Bone Marrow Diagnostic Discordance Determination: A Foundation for Clinical Decision Support

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Andrea Renee' Pitkus

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Donald P. Connelly, Advisor

May 2010

Acknowledgements

I wish to express sincere appreciation for the support and guidance of my advisor, Dr. Don Connelly, over the course of this research project. In addition, many thanks are due to Dr. William Swaim, Dr. Sandy Potthoff, and Dr. Stuart Speedie for the support, advice, and time serving as my committee members. Heartfelt gratitude is also due to Dr. Lael Gatewood for the support and opportunities she has provided in international informatics education. I would like to thank Dr. Kathy Foucar, Dr. Elliot Foucar, Dr. Robert Skeate and Christine Spisla for their advice and consultation. I also appreciate the support of colleagues at the University of Minnesota and the University of Heidelberg, especially Dr. Thomas Wetter and Dr. Diana Schmidt during this research project.

Thank you also to my family and friends who have provided needed support during this time. This dissertation is written in memory of Bernice Anderson, who succumbed to her battle with multiple myeloma while this dissertation was being written. The author also wishes to acknowledge those who have fought the valiant fight against hematological disorders in the past, the present and in the future and hopes that cures are found sooner, rather than later. Lastly, thanks be to God for the grace and blessings to be able to undertake the vocation of a graduate student culminating in this thesis.

I also would like to acknowledge the Minnesota Supercomputing Institute for their technical support of this work and use of their computer facilities. Thank you to the University of Minnesota Cancer Center Informatics Resource for the use of the Nucleus Database. This research was supported by the National Library of Medicine Grant #LM070401.

Abstract

Bone marrow testing by the hematopathology, flow cytometry and cytogenetics laboratories provides valuable information utilized in the diagnosis, prognosis and treatment of leukemias. Not much is known about unexpected informatics issues which arise during the analysis of bone marrow, which impact information about the patient's hematological status. This status needs to be clearly communicated to the clinician since it impacts clinical decision making and patient care. This research addresses whether bone marrow diagnostic discordance can be utilized as an indicator of issues in the bone marrow information process, providing the foundation for clinical decision support tool development.

The study first measures disagreement in the diagnoses reported by the three laboratories, on bone marrow specimens collected at the same time, to determine lexical diagnostic discordance. Semantic diagnostic discordance is determined utilizing the 2001 World Health Organization leukemia classifications. Statistical significance of diagnostic discordance is measured with Cohen's Kappa statistic.

The second research phase categorizes factors contributing to the discordances found in the first phase to further understand the etiology of the discordances. It is important to distinguish discordances due to expected testing process limitations from unexpected discordances due to other etiologies. It is also vital to denote which are clinically significant and likely to impact patient care. These factors are critical in designing an effective decision support tool which alerts the clinician appropriately.

Results of the first research phase show lexical and semantic discordance can be measured successfully from three laboratories reporting on bone marrows. Cohen's Kappa statistic also provides an automatic means of detection and measurement of semantic discordance. Categorization of discordances distinguishes which discordances are due to limitations in laboratory testing. Categorization also indicates where in the testing process interventions such as a decision support tool are optimally placed in alerting pathologists of problems in the information process needing further assessment.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Every year adults and children are diagnosed with leukemia, one of the blood cancers that comprise almost 10% of the total cancer deaths in 2009. Leukemia comprises 1/3 of pediatric cancers, with acute lymphoblastic leukemia the most common amongst 1-7 year olds. The leukemia diagnostic process involves decision making by pathologists as they interpret the findings of laboratory and pathology results. Many steps are involved in the information process from the origination of the findings during laboratory analyses to the interpretation of the results by the clinician. Problems may occur at any step in the information process that can affect clinical decision making and patient care. This study assesses the information process associated with the analysis of bone marrow biopsy specimens utilized to diagnose acute myeloid and acute lymphoid leukemias. Information learned from this study will provide the foundation for the development of a clinical decision support tool designed to alert pathologists about these issues arising from bone marrow biopsy analysis. This will allow clinical decision making to be performed with quality patient information so the appropriate and often personalized diagnosis and treatment can be made for optimal patient care.

The research hypothesis addressed in this study is: bone marrow diagnostic discordance can be used as an indicator of potential issues in the bone marrow analysis information process. To test this hypothesis, the research has been designed in two phases.

The aim of the first phase of research is to detect and measure diagnostic discordance among the cytogenetics, hematopathology and flow cytometry laboratories performing bone marrow biopsy analyses on the same specimen. Diagnostic discordance is a quality indicator used in anatomic pathology to make the pathologist cognizant of quality issues in laboratory testing. Research is lacking concerning the use of such alerts with bone marrow testing. The hypothesis in this phase is that bone marrow

biopsy diagnostic discordance can be detected and measured. Bone marrow biopsy diagnostic discordance is measured both lexically and semantically. Lexical diagnostic discordance occurs when there is not word for word agreement in each laboratory's diagnosis. Semantic diagnostic discordance occurs when there is disagreement in the reported disease concept. Cohen's Kappa provides a statistical measure of agreement among the diagnoses. Detection and measurement of disagreement among the diagnoses is an indicator of a potential problem in the testing process. Whether this detection and measurement is performed lexically, semantically or via Cohen's Kappa statistic, it can be utilized in a clinical decision support tool to alert the physician of a problem with the bone marrow testing process prior to patient care decisions being made.

The second phase of research aims to classify the factors contributing to the discordances found in the first phase. The results of this phase of research will distinguish between those factors which impact clinical decision making and patient care the most from those impacting patient care the least. It is also important to distinguish which factors are due to limitations in the testing process from those factors that may indicate error. Adoption is fostered by accurate identification of true positives and true negatives, while ensuring false positives and false negatives are negligible. A key consideration in designing a clinical decision support tool is to incorporate the appropriate level of physician alerting. Maximum physician adoption of a decision support tool will occur with alerts that are not too frequent or infrequent. In addressing these concerns, the second phase of this research categorizes factors contributing to discordance. The hypothesis is that these categorizations distinguish those discordances requiring a physician alert and those which do not. The capability to distinguish which factors have a high impact on patient care and those which are false alarms due to laboratory testing limitations is crucial in designing an effective clinical decision support tool which aids, rather than disrupts the physician's workflow.

Chapter 2

Background and Literature Review: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Origin of Diagnostic Discordance

Diagnostic discordance has always occurred. However, it has only recently been studied and reported in the literature. Studies about diagnostic discordance have significantly increased since the advent of the Institute of Medicine's (IOM) *To Err is Human* report (1). Almost all of these studies in pathology have been limited to anatomic pathology, while few studies even mention bone marrow diagnostic discordance. None of these studies analyze this topic from an informatics point of view. Perhaps this is due to the complicated nature of studying this topic.

Several factors contribute to the challenges of studying bone marrow diagnostic discordance. To understand these factors, diagnostic discordance is discussed from two perspectives. The first is how diagnostic discordance is defined and the second is how it is detected, both of which shed light on the origins of the problem. Next, information about what is known about the problem is discussed from several perspectives, all of which provide the foundation upon which this research is based. These perspectives include what normally occurs in the bone marrow testing process, what happens when issues occur in this process, what informatics knowledge and tools are involved in the clinical laboratory, and what is known about clinical decision support tools, especially those used in the laboratory environment. Each is explained, including how these factors impact the analysis of the bone marrow specimens and resulting information. Lastly, methods that have been utilized to solve the problem are presented. The focus of this research is on informatics and clinical decision support tools, as well as quality assurance methods, determining where they all are most and least effective. However, more recent advances in clinical decision support tools are discussed, especially in how they can be utilized in detecting bone marrow diagnostic discordance and alerting pathologists to a potential problem related to the information process.

Beginning with the definition, this author defines bone marrow biopsy diagnostic discordance as a disagreement among any of a laboratory's pathologists in their reported diagnoses on bone marrow biopsy specimens collected at the same time. Raab similarly defines a discrepancy in the field of pathology as, "a difference in interpretation of reporting between 2 pathologists," (2). The term discordance is utilized throughout this research to avoid the negative connotation of similar terms such as discrepancy, error, disagreement and the like.

Discordance in bone marrow aspiration specimens was reported in the literature as early as the 1970s by Jacobs, Hann, and by Golembe at the University of Minnesota (3-5). The discordance described in these studies involved two separately sampled bone marrow specimens. Diagnoses at that time were made utilizing the gold standard technique of morphological testing. Conventional cytogenetics, molecular diagnostics, and flow cytometry testing were either in their infancy or not yet utilized clinically at that time (6). Furthermore, early testing methods did not provide as much information about the disease process as they do today (6-8). The prevailing professional culture was that pathologists rarely make errors so the topic was, "hardly worthy of discussion," (9).

Regulatory agencies have kept pace with advances in testing with requirements designed to ensure quality. These include the correlation of testing performed on the same specimen in different laboratory areas. The Clinical Laboratory Improvement Act of 1988 (CLIA '88) requires the measure of agreement between the cytologic and histologic diagnoses as a measure of laboratory quality (10-13). The Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) and the College of American Pathologists (CAP) require adherence to the CLIA '88 regulation in addition to their own requirements as part of the inspection process for laboratories and healthcare organizations (14). The CAP HEM.36250 checklist item requires that laboratories evaluating bone marrow specimens in different areas have the capability to compare the data and interpretations from all these specimens (14). The CAP Flow.20100 and CYG.30200 checklist items state that the laboratory must have a documented system to detect and correct significant clerical, analytical errors and unusual results in a timely

manner. Timely is defined as prior to results becoming available for clinical decision making (15, 16). CAP checklist item ANP.12400 requires a mechanism to correlate flow cytometry, cytogenetics, and morphology results (17). Even though detection of diagnostic discordance is implied with the required correlations between laboratory diagnoses, the regulatory agencies do not dictate how discordance should be measured or even defined (14).

Advances in medical science have only intensified the need for a standard to measure diagnostic discordance. The explosion of research in human genomics, the acute leukemias, laboratory testing methodologies, leukemia treatment and prognosis, and the translation of this research into clinical practice have provided even more information about hematological neoplasms. Subsequently, the World Health Organization (WHO) Classification criteria emerged in 2001 to include categorizations based upon morphological, cytogenetic, flow cytometry and molecular diagnostic testing for the diagnosis of these malignancies (7).

At about the same time the WHO classification criteria were placed into clinical practice, the IOM *To Err is Human* report was published (1). This report heightened interest in patient safety and quality monitoring. Subsequently, there has been an influx of articles on diagnostic discordance as a quality and safety issue in anatomic pathology. These articles not only began to define discordance, but also focused attention on methods of its detection and measurement. Details about how these quality assurance methods are utilized in further addressing diagnostic discordance in anatomic pathology are discussed later.

Despite growing interest in patient safety and quality healthcare, few studies have been reported that define, detect or measure bone marrow diagnostic discordance. Despite increased awareness and publication on quality assurance contributions to the detection of bone marrow diagnostic discordance, standardization is lacking in its definition and measurement. To further understand this problem from an informatics point-of-view, one needs to know what has been published about bone marrow biopsy testing issues.

What is Known About Bone Marrow Diagnostic Testing?

