338-0