This question is first addressed by stating what is known about the normal bone marrow testing process as stated in the literature. This includes literature from an informatics perspective describing aspects of the testing process from the generation of laboratory data to the results and diagnosis reported to the clinician detailing the patient's disease status. The informatics perspective also includes how reported information is utilized by the clinician in the diagnosis, prognosis and monitoring of a patient with hematological abnormalities.

Normally, the bone marrow testing process begins with a physician order when a hematological abnormality is suspected (18, 19). The order may or may not originate as a result of an abnormal complete blood count (CBC) result that is routinely part of the examination process (19). The physician typically orders bone marrow analyses for new cases of suspected hematological disorders. These analyses include testing of the bone marrow by the cytogenetics, flow cytometry, and hematopathology laboratories to gain information about the hematological status of the patient. These orders are considered part of the pre-analytical phase of the testing process that occurs prior to the arrival of the specimen in the clinical laboratory.

The next step is collection of the bone marrow, also part of the pre-analytical phase of testing. A pathology resident, pathologist or other clinician obtains bone marrow samples from the patient as described in Riley's bone marrow biopsy guide and included in each laboratory's standard operating procedure (18, 19). Often a medical technologist accompanies the clinician to make slide preparations at the bedside and ensure that all specimens are correctly labeled (18, 19).

After collection, specimens are delivered to the appropriate laboratories for testing, thus ending the pre-analytical phase of testing and beginning the analytical phase of testing. Hematopathology or morphology is the classical gold standard of hematological neoplasm classification. Slides prepared from bone marrow speciments are stained with stained with Wright-Giemsa, cytochemical or immunohistochemical stains. A

peripheral blood sample is collected from the patient at the time of the bone marrow biopsy and evaluated in conjunction with the bone marrow preparations. The pathologist evaluates all samples that were obtained at the time of collection for cellular architecture, cellular distribution, cellularity, presence of megakaryocytes, abnormal aggregates or infiltrates, fibrosis, abnormal intracellular and extracellular material, and morphology. A 500 cell differential is performed on the best slides and all morphological aberrations such as cellular inclusions, mitotic figures, abnormal lobulation, the presence of multiple nuclei, and cellular structure or size aberrations and cell types are noted (20).

In addition to the evaluation above, the medication list of the patient and recent hematology parameters such as the CBC in the report are noted. Upon review of clinical notes and morphological evaluation, a determination is made if further testing is needed. Tests, such as special histochemical or immunochemical staining on the slides are ordered to aid the diagnostic process. The data are assimilated to narrow a list of differential diagnoses down to a single diagnosis to be reported either in the electronic or paper medical record.

Similarly, the immunophenotyping or flow cytometry laboratory receives a bone marrow specimen for analysis and reporting of information about the Cluster Determination (CD) antigens on the surface of the white blood cells. Certain CD antigens are indicative of cell lineage, while others are indicative of the maturity of the cell. Counts and percentages of each CD antibody are noted by the technologist. Plots produced by the flow cytometer also yield information about the size, structure and maturity of the white cells. This information is forwarded to the pathologist of the flow cytometry laboratory for interpretation and rendering of a diagnostic report.

The flow cytometry report contains a descriptive summary of neoplastic cells, including the identification of type, size, granularity, lineage assessment, and immunophenotype according to the United States-Canadian Consensus. The report also includes an assessment of the heterogeneity of the cells, which further distinguishes subsets of cells

in diagnosing and classifying hematological malignancies (21, 22). Finally, the pathologist synthesizes relevant patient information available to issue a final diagnosis via the paper or electronic report to the clinician.

The purpose of cytogenetics analysis is to identify chromosomal abnormalities present in the bone marrow specimen. In the laboratory, cultures are set up to grow the cells present in the specimen. An aliquot of specimen is added to a culture bottle with appropriate culture nutrients, antibiotics, and media (23). After several days of culture, the cultured cells are removed from the media, washed, resuspended, and fixed (23). Next, slides are made with the fixed cells and stained. A medical technologist examines the slides under a microscope, noting morphology of the chromosomes, quality of the slides, quality of the banding, ploidy, and karyotype of the patient. Abnormalities detected by cytogenetics testing include: translocations, deletions, inversions, duplications, and numerical aberrations (23). Another cytogenetics test is Fluorescent In-Situ Hybridization or FISH, which involves utilizing a specific probe to fluorescently label a unique segment of a chromosome, many of which are diagnostic of specific diseases, like leukemias.

Cytogenetics diagnoses are frequently characterized by specific translocations or other cytogenetics abnormalities. Recent advances in genomic research are incorporated into some of these diagnoses which include the cytogenetic characteristics of bone marrow that are indicative of the prognosis and diagnosis of hematological disorders (24-26). The pathologist communicates this final diagnosis to the clinician including notations as to which leukemia(s) are consistent with the karyotype, in the electronic or paper cytogenetics report.

Diagnostic determination performed by the laboratory pathologist is dictated by the 2001 World Health Organization (WHO) criteria system, a peer consensus of leukemia and lymphoma classifications. The WHO system is an updated version of the French American British (FAB) and the Revised European/American Lymphoma (REAL) systems, previously validated as the scientifically accurate, clinically relevant, and

reproducible gold standard (27-30). Hematological neoplasms as defined by the WHO criteria, are a "combination of morphology, immunophenotype, genetic features and clinical features," (30), rather than a single gold standard test (25, 27, 28). The WHO categorizations even contain categories that are specific for certain cytogenetic translocations that occur in specific leukemias (31).

All phases of laboratory testing and reporting are affected by the requirements of the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), the College of American Pathologists (CAP), the Clinical Laboratory Improvement Act of 1988 (CLIA '88), and the Centers for Medicare & Medicaid Services (CMS). Funding and operations of health care organizations are impacted by the fulfillment of the requirements of these organizations, so it is in the best interest of health systems to utilize quality measures to provide quality patient care, participate in survey and checklist programs, maintain accreditation, and participate in regular inspections (14-17, 32, 33). There are also specific requirements and measures of quality influencing bone marrow testing and reporting which are dictated in the CAP checklists.

Communication of the final diagnosis via the report to the clinician is the end result of the analytical phase of each laboratory's testing process. All aspects of the testing process occurring after the issuance of the report is considered part of the post-analytical phase of testing. The post-analytical phase includes physician decision making in patient diagnosis, prognosis, treatment, monitoring, the detection of minimal residual disease, and other aspects of patient care (34). Diagnoses impact clinician decision making since physicians do not order risky treatments such as chemotherapy, radiation or bone marrow transplantation without a certain level of confidence in a final diagnosis (34). Cytogenetics and molecular diagnostics findings are often predictors of patient prognosis, outcomes, and dictate the use of specific treatment regimens (34).

Each laboratory type performs bone marrow analysis utilizing a different testing methodology to provide information describing the patient's disease process. However quality control measures regulate most aspects of the testing process. Complete details

of each procedure utilized in the processing and analysis of the bone marrow can be found in the procedure manual of each laboratory.

What is Known About Issues Impacting the Diagnostic Information Testing Process?

Normally, the clinician is dependent on the findings of bone marrow analyses in determining patient care. However, internal and external laboratory issues occurring with the bone marrow analysis process often impact the quality of the information reported. These issues, where they occur in the testing process, and how they impact the information process, are all reviewed to understand how all these factors contribute to diagnostic discordances.

In the pre-analytical phase of testing, the two main processes are ordering and collection. In the ordering process, the most problematic step is ordering the wrong test, which at times may also be on the wrong patient (9). Bone marrow testing is not ordered without a clinical indication, as the collection process involves discomfort and some risk to the patient (9). Consequently, all test orders associated with a bone marrow procedure should be placed at the time of the initial order to ensure enough sample is obtained (9).

Ordering and specimen collection usually do not usually present information problems, but problems may occur when unexpected information is present in the wrong patient's chart or expected information is not in the correct patient's chart. If this error is not noticed by the clinician, clinical decision making and treatment can occur based upon incorrect information. Furthermore, delays or incorrect tests ordered on the correct patient may cause diagnosis and treatment delays until the errors are caught and rectified.

During bone marrow collection, specimen collection issues may also arise. One study indicates that the most common cause of diagnostic discordance in anatomic pathology is attributed to sampling issues (35). Poor sample quality issues include collection of a fibrotic specimen, dilution of the specimen with blood, tissue destruction artifact, or

collection of an insufficient volume of specimen. The disease process causes the specimen to be fibrotic. Inadequate volume can be a result of human judgment or technique, or due to disease processes causing a fibrotic marrow, or insufficient marrow at the collection site. A sample diluted by blood can also be affected by collection technique when a bloody area is sampled, the wrong site is sampled, or the sample is not deep enough to reach the marrow (19). A diluted sample may also be caused by a dilute and bloody marrow that is truly representative of the disease status of the patient. All of these issues pertaining to poor sample quality can cause bone marrow artifacts that are revealed during the testing process. Issues may be so severe as to produce little or no information about the patient's hematological status during analyses (9, 19). Sample mislabeling is another common issue occurring during collection.

Another pre-analytical issue is uneven distribution of disease at the sample site. Several clinicians have noted this finding when one specimen indicates disease and another specimen does not, both from the same collection site and at different collection sites (5). Additionally, unexpected manifestations of the disease process, differences in disease dissemination, influences of the micro environment, differential responses to therapy, recirculation of small tumor cells, transformation of a low grade disease process, presence of more than one unrelated neoplasm, or evolution of the tumor showing histological transformation resulting in a change of antigen expression can lead to diagnostic discordance (36, 37). All these effects of the disease process can contribute to bone marrow diagnostic discordance.

Issues with preparation of the specimen at the bedside also occur in the pre-analytical phase. Human impact is evident when poor or an inadequate number of slides or crush preparations are made due to poor technique. Whereas, when one is unable to make enough of these preparations or slides due to an insufficient, poor, or clotted specimen, the disease impact is evident. Detection of these problems usually does not occur until after the specimen has reached the laboratory and microscopically viewed slides are noted as having different cellular numbers or morphologies (9, 19, 38).

Disease effects that contribute to these different numbers and morphologies include an uneven distribution of the sample due to clumping, fibrin, a fibrotic specimen, or insufficient marrow and marrow volume. Effects of the disease may also result in too many cells or too few cells creating too thin or too thick of a slide where cells overlap and morphology is indistinguishable (9, 19). Each of these issues with sample preparation or disease distribution can contribute to analytical issues later in the testing process.

Next, the analytical phase of testing may also contain informatics issues arising from the analysis of bone marrow specimens. Disease effects that become evident during testing include rare staining of other lineage cells, loss of antigen expression during processing, loss of architecture during processing, damage of a specimen or slide, and presence of too few cells for adequate analysis (19, 38, 39). In cytogenetics, detection of chromosomal abnormalities in normal patients with unknown significance is yet another disease effect (9). Analytical issues caused by humans can also include loss of antigen expression with processing, availability of too few cells for adequate analysis, loss of architecture in the process or a damaged specimen or slide (19, 38, 39), and inappropriate utilization of testing and quality assurance practices. During testing these all impact the quality of information generated.

The analytical phase also includes the bone marrow diagnostic process. The following disease factors impact the diagnosis and information reported to the clinician. They include morphological diagnosis variations, cases fitting into multiple categories, the reliance on surrogate markers instead of real disease markers, and unexpected variability in disease manifestations such as antigen expression or response to treatment. In addition, indistinct category boundaries may occur when myelodysplastic syndrome is evolving into acute myeloid leukemia. Variability in where a case fits into a categorization bin occurs when cytogenetic features place the disease in a different categorization bin than the morphological features of the disease (9). Human factors, such as judgment, affecting bone marrow findings include the subjective variability of categorizations (38, 40), variability in where a case fits into a categorization bin, cases

fitting into multiple categories, and differences in opinion about which results most influence categorization (38, 40). Utilization of subjective decision making in the diagnostic process (38, 40), the influence categories have on the diagnostic process (38, 40), and changes in treatment protocols which affect classification to the extent that measures of "clinical relevance" influence classification are also human factors. The impact of new diagnostic testing such as molecular diagnostics and cytogenetics combined with old categories, political and social factors in each department affecting the acceptance of classifications, "tacit knowledge influences" (38, 40), and reliance on surrogate markers instead of real disease markers (9)also are human factors. All of these factors impact the bone marrow information analysis process and may result in diagnostic discordance.

Report generation is also included in the analytical phase of testing. The pathology report from each laboratory includes notations about the bone marrow examination findings, aspiration site, and specimen quality such as whether the specimen was clotted, dilute, or otherwise normal (19). The report should include references to additional specimen testing, but this does not always occur. The report typically concludes with a definitive diagnosis reflecting the hematological status of the patient's bone marrow (19). The report, therefore, is the medium by which the hematological status of the patient is communicated to the clinician. Pathology reports of poor quality or with missing information can negatively impact clinical decision making, since the subsequent treatment plans are based upon the reported findings.

In addition, CAP requires correlation of bone marrow biopsies with previous bone marrow biopsies, other specimens, clinical information, and other ancillary or special studies such as immunophenotyping, flow cytometry, and molecular diagnostics (41). However, CAP does not dictate how this correlation should occur, but requires that "clinically significant discrepancies be reconciled and documented" (14). There are few studies about diagnostic discordance, especially in reducing error and improving patient care (13). Detection and measurement of diagnostic discordance forms the foundation for this research.

Post-analytical activity is the last phase of the testing process. The clinician reads the pathology reports, synthesizes the information, makes decisions upon, and provides patient care, in this phase. Issues and errors that commonly occur in this phase are those involving turnaround time, misinterpretation, miscommunication (42), missing information, report upload errors, report formatting errors (39), typing errors, dictation, transcription, and report verification problems (43). Cytogenetics testing on bone marrows is frequently reported a week or more after the morphology and immunophenotyping are reported to the clinician. The clinician typically acts upon the initial results, but cytogenetics results reported later may also impact prognostic or treatment decisions and cause the clinician to modify patient care. The clinician may misinterpret the initial reports and act accordingly causing unnecessary treatments, costs or patient harm (42).

The impacts these information issues have on patient care have not been widely studied. Clary has studied effects of diagnostic discordances on outcomes such as diagnostic delays, morbidity, mortality or the necessity of additional procedures occurring (13). Clary also found that 63% of gynecologic cytohistologic discrepancies were of no clinical significance, but 37% of these discrepancies resulted in a delay in treatment with no deaths due to errors being reported (13). Raab admits that although diagnostic discrepancies have been studied, their impact on patient outcomes have been studied in less detail (2).

In short, the most frequent and significant errors have been found to occur in the pre-analytic and post-analytic phases (43, 44). Another study pointed to collection problems as contributing the most to laboratory errors, rather than problems in laboratory performance (35). The pre-analytic and post-analytic phases should therefore be targeted to reduce error, even though errors and information issues can occur in any phase of laboratory testing.

Despite regulatory agencies requiring the concordance of bone marrow biopsy testing results from the cytogenetics, hematopathology and flow cytometry laboratories, there

are no standards on how concordance is performed. Standards are also lacking in how laboratories should respond to discordances detected. This study addresses these questions to provide data on how often diagnostic discordance occurs in bone marrow analysis as there are few studies in this area.

Informatics Knowledge and Tools in Bone Marrow Reporting

Next, the information systems utilized for bone marrow reporting will be reviewed. Laboratory information systems (LIS) are the main repositories for laboratory generated information concerning bone marrow testing. Although some laboratories may still utilize a paper based process for all or part of the bone marrow reporting process, most have a laboratory information system. Use of the LIS in bone marrow analysis and reporting is described next.

Primary LIS usage is for storage and management of all information generated in the bone marrow testing process. The information process begins when the clinician generates an order for a bone marrow procedure and testing. An electronic medical record (EMR) or clinical provider order entry (CPOE) system is utilized to initiate the order, which is then transmitted to the LIS. Receipt of the order in the LIS communicates to laboratory personnel that a bone marrow specimen needs to be collected.

Specimen analysis results in the generation of information about the patient. The information is stored in the LIS or in hard copy format from laboratory instrumentation. The pathologist synthesizes these results in producing the bone marrow biopsy report. The bone marrow biopsy report is communicated to the ordering clinician via transmission to the EMR, fax, or direct transfer to another LIS if the specimen was analyzed at a reference laboratory. Regardless, CAP Checklist item GEN.41300 requires patient results to be easily retrievable (32). A contributing factor to Diagnostic discordances may be a result of not having access to the right information at the right time and place

The LIS also functions to store and manage quality control information performed during all aspects of the testing process. Quality control processes often dictated by the CAP ensure that "good" results are reported and laboratory processes are performing as expected. Quality control procedures affect the quality of information reported as well since results cannot be reported unless quality control measures are in control. CAP item GEN.20377 requires patient test results and reports and quality management records to be accessible for at least 2 years (32). The LIS is utilized to document these quality control processes and measures, comply with CAP regulations, and ensure quality results are being reported by each laboratory.

The LIS is also utilized as an internal laboratory communication tool for various aspects of the bone marrow testing process. Communication occurs within the laboratory between the different people handling the specimen and information generated from the testing process. The LIS is used to document and communicate to staff when the specimen was collected, received by the laboratory and the completion of tasks in the testing workflow. The LIS may also contain notations about the quality of the specimen or testing results that affect subsequent tasks in the laboratory workflow, such as an insufficient volume to perform additional testing or a specimen of poor quality. As tasks are completed in the LIS, laboratory personnel usually have access to these preliminary results. Laboratory personnel may have access to findings reported by other laboratory areas impacting their work and the patient's hematological status. Unfortunately, this often does not include pertinent patient information from the EMR impacting bone marrow such as iron, Vitamin B12, and folate levels; chemotherapeutic agents, G-CSF (Granulocytic-Colony Stimulating Factor) and other medications with hematologic effects. These are necessary pieces of information for the pathologist to consider in making their preliminary diagnosis. Often the preliminary diagnosis is reported to the ordering clinician via phone, documented in the LIS, and available to other laboratory personnel. Once testing has been completed, a final diagnosis is issued by the pathologist and is included in the final report communicated to the clinician via the EMR. The final report may also be transmitted to other information databases such as a data repository, research database or database for quality tracking.

More recently, the LIS and its connected databases are being utilized more frequently in clinical research, data mining, and clinical decision support. Historically, the laboratory was one of the first areas of a hospital to be computerized, and therefore has the advantage of producing a wealth of clinical information. Only recently has this rich source of patient and quality data been utilized for data mining and research. Likewise, information provided by bone marrow testing and stored in the LIS can be tapped for cancer, translational, clinical, quality, and patient safety research to improve patient care. There are many uses for this data such as retrospective analysis for diagnostic discordance measurement.

Use of Quality Methods

One potential solution that has emerged to address the problem of diagnostic discordance is the use of quality assurance methodology. Although laboratory testing is rigorously regulated by quality control measures, there has been an influx of additional studies and methods developed as a result of the IOM report. Despite that, there are still issues that have arisen, such as bone marrow diagnostic discordance. The following looks in more detail at quality methodologies as a solution.

Correlation of laboratory testing is a requirement of CAP and CLIA '88. This practice includes laboratories that perform bone marrow biopsies (10-17). Foucar indicates that "assessing diagnostic concordance among pathologists, or between pathologists and cytogeneticists," is included in bone marrow quality assurance measures and "an ideal procedure for monitoring institution-wide quality assurance activities" (9). Normally, laboratories perform diagnostic concordance by manually comparing diagnoses on paper reports or electronically comparing diagnoses with the aid of a module within the LIS or EMR.

Although this regulatory method is designed to detect potential quality issues, the method also has several limitations. The first limitation is the lack of standards or criteria to measure or define diagnostic discordance, (45, 46) including "what level of diagnostic disagreement is clinically acceptable," even with Cohen's Kappa statistic (9).

Hence by lacking standards, there are many ways laboratories perform diagnostic correlations, many means of defining correlation or diagnostic discordance, and many ways to interpret the results. Furthermore when discordance data are inspected by regulatory agencies, there is no way to compare hospital discordant rates without standards (45, 46).

Lack of long term studies and monitoring standards is another limitation which is problematic, even though correlations are performed. Monitoring reveals recurrent problems contributing to discordance such as a clinician with poor sample collection technique. Furthermore, monitoring indicates whether discordance has improved or worsened over time as part of quality improvement initiatives (13, 46, 47). Although CAP inspections may reveal discordances, detection may occur as long as two years after the precipitating event. Long term monitoring is another important component in detecting potential quality issues with the information process (47).

The last limitation of regulatory correlations and diagnostic discordance detection, is the lack of information concerning how diagnostic discordances impact clinical decision making and outcomes. Most methods of discordance detection during quality monitoring occur well after laboratory results have been reported (46). Unfortunately, the timing of these quality measures does not prevent the clinician from acting upon discordant results discovered.

Many laboratories do participate in their own quality improvement or quality management programs to ensure quality laboratory testing in addition to those required by regulatory agencies. One type of voluntary peer based survey is the CAP's Q-Probes studies. Several Q-Probes studies focus on diagnostic errors, diagnostic correlations, discrepancy frequencies, and effects on patient outcome in surgical and anatomic pathology (2, 39, 48).

As Q-Probes findings appear in the literature, some authors have developed standard quality measures from the data. Raab is one of the first to report about the quality of

bone marrow testing, define discrepancies and their effects, and correlate discrepancies with patient outcomes in his key study of anatomic pathology diagnostic discordances (2). Furthermore, Raab indicates Q-Probes is a multi-institutional study measuring anatomic pathology discrepancies detected by secondary review and their effects on patient outcomes (2). Raab's study notes bone marrow diagnostic discrepancy rates by this method to be 6.6%, with almost 86% of discrepancies being due to a change in the same category of diagnosis. The study indicates that bone marrow diagnostic discrepancies also result in no patient harm 71% of the time, but changed reports occur about half the time (2).

One study surveyed hospitals to determine how cytologic-histologic correlations are performed in order to establish standards and tools for this process (46). Another survey focused on long term monitoring of frozen versus permanent section discordant diagnosis frequencies, indicating longer monitoring decreases the frequency of discordance (47). Foucar stated that one way to evaluate pathology diagnoses and their relationship to patient outcomes is to, "consider the diagnosis itself an outcome."(9) Several studies in anatomic pathology also utilized this methodology. Cioc describes how her institution implemented a method for performing and documenting correlations prior to finalization of pathology reports (49). Concurrent review of cervical biopsies and cytologies were performed and discrepancies were codified and documented on the final report for clinical practice (49).

Blinded secondary review is another quality method that is ideal for detecting diagnostic discordance. However, it is only available in institutions having more than one pathologist. One study utilizing this method found 2% of the total pathology cases were hematological in nature, but 10% featured diagnostic disagreement (50). However, as Renshaw states, secondary reviews are impossible to do as there, "are not enough time and resources to review all cases twice," (51). Renshaw encourages investigation of other quality assurance review methods.

Many quality assurance articles go further and include diagnostic discordance with error detection methodology as a type of error (2). One basis for error detection mentioned by Foucar is, "it is reasonable to assume that when the same case receives two different pathology diagnoses, at least one of these diagnoses is wrong," (9). This author respectfully disagrees that one of the diagnoses is wrong or diagnostic discordance is a form of error. Diagnostic discordance can be an indicator of a potential problem such as a wrong diagnosis or error (45). There are also instances when diagnostic discordance is not an error, rather, an indicator of limitations with the laboratory testing process as seen when morphology is unable to detect cytogenetic translocations. The morphological testing process limits the diagnosis to a noncytogenetic WHO classification, while the results of the cytogenetic testing process allows the pathologist to issue a WHO classification that incorporates the cytogenetic results. Neither classification is "wrong", but is a result of testing limitations. Detection of diagnostic discordance followed by classification of discordances found allows one to distinguish errors from testing limitations and aberrant disease expression.

For discordances resulting from a variety of errors, systems analysis can identify and provide information on factors contributing to error, as well as their effect on patient outcomes(9). Many times, errors do not result in patient harm (2). Identification of factors contributing to error is a key component of a continuous quality improvement plan. Automatic identification of contributing factors integrated into a decision support tool can be utilized to be proactive in quality and error prevention (9). Zarbo looked at error detection classification in surgical pathology to measure error frequencies, clinical impact in the described taxonomy of patient outcomes, and impact of reduction and prevention methodologies (39). Another means of error detection involves examining amended reports (48). The study revealed that amended reports review was most commonly initiated by clinicians who noticed diagnostic discordance (48). Amended reports can also be an indicator of error, but amended reports most often include information resulted after the final report has been issued.

Nakhleh described factors contributing to error, and proposed solutions to reduce error (52). Interestingly, several of these solutions are rooted in informatics. They include a reduction in the reliance on physician memory, improving access to information and standardization of tasks and language, including terminology. The proposed solutions continue with forcing users to use tools which aim to eliminate error in processes such as completely collecting report information. It is recommended to utilize systems which identify errors, such as report proof reading and case reviews. Lastly, there should be a decreased reliance on human vigilance and use of tools such as an EMR (52).

Rarely have these types of studies looked at bone marrow testing since their main focus has been anatomic pathology. However, once studies on bone marrow discordance have been performed, standards can then be established by the hematology community and integrated into clinical decision support tools to address this problem.

Use of Informatics and Decision Support Methods

Although the LIS is utilized for the management and storage of laboratory generated information, clinical decision support tools are frequently embedded in the LIS to aid users with managing the wealth of information. Clinical decision support tools provide the right information at the right time and place and are a viable solution in alerting users to diagnostic discordance from bone marrow testing.

The wealth and complexity of information produced by the clinical laboratory has grown rapidly, especially with the translation of genetic testing from research into clinical testing (53). Bone marrow biopsy testing is no exception. Both pathologists and clinicians today consider a plethora of information in making a diagnosis and providing patient care. Both the complexity of medicine (52) and cognitive overload (54) contribute to errors. Bone marrow biopsy diagnosis is a complex activity involving

the synthesis of information from several areas (53, 55) and therefore is susceptible to error. However, few studies document the frequency of errors in bone marrow testing.

Clinical decision support tools allow clinicians to manage the overload of information in real time. Clinical Decision Support (CDS) is defined as, "providing clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare. It encompasses a variety of tools and interventions such as computerized alerts, reminders, clinical guidelines, order sets, patient data reports, dashboards, documentation templates, diagnostic support, and clinical workflow tools" (56). CDS is integrated into computerized provider order entry, treatment reminders, autoverification, and facilitates access to the answers for clinical questions (57, 58).

Computer based CDS utilizes a variety of methods for its functionality. The foundation often includes rules based systems, fuzzy logic, statistical models such as linear and logistic regression, neural networks, decision trees, genetic algorithms, Bayesian networks or support vector machines (58, 59). Utilization of these methods is frequently determined by the problem of interest or the ability to implement them as part of the larger LIS or EMR.

CDS is an effective tool for improving health outcomes and making knowledge available to users at the right time in order to improve health care (56). CDS has been shown to, "improve prescribing practices, reduce serious medication errors, enhance the delivery of preventive care services, and improve adherence to recommended care standards, "(60). Successful CDS features automatic aid to the clinician and integration within EMR functionality, such as aiding order entry. Additional features of a successful CDS include requiring a clinician, "to record a reason when not following the advised course of action," occurring at the time and place of decision making, and providing an assessment and corresponding care recommendation, rather than an

assessment alone" (60). In the clinical laboratory, CDS improves the efficiency of the laboratory and doesn't add to the workload (61).

In the field of hematology and laboratory medicine, clinical decision support is frequently integrated into middleware, the LIS or the EMR (57). Autoverification is one type of CDS frequently utilized in the validation of normal instrument results. Autoverification transmits the results to the EMR without human interaction (57, 62). Flagged results are held for user intervention and not autoverified. Hematology use of CDS includes a discriminant function classification tool for the diagnosis of acute leukemias by flow cytometry (59) and graphical reporting formats for flow cytometry results (63). Clinical Provider Order Entry (CPOE), a pre-analytic CDS, and a treatment protocol, post analytic CDS, both aid the clinician with hematology testing (64). Hematology guidelines aid users in the utilization of evidence based medicine in clinical practice.

One of the most common CDS embedded within the LIS allows users to write custom rules for their needs (57). In hematology, rules are frequently utilized to cancel duplicate tests, add tests for certain patient populations or to order reflex testing. Rules allow decision support to be custom built for the specific needs of each laboratory. The report generating module is an example of CDS utilized in more specialized hematology testing. This CDS is integrated within the Anatomic Pathology (AP) Information System, this CDS allows users to create templates which assist in standardized report writing and can include drop down menus with phrases to guide the writing process (57). Flow cytometry, hematopathology, cytogenetics and molecular diagnostics are laboratory areas that frequently utilize this feature to customize their reports.

CDS is rarely utilized in bone marrow testing, except in quality assurance reporting. One way users comply with CAP diagnostic correlation requirements is to employ some of the previously described CDS. Custom rules can be developed that compare diagnoses from the clinical laboratories performing bone marrow testing. CoPath Plus

from Cerner Corporation in Kansas City, MO, features the "Diagnostic Correlation" function that prepares reports for the pathologist to determine diagnostic discordance (65, 66). This application allows the user to produce a list of patients with specimens analyzed by two different laboratories in order to compare the diagnoses manually. The pathologist determines the presence and reason for the discordance between the diagnoses instead of the computer. More commonly, this function is used for quality assurance reports.

Although basic CDS often aids the end-user, it may have issues and limitations. One limitation is that all CDS tools are not completely automated as described with the "diagnostic correlation" function. Second, some CDS lacks standards and definitions concerning discordance, leaving that decision up to the end user. Another key limitation with these examples of CDS is that they are designed to trigger an alert after laboratory results have been reported to the clinician. These quality reports may be generated hours or months after the clinician has received the results and had the opportunity to act upon them. One quality study featured a tool developed to list surgical pathology cases and any corresponding cytopathology cases from the previous six month period. Ohori reported quality review results from weekly use of the tool to survey cases (66).

It has been shown that current hematological CDS tools have many issues and limitations with regards to the detection and management of bone marrow biopsy diagnostic discordance. In order to develop a more functional decision support tool to aid the detection and measurement of diagnostic discordance associated with bone marrow pathology, some decision support development questions need to be addressed. In developing tools for decision makers, the focus should be made on the decisions and the technological capabilities of the tools, instead of hardware and software issues as typically found in CDS development according to Nykaenen (61). In addition, environmental factors, "the domain problem and the purpose of the planned system," need to be accounted for in the development and design of CDS, as shown by

Nykaenen's research (61). This research focuses on the information and decision processes involved in bone marrow biopsy testing from an informatics perspective to address the key research questions within the domain and within the decision making context that form the foundation for CDS development.

Chapter 3

Materials and Methods: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Specific Aims and Hypotheses

This two phase project is designed to address the research question of, "Can bone marrow biopsy diagnostic discordance be used as an indicator of issues with the diagnostic bone marrow information process?" The aim of the first phase of this research is to detect and measure discordance among the reported diagnoses generated by the cytogenetics, hematopathology and flow cytometry laboratories. The hypothesis for this first phase is that bone marrow biopsy diagnostic discordance can be detected and measured. This hypothesis is tested for lexical diagnostic discordance, or word for word comparison of the diagnoses, and semantic diagnostic discordance, or comparison of the concepts described by the diagnoses. Cohen's Kappa statistic is used to measure diagnostic agreement statistically. Information learned from the first phase of research will guide which methodology would best detect and measure diagnostic discordance in a decision support tool.

The second phase of research aims to characterize and categorize the factors contributing to diagnostic discordance to distinguish which factors have the greatest impact on clinician decision making and ultimately patient care. In addressing this goal, these categorizations also need to distinguish which diagnostic discordances are due to limitations in the testing process versus those diagnostic discordances due to other causes. The hypothesis tested in the second phase is categorization of factors contributing to diagnostic discordance allows one to distinguish between discordances requiring pathologist alerting and those that do not. Information learned from this aspect is critical in designing decision support which appropriately alerts pathologists to diagnostic discordances.

Experimental Design and Data Collection: Subjects and Cases

This study examines final reports resulting from analysis of bone marrow biopsy specimens by the Fairview Diagnostic Laboratories at University of Minnesota Medical Center, Fairview (UMMC). The specimens analyzed are from inpatients from Fairview Health Services Hospitals, outpatients from the Fairview Clinics or University of Minnesota Physicians Clinics and referrals from outside institutions. The patient population of analysis includes adult and pediatric males and females, where their physician has requested bone marrow testing to investigate the hematological status. This population also includes patients with bone marrow analyses performed at outside laboratories, but have been referred to the University of Minnesota. A University of Minnesota pathologist performs an independent review of the bone marrow slides and data, and issues an independent report for these referrals.

Each bone marrow specimen obtained from a single biopsy procedure has been analyzed by at least two of the three laboratories: flow cytometry, cytogenetics and hematopathology. Each laboratory utilizes the 2001 WHO criteria and diagnostic categorizations for hematological neoplasms in reporting a bone marrow diagnosis (30). The case population studied is limited to those neoplasms within the Acute Myeloid Leukemia and Acute Lymphoid Leukemia categories of the 2001 WHO classification criteria as indicated in Appendix A. These two diagnostic groups have been chosen since they both make use of cytogenetics, hematopathology, and flow cytometry findings in the determination of the diagnoses. In addition, these leukemia diagnoses encompass the most prevalent leukemias in children and adults together comprising 55% of all leukemias (67). An initial survey of the leukemias presenting at UMMC indicated there were a sufficient number of cases available for analysis as shown in Appendix B.

The unit of analysis is the initial bone marrow biopsy diagnostic text reported by each laboratory originating from the same bone marrow procedure. Each case may be comprised of up to three separate reports, one from each laboratory. Subsequent patient

bone marrow examinations are excluded to avoid treatment effects from chemotherapy or bone marrow transplantation. The exception occurs if a patient develops a subsequent hematological neoplasm that is a different disease process. The first initial diagnostic bone marrow specimen of the new disease process is considered a separate case with these exceptions.

Experimental Design and Data Collection: Data Acquisition and Data Sources

After obtaining approval from the University of Minnesota Institutional Review Board (IRB), including an exemption from the Health Insurance Portability and Accountability Act (HIPAA), data collection began. All data in the Fairview Clinical Information System (FCIS) from patients seen from 2000 until 2005 were retrospectively analyzed after obtaining access to this electronic medical record (EMR). A search of patient bone marrow reports was attempted by diagnosis, but limitations of the EMR search engine required use of the patient medical record number (MRN) instead.

To find the MRNs associated with the bone marrows reported in FCIS, the University of Minnesota Cancer Center's Nucleus Database System (NDB) was utilized. The NDB is searchable by diagnosis and populated with all UMMC cancer patient diagnoses after January 1, 2001 (68). After IRB approval was received for the addition of the NDB to the study, the database was searched by ICD-O-3 diagnostic codes corresponding to the WHO Myeloid and Lymphoid Neoplasms Classifications as indicated in Appendix A. A list of MRNs was returned and employed in searching FCIS for the initial bone marrow reports for each patient.

Experimental Design and Data Collection: Data Storage

Once bone marrow biopsy reports were located in FCIS, an electronic copy of each report was made for storage and further reference. For storage of case reports over the course of the research project, a secure database conforming to University and HIPAA regulations was established at the Minnesota Supercomputing Institute (MSI). Population of the database was facilitated through the use of a secure web-based user

interface. A sequential case number, medical record number, accession number, WHO diagnostic category, specimen collection date, and final diagnosis and report copy were all entered into the database for each new case. The database was searchable by medical record number, accession number, case number, reporting laboratory, WHO diagnostic category, and final diagnosis. The MSI database has been maintained and supported by the Minnesota Supercomputer Institute, including daily data backups.

Phase I: Determination of Lexical Diagnostic Discordance

The operational definition of diagnostic discordance for this research is any disagreement in the diagnoses determined by any of the laboratories which issued a bone marrow biopsy report on a specimen collected from the same bone marrow procedure. As the data were reviewed, it became apparent that diagnostic discordance would have to be further subdivided into two groups. The first group is lexical diagnostic discordance or disagreement in word for word comparison of the laboratory diagnoses. The second group is semantic diagnostic discordance, disagreement in the disease concept described by each diagnosis. The hypotheses previously described for this first phase were tested for both lexical and semantic diagnostic discordance.

Word for word comparison of the diagnoses was performed to test the hypotheses for lexical discordance. Different lexical comparison techniques are employed in laboratory information systems. Some lexical comparisons integrate a rules-based methodology whereby text strings are compared and the result is either complete agreement or complete disagreement of said strings. Although automated tools were not employed for this research, lexical comparison methods in this research are based upon the premises utilized by some anatomic pathology clinical decision support tools.

Data organization and analysis was performed with Microsoft Excel spreadsheets. One spreadsheet encompassing diagnoses for all three laboratories side by side was created as well as individual spreadsheets for comparing each laboratory pair: hematopathology versus flow cytometry (HF), hematopathology versus cytogenetics

(HC), and flow cytometry versus cytogenetics (FC). These spreadsheets were created for both lexical diagnostic comparisons and semantic diagnostic comparisons.

Focusing on lexical agreement, case numbers and laboratory diagnoses reported by each laboratory from the MSI database were entered into the spreadsheet. Next, diagnoses for each case were compared to determine if lexical agreement or disagreement was present and the result recorded. The number of cases with lexical agreement was tallied, as were the cases with lexical disagreement. These data were utilized to calculate the percent agreement for each laboratory pair. The percent agreement, a reliability measure, is the percent of all the cases in which judges, or in this case, pathologists, determine the same diagnosis. The percent agreement can be calculated among all cases, for each lab, and across each diagnostic category, but does not indicate if the agreement is due to chance or not (69). The percentage of agreement or disagreement was determined by dividing the number of cases in agreement (or disagreement) by the total number of cases that were compared for that laboratory pair.

These calculations were performed for each laboratory comparison for lexical diagnostic agreement and lexical diagnostic disagreement or discordance. Patterns in agreement among the laboratory comparisons were noted. Measurement of percent agreement was performed in the testing of the hypothesis for lexical diagnostic discordance.

Phase I: Determination of Semantic Diagnostic Discordance

Semantic diagnostic discordance determination was performed similarly to the method used for lexical diagnostic discordance. Diagnoses reported by each laboratory for a case were compared to determine if they described the same disease process and had semantic agreement. The 2001 WHO Hematological Neoplasms Classifications was the disease classification and naming convention standard utilized. Pathologists may not describe a disease process with the same terminology despite the recognition of the WHO Classification Criteria as the accepted classification and terminology standard. Measurement of semantic diagnostic discordance attempts to account for this practice.

Each case was assessed and again a tally was made of the number of cases with semantic agreement or semantic disagreement or discordance.

The author is unaware of any fully automated semantic comparison decision support tools in use for determination of bone marrow biopsy diagnostic discordance. There are however, some semi-automated clinical decision support tools in some laboratory or anatomic pathology information systems which aid the user in determination of semantic discordance. These tools function by presenting the user with a list of diagnoses on specimens collected at the same time for the user to determine diagnostic concordance or discordance with their clinical judgment. One example of such a program is the "Diagnostic Correlation" module offered in some Cerner products. Determination of semantic diagnostic agreement is closely related to how pathologists function in the laboratory.

In making these case assessments, it was found that in some instances agreement determination was not so clear cut. For example, "sibling" relationships between diagnoses such as acute monocytic leukemia and acute monoblastic leukemia were deemed an example of semantic diagnostic discordance. Another instance involved a specific diagnosis being compared to a more general diagnosis. However there was insufficient information reported to determine if the general diagnosis was the same disease or not. These cases were designated as unclassifiable and the number of these cases was also tallied. A diagnosis reported as acute myeloid leukemia with maturation by one laboratory and as acute myeloid leukemia by another lab is an example of an unclassifiable case. If additional diagnostic information was in the report, reclassification could occur so both diseases would be reported as acute myeloid leukemia with maturation and thus be in diagnostic agreement. However, insufficient report information permitted the diagnoses of acute myeloid leukemia without maturation and acute myeloid leukemia with maturation to be considered in the differential diagnosis and thus be considered in diagnostic disagreement. Many of these cases do not contain enough information in the diagnostic report to determine whether the diagnoses agreed or not.

The percent agreement, a reliability measure, was calculated for each of the agreement designations indicating the percentage of all cases where the compared case diagnoses agreed, disagreed or were unable to be ascertained and were designated unclassifiable. Trends in diagnostic agreement were noted, including cases where agreement was unable to be determined and were designated unclassifiable. These methods of measuring agreement tested the hypotheses previously described.

Phase I: Kappa Statistic

The primary statistical method employed in this research is Cohen's Kappa statistic, a measure of inter-observer agreement or reliability. Kappa measures reproducibility of two raters that independently categorize nominal data, such as if a case is positive or negative for a disease category. The assumptions of Kappa are that the cases "are independent, the categories or the nominal scale are independent, mutually exclusive and exhaustive, and the judges operate independently," (70). Landis and Koch successfully measured agreement with Kappa among clinicians in two different geographic locations diagnosing multiple sclerosis to study if differences in the diagnoses are significantly different from chance alone. The hypothesis proven was that Kappa was equal to zero, indicating poor agreement and diagnoses by the clinicians were not significantly different than by chance alone (70). Kappa has been successfully applied to measure inter-rater reliability among hematopathologists assessing bone marrow specimens (71).

SAS statistical software can compute Kappa by using the procedure PROC FREQ with the AGREE option. PROC FREQ also provides the user with the standard error, confidence limits for Kappa, and "an exact p-value testing Kappa <= 0 versus kappa > 0," (72, 73). A p-value < 0.05 is considered statistically significant when measuring Kappa (69). There are some limitations of calculating Kappa with SAS. Kappa is unable to be calculated in a non-square contingency table such as those lacking an observation in each cell. Kappa also will not be calculated if all values of a row or column are equal to zero. One way to avoid these issues is to add additional lines to the

SAS code that change any blank or zero value observations to a negligible value so that Kappa can be successfully calculated (74). This method is more commonly referred to as calculating the weighted Kappa value.

Phase I: Calculation of Kappa

For this research project Kappa could not be calculated for lexical diagnostic comparisons due to great variability among diagnoses. Kappa was successfully utilized for semantic diagnostic comparisons. Diagnoses from each case were transferred from the MSI database to a spreadsheet for analysis. One column indicated the case number, the second column contained the diagnosis issued from the first laboratory and the third column contained the diagnosis reported by the second laboratory. Due to SAS limiting each observation or diagnosis to eight characters, each diagnosis was converted to an eight character code. Each code was linked to the appropriate WHO categorization concept or to the diagnostic text listed on the pathology report. The coded spreadsheet for each laboratory pair was then saved in comma delineated format for SAS analysis.

Each spreadsheet was uploaded to the MSI UNIX cluster for statistical analysis utilizing SAS version 9 (The SAS Institute, Cary, NC). Each spreadsheet and corresponding SAS code file was imported into SAS and analyses were performed. SAS produced several output files containing the data analyses including Kappa. Once Kappa had been obtained for each laboratory pair, this phase of research was considered complete.

Phase II: Categorization of Discordant Cases

Experiments from the second phase of this research project were designed to test the hypothesis that categorization of factors contributing to diagnostic discordances allowed one to distinguish between those discordances which require a physician alert and those discordances that did not. This phase also collected information concerning those issues or factors contributing to diagnostic discordance to gain additional insight as to the nature of diagnostic discordance. All these points are critical in designing an effective decision support tool which aids the clinical workflow instead of disrupting it.

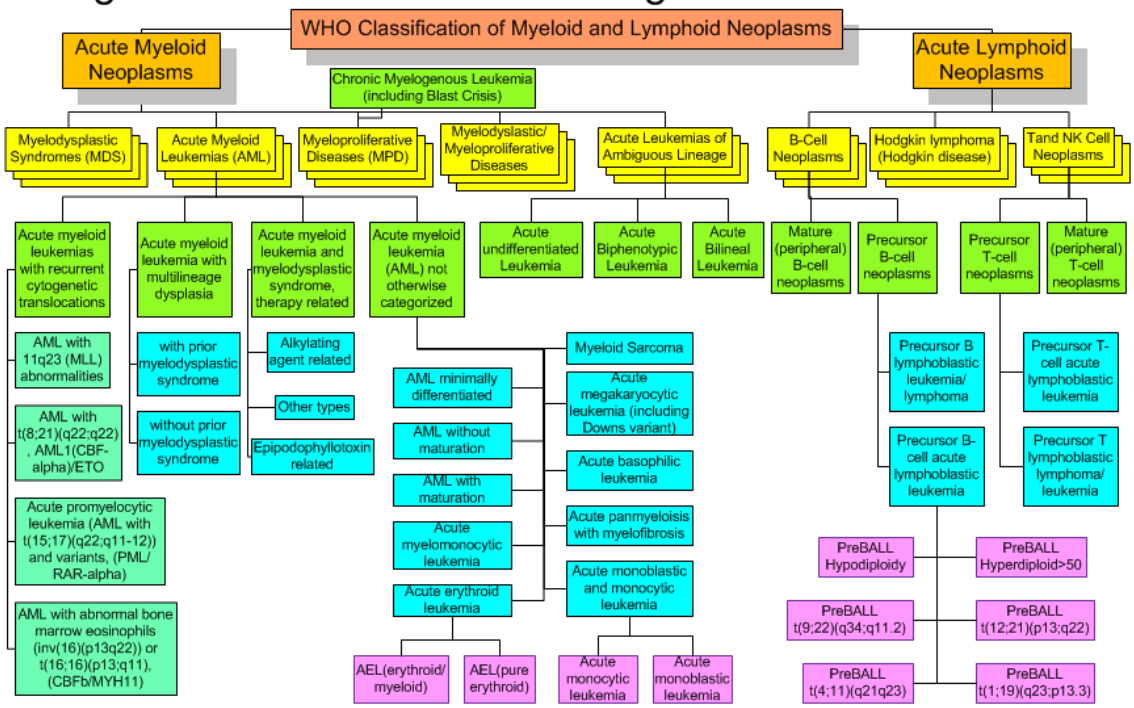
Cases from the first phase of this research which exhibited semantic diagnostic discordance became the basis for study in the second phase of research. Lexical diagnostic discordance cases were not utilized since a number of these cases had slight differences in diagnostic text and were not truly discordant semantically.

The first means of categorization of discordant cases was designated, "Categorization By Level," and referred to the relationship of the diagnoses to the WHO classification criteria levels. The second categorization of discordant cases was, "Categorization by Limiting Factors," and described the contributing factors by laboratory type. The third discordance classification was "Categorization By Where Issues Occur in the Testing Process." This third method pinpointed where the limiting factors originated in the three major phases of the testing process, while the fourth classification addressed more specific causes of these factors in "Categorization by Etiology." This fourth method analyzed both general issues and specific issues contributing to discordance in each phase of laboratory testing. The fifth and last method of categorization of discordant cases, "Categorization by Reporting Errors," gathered information about errors that were noted and amended in the report and those errors which were not noted. Details about how each categorization was performed, follows.

Phase II: Categorization By Level

"Categorization By Level," utilizes the 2001 WHO categorizations as shown in Figure 1 to categorize the relationship between the diagnoses reported by each laboratory. These relationships indicate whether the diagnoses fall into a more general or a more specific categorization. Diagnostic categorizations that are too general can occur when insufficient information is available at diagnosis, resulting in a more general diagnosis characterizing the disease status. General diagnostic categorizations may be an indicator of an issue in the information process.

Figure 1: WHO Leukemia Categorizations and Levels



These relationships may also indicate quality issues. Raab describes an error in disease categorization or diagnosis as an "interpretation error," (12). He further subdivides these errors into an overcall, where the "review diagnosis was categorically lower than the original diagnosis," or an undercall, where "the review diagnosis was higher than the original diagnosis." (12) This research does not classify undercalls or overcalls, as they are both more commonly employed in secondary pathology case review. This research does look at the diagnostic relationships similarly as described below.

This categorization began with discordant cases found in the first phase of research. A line was drawn between the two WHO diagnoses reported for each case in order to diagram the diagnostic relationship. When a modification of the WHO diagnosis was reported by a laboratory, this diagnostic text was added to the figure and connected via a line to the other reporting laboratory's diagnosis. The number of cases exhibiting the relationship depicted by each line is indicated on that line. Diagramming these relationships provides a visual indicator of which diagnoses are more general

categorizations, and which are more specific categorizations. The case counts indicate which relationships occur more frequently among the data.

As the cases are diagrammed, they are also assessed as to which of the four relationship "levels" they exhibit as explained in the following operational definitions. Diagnoses that are related vertically in a parent-child relationship only are said to exhibit a "Level I" or "Vertical" relationship. This definition corresponds to Raab's undercall or overcall designations, exhibiting a general to more specific diagnostic relationship or vice versa. An example of this type of relationship is the subcategory "B-Cell Neoplasms" as related to the subcategory of "Precursor B-Cell Neoplasm."

Diagnoses that are related in a "sibling-to-sibling" relationship only are said to be in a "Level II" or "Horizontal" relationship. These include the relationship between the categories of "Acute Myeloid Neoplasms" and "Acute Lymphoid Neoplasms," as well as the more specific subcategories named for each of the cytogenetic abnormalities as compared to each other. Relationships that are not in a direct vertical or horizontal relation, such are designated in the "Level III: Other" relationship. Relationships of this type may be diagonal or more complex. An example of this type of relationship is that between "Acute Biphenotypic Leukemia" and "Acute Myeloid Leukemia with Maturation." Lastly, case relationships that could not be determined are designated to the "Unable to be Determined" category. These include cases in which there is insufficient information to determine if there is a vertical, horizontal or other relationship, but if additional information were provided the relationship could be determined.

The data were analyzed by tallying the number and percent of cases exhibiting each relationship type. Differences among laboratory comparisons are also noted. These experiments provide further information characterizing the relationships of the discordances.

Phase II: Categorization By Limiting Factors

"Categorization By Limiting Factors," is the next categorization performed on the discordant cases from the first phase of research. This categorization indicates the limiting factors that contribute to the discordance, whether these are laboratory based issues or due to other factors. Descriptions of these limiting factors further characterize the diagnostic discordances. These limiting factors may be indicators of issues with the information process and indicate where a decision support tool may be best implemented to alert the clinician to these issues.

This categorization also aims to provide further information about how limiting factors contribute to diagnostic discordances. For example, was the discordance due to the limitations in laboratory testing and in which laboratory? What factor(s) led to the limited information being reported? Was the discordance due to a non-laboratory source of insufficient information? The operational definition for laboratory limitation is any factor occurring as a result of laboratory testing that resulted in a lack of information necessary for the specific diagnosis of the disease process utilizing the WHO criteria. Examples of laboratory limitations include insufficient laboratory testing performed on a specimen resulting in only enough information to issue a general diagnosis, rather than a specific diagnosis. Laboratory test results lacking an interpretation are also considered a laboratory limitation. An "Other Limitation" example is not having prior patient treatment information available for the pathologist at the time of diagnosis, especially in cases where the differential diagnoses are dependent on that information such as with "Acute Myeloid Leukemia and Myelodysplastic Syndrome, Therapy Related, Alkylating Agent Related." These information limitations are assessed for each case of diagnostic discordance found in the first phase of research and with each laboratory pair. Each laboratory report is reviewed for additional information indicating which limiting factors are involved in the resulting diagnosis.

Once a list of limiting factors has been obtained, it is analyzed to determine if there are any common patterns such as recurring factors and their frequency. The analyses also indicate whether these factors are in common for all laboratories or just prevalent in certain ones. Recurring limiting factors and other patterns are potential indicators of

where decision support alerts or quality measures may be implemented to improve the information processes.

Phase II: Categorization By Where Issues Occur in the Testing Process

"Categorization by where issues occur in the testing process," provides information concerning where testing process issues contribute to diagnostic discordances. Discordant cases from the first phase of research are classified according to which testing process phase the limiting factors from the previous section occurred. Knowledge about where an issue contributing to discordance occurs, allows one to know where to best place a decision support alert or other intervention in preventing future information process defects.

One example occurs in the flow cytometry laboratory. An insufficient specimen volume available prevents testing the full panel of antibodies necessary for subcategorizing an "Acute Myeloid Leukemia." The result is a general diagnosis of "Acute Myeloid Leukemia," issued by the pathologist. Categorization of where in the testing process the limiting factors occurs, reveals that a partial panel of antibody testing occurred in the analytical phase. Categorization also reveals that insufficient specimen volume occurred in the pre-analytical phase resulting in the analytical phase effects. Both factors contribute to incomplete information being reported about the disease process. A simple intervention in preventing this information defect is education on the minimal sample volume requirements of each laboratory.

The three phases of laboratory testing, pre-analytical, analytical, and post-analytical, have been previously described. For this research they were also subdivided into their key components in order to further classify where issues occur in the testing process. Pre-analytical testing was subdivided into "test ordering" and "specimen collection." The analytical phase was subdivided into the "testing," "diagnosis/classification," and "reporting." The post-analytical phase of testing is defined as processes occurring with test results after they have been reported in the LIS including clinical decision making. The post-analytical phase was not studied in this research since the ultimate goals is to

implement a clinical decision alert prior to the reporting of test results and the post-analytical phase.

Each discordant case report is reviewed to determine in which testing phase limiting factors or issues originated. Next, the appropriate subdivision(s) of origin for the issue(s) are noted. Many times, there may be multiple contributing factors or testing phases affected resulting in discordance. These classification methods are applied to each of three laboratory comparisons.

The data are analyzed by counting the number of limiting factors or issues in each testing phase and subdivision. These factors are also compared among the three laboratory pairs for trends and further characterize the discordances.

Phase II: Categorization By Etiology

This fourth method of discordant case classification is called, "Categorization By Etiology." The discordances are reviewed and classified into one or more of the general etiology categories and by any specific etiology issues that contribute to the discordances. The operational definition of etiology used in this project is the source or cause of the diagnostic discordance or issues contributing to the discordant cases. Classification by etiology provides information concerning how discordance occurs. This classification indicates whether a human or the disease process contributed to these issues, as well as the specific causes listed in Appendix C: Diagnostic Discordance Classification by Specific Etiologies. This information aids the determination in where and what type of intervention or decision support would work best for the situation.

Each discordant case from phase one was reviewed to determine if the etiology of the discordance was due to a human issue, the disease process, both a human issue and the disease process, or unclassifiable with the reported information. Some etiologies, such as a fibrotic specimen, are classified as a "both" category since it can be caused by the disease process or by poor sampling technique.

Once these general etiology classifications were noted, then specific etiologies as listed in Appendix C were noted as well. Foucar describes many of these etiologies and their role in diagnostic discrepancies (9). Etiologies occurring earlier in the testing process such as a fibrotic specimen also contribute to etiologies in later phases of the testing process further compounding the problem. Therefore several etiologies can be listed for a case.

Data analysis entails counting the number of etiologies, both general and specific for each case and laboratory pair. Patterns are noted such as which general and specific etiologies occur most often, as well as differences between laboratory pairs. These data provide information as to which etiologies may be rectified via an intervention such as education or a clinical decision support tool or those which an intervention may not be helpful such as the manifestation of the disease process.

Phase II: Categorization By Reporting Issues

"Categorization By Reporting Issues," is the last method utilized to classify discordances. Each discordant case from the first phase of research is reviewed for errors. The review notes any typographical errors, amended reports, contradictory information, or error corrections found within the report as a means to classify both errors that are known and unknown. These "report defects" have been indicated as one of the major categories and indicators of error in a standardized taxonomy of anatomic pathology errors (39). A review of amended report rates and diagnostic discrepancy rates in anatomic pathology are two surveillance methods that have been utilized to detect error in pathology reporting, although not with bone marrow biopsy reporting (39). Furthermore, Zarbo states that, "some errors that arise from nondiagnostic information defects in the pre-analytical aspect of the test cycle, may be evident, if initially uncorrected,"…"as a defective report requiring emendation." (39)

Detection and prevention of errors can be aided by different clinical decision support tools. Knowledge of the type of error guides the informaticist in determining which

informatics solution works best for the situation. For typographical errors, a spellchecker or medical ontology are examples of two informatics based solutions. For other errors, alerts may be the best solution. This project does not specifically study the impact of different report formats on clinical decision making. Studies have been performed on clinical decision making information and what the clinician considers extraneous or less commonly utilized so that reports can be tailored accordingly (75).

"Categorization By Reporting Issues," begins with a review of discordant case reports for errors. The first criterion noted for each report is the presence of a changed or amended report or not. The operational definition of an amended report is one with a notation indicating the correction of an error or addition of laboratory information after the initial report was issued. Reports not amended are noted as "unchanged." Amended reports indicate errors that have been detected, while unchanged reports indicated errors that have been missed or have gone undetected. These categorizations also are a measure of how errors are detected (48).

Each report classification is further subdivided by types of reporting error. The two types of unchanged report errors are typographic errors and contradictory information found on the report. The laboratory issuing the erroneous report was noted as well. An example of contradictory information is one laboratory reporting myeloid blasts, while the other reports lymphoid blasts.

Amended reports are also subclassified by type of reporting error. Information noted includes the laboratory issuing the amended report, and why the report has been amended. Reasons for amended reports include: to include additional or delayed information or to make a correction. Lastly, why a correction was made is noted. All these classifications characterize the presence or absence of an error and the type of error made.

Cases in the first subdivision, the inclusion of additional or delayed laboratory information, are not indicative of error as reports are frequently amended to include additional laboratory test results such as special stains. The second categorization,

corrections on an amended report, is considered a measure of an error that was noted and corrected. The reason for the correction, for patient information or demographics, for specimen information, or for diagnostic information, all further characterize the error.

Analysis of these data includes the quantification of amended and unchanged reports by laboratory type. Trends, such as which laboratories report the most errors, amendments, or typographical errors, are noted. These classifications provide information as to which discordances are true errors, as opposed to amendments. Further information as to which errors remain undetected is also gathered.

Chapter 4

Results: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Phase I: Determination of Lexical Diagnostic Discordance

Assumptions made about data prior to measurement include that each laboratory's testing process is error free and performed under strict control of regulatory and quality assurance measures. For those specimens initially analyzed elsewhere and referred to the University of Minnesota, it is assumed that a University of Minnesota pathologist has reviewed the specimens and related data independent of the previous diagnosis. It is also assumed that when University of Minnesota pathologists reviewed the case, they issued their own diagnosis, which may or may not be the same as the initial diagnosis.

Examples of real case data for the determination of lexical and semantic diagnostic agreement are shown in Table 1. Beginning with lexical discordance, case one illustrates the different diagnostic terminologies reported that are considered lexically discordant. Of the example cases, only case four exhibits exact agreement of the diagnoses issued in each laboratory report and therefore is not considered discordant. The remaining cases indicate the variability typically seen with diagnostic text, which would be flagged lexically as discordant by some clinical decision support tools.

Figures 2, 3, and 4 display the results and comparative analyses performed for each laboratory pair in the first phase of research. Each box indicates the number of cases and percentage of cases with the criterion indicated. Figure 2 illustrates that for the 140 cases with the hematopathology and flow cytometry diagnoses, 94% displayed lexical discordance. Similarly in Figure 3 with 143 cases, 95% exhibited lexical disagreement. Likewise, Figure 4's 92 cases show 92% disagree lexically. As illustrated in Table 1, the variety of diagnoses contributed to the elevated number of discordant cases measured with this methodology. Consequently due to this large variety, Kappa could not be calculated for lexical discordance.

Table 1: Examples of Lexical and Semantic Diagnostic Comparisons

Case	Hematopathology Diagnosis	Flow Cytometry Diagnosis	Lexical Determination	Semantic Determination
1	Common Precursor B-cell acute lymphoblastic leukemia	Precursor B cell lymphoblastic leukemia	Disagree	Agree
2	Acute myelomonoblastic leukemia	Acute myeloid leukemia	Disagree	Unclassifiable
3	Acute myeloblastic leukemia without maturation	Acute myeloblastic leukemia with maturation	Disagree	Disagree
4	Precursor B cell lymphoblastic leukemia	Precursor B cell lymphoblastic leukemia	Agree	Agree
5	Acute myelomonocytic leukemia	Acute myelomonocytic leukemia or acute monocytic leukemia	Disagree	Unclassifiable
6	Acute leukemia, favor biphenoypic, precursor T-lymphoblast and myeloblast	Acute leukemia, favor biphenoypic, precursor T-lymphoblast and myeloblast or Precursor T-lymphoblastic leukemia	Disagree	Unclassifiable
7	Acute myelomonocytic leukemia	Immunophenotypic analysis shows M4*	Disagree	Agree

*M4 is the old French-American-British (FAB) classification for acute myelomonocytic leukemia (30)

Figure 2: Phase I: Determination of Diagnostic Discordance
(Hematopathology versus Flow Cytometry, 140 total cases)

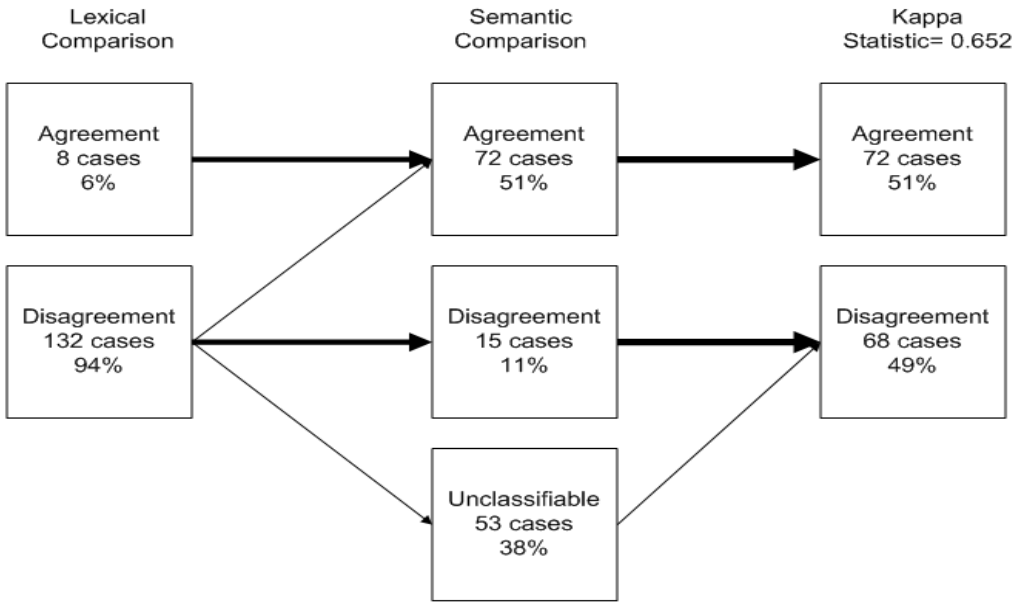


Figure 3: Phase I: Determination of Diagnostic Discordance
(Hematopathology versus Cytogenetics, 143 total cases)

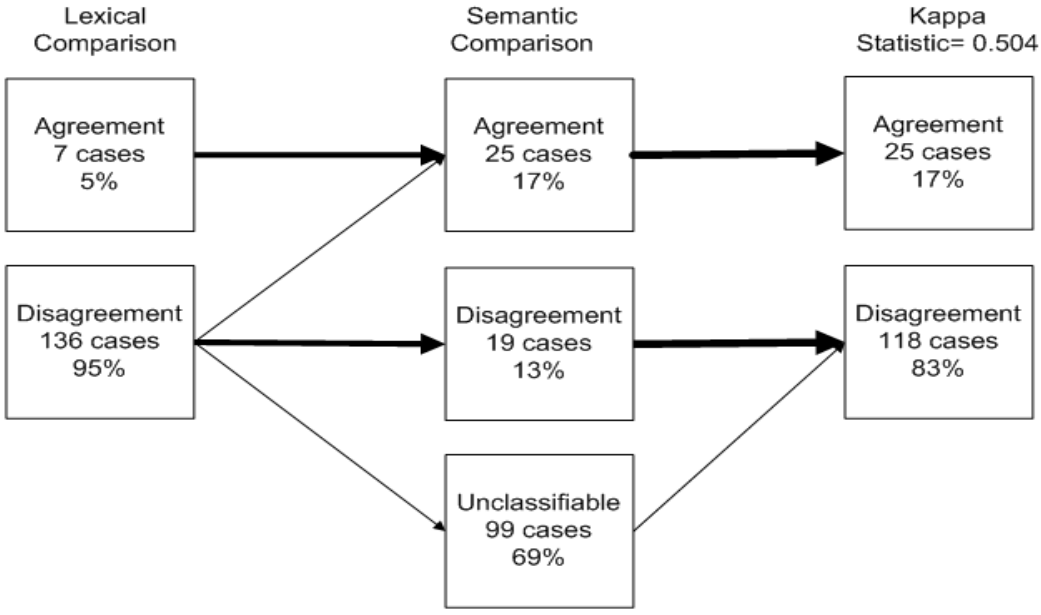


Figure 4: Phase I: Determination of Diagnostic Discordance (Flow Cytometry versus Cytogenetics, 92 total cases)

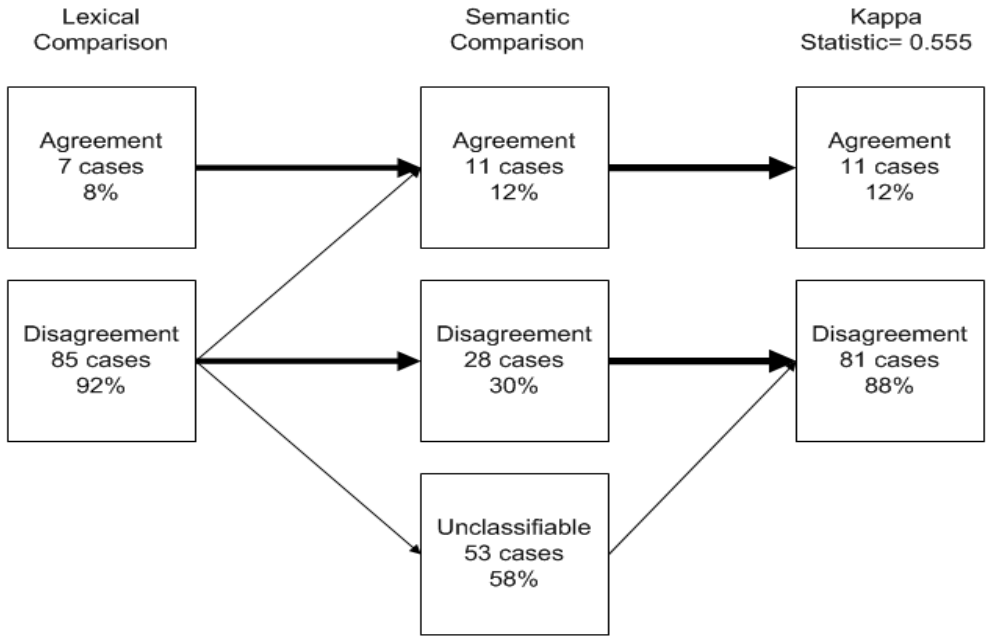


Table 2: Lexical and Semantic Discordance Comparison of Paired Laboratory Reports				
Case Comparison Totals Types	Lexical	Semantic	Lexical % of Total	Semantic % of Total
100% Agreement: 2 labs, 1 unique diagnosis	2	40	1.0%	20.6%
100% Agreement: 3 labs, 1 unique diagnosis	3	11	1.6%	5.7%
Partial Agreement: 3 labs, 2 unique diagnoses	17	36	8.8%	18.6%
No Agreement: 2 labs, 2 unique diagnoses	93	65	47.9%	33.5%
No Agreement: 3 labs, 3 unique diagnoses	71	42	36.6%	21.7%
Total Cases with 1 performing lab (excluded)	8	0	4.1%	0.0%
Total Case Comparisons, 2 performing labs	103	102	53.1%	52.6%
Total Case Comparisons, 3 performing labs	91	92	46.9%	47.4%
Total Case Comparisons, 2 or more performing labs	194	194	100.0%	100.0%

Lexical agreement was also determined by comparing all diagnoses for a case side by side as shown in the lexical columns of Table 2. Cases with only a single laboratory diagnosis were excluded from this analysis. However, cases with two of the three or all three laboratories issuing reports were analyzed. It was found that in 2.6% of case comparisons, the laboratory reports featured an exact match in lexical diagnoses and all

the laboratories agreed upon a single diagnosis. Conversely, in 84.5% of comparisons, each laboratory reported a different diagnostic text. That is, all three laboratories were in complete disagreement with each other and each issued a unique diagnosis resulting in three diagnoses for the specimen. In short, complete lexical diagnostic discordance was more common than complete lexical agreement.

Phase I: Determination of Semantic Diagnostic Discordance

Semantic diagnostic discordance among bone marrow specimens was also detected and measured in this part of the first phase of research. The same cases from the lexical comparisons were semantically compared. The same assumptions described above hold for the determination of semantic discordance.

Table 1 indicates real case data examples illustrating determination of semantic diagnostic agreement. Cases one, four, and seven, all show semantic diagnostic agreement. Case three is an example of semantic disagreement or discordance. Cases two, five and six are cases which are considered unclassifiable by our methods. Case two exemplifies this distinction. If additional information was provided for the acute myeloid leukemia diagnosis, it is unclear if this diagnosis would be reassessed as acute myelomonoblastic leukemia and thus be in agreement with the other laboratory reported diagnosis. Without additional information, these diagnoses would be considered semantically discordant. There is insufficient information reported to determine semantic agreement. Cases five and six demonstrate one laboratory reporting two distinct diagnoses. It is unclear from the report as to which diagnosis best describes the case. Again, with sufficient information, one may be able to determine that one diagnosis better describes the case and is either in agreement with the other laboratory diagnosis or discordant. Current decision support tools lack the capability to make an adequate assessment of semantic agreement.

Figures 2, 3 and 4 also illustrate semantic agreement and disagreement among the laboratory comparisons. Semantic agreement between the hematopathology and flow cytometry laboratories in Figure 2 was 51%, but 38% of cases were unclassifiable. Many of the unclassifiable cases reported a more general WHO classification as illustrated with

case two in Table 1. Similar results are shown in Figure 3 with 17% of cases in semantic agreement and 69% of cases unclassifiable. Likewise, Figure 4 shows 12% of cases in semantic agreement and 58% deemed as unclassifiable.

Semantic comparison of all diagnoses reported on each case was performed similar to lexical diagnostic comparison as shown in Table 2. The data indicate that all the laboratories agreed on a single semantic diagnostic concept in 26% of cases. Compared to the results for lexical diagnostic discordance, there is a 30% drop in discordance by semantic comparison. Semantic agreement also showed a 23% increase over lexical diagnostic agreement. Overall, the semantic methodology resulted in better diagnostic agreement and less diagnostic discordance than the lexical methodology.

Phase I: Calculation of Kappa

Cohen's Kappa statistic was calculated only for the semantic diagnostic comparison for each laboratory pair in measuring diagnostic agreement statistically. Kappa is an appropriate statistical measure since it measures agreement of independently categorized nominal data. Corresponding Kappa reference ranges and interpretations are depicted in Table 3.

Table 3 Kappa Reference Ranges and Interpretations (70)

Kappa value	Interpretation
< 0	No Agreement
0.0-0.19	Poor agreement
0.20-0.39	Fair agreement
0.40-0.59	Moderate agreement
0.60-0.79	Substantial agreement
0.80-1.00	Almost perfect agreement

Kappa results for semantic diagnostic agreement measurement for each laboratory comparison are in Figures 2, 3, and 4. Kappa for the hematopathology and flow cytometry comparison showed substantial agreement with a value of 0.652 as shown in Figure 2. Figure 3 shows a Kappa of 0.504 indicating moderate agreement over chance

alone for the hematopathology and cytogenetics diagnoses. Likewise, a Kappa of 0.555 indicates moderate agreement in Figure 4 in comparing the flow cytometry and cytogenetics laboratories. In Figures 2, 3, and 4, manual calculation of semantic agreement resulted in the same number and percent of cases determined to be in semantic agreement automatically by Kappa. However, Kappa is only able to determine agreement or disagreement and not further classify the cases into the categories we used. Therefore, disagreements and unclassifiable cases are designated as disagree via Kappa methodology.

Phase II: Categorization of Discordant Cases

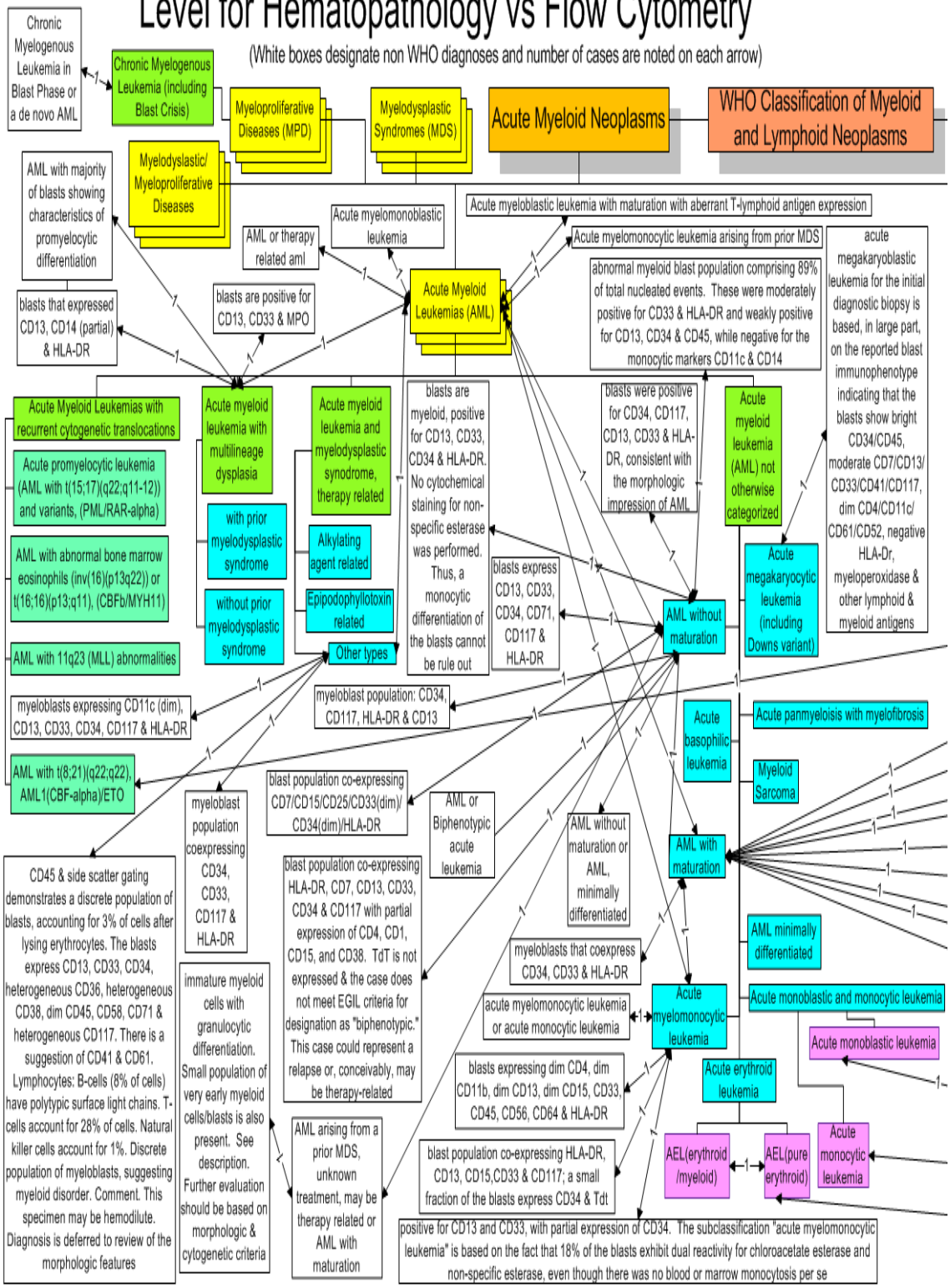
Discordant cases from phase one were utilized in this second phase of the research. Only semantically discordant cases with the same first phase assumptions were analyzed. These second phase results provide information concerning the classification of factors contributing to diagnostic discordance. The analyses distinguish anticipated issues resulting from laboratory limitations from those which are unanticipated, such as errors impacting patient care. Key factors contributing to diagnostic discordance and where interventions such as clinical decision support tools might best aid the clinician are also determined.

Phase II: Categorization By Level

"Categorization by Level," classifies each pair of discordant cases according to diagnostic relationships utilizing the 2001 WHO diagnostic categorizations in Figure 1. Results for each comparison pair are diagrammed in Figures 5, 6, and 7. Arrows connect each of the two diagnoses reported for a case. Case counts for each relationship are indicated by the number on the arrow. For cases not utilizing a WHO diagnostic category, diagnostic text was listed in its own box and linked to the second diagnosis for that case.

Figure 5: Discordant Case Categorization by WHO Classification Level for Hematopathology vs Flow Cytometry

(White boxes designate non WHO diagnoses and number of cases are noted on each arrow)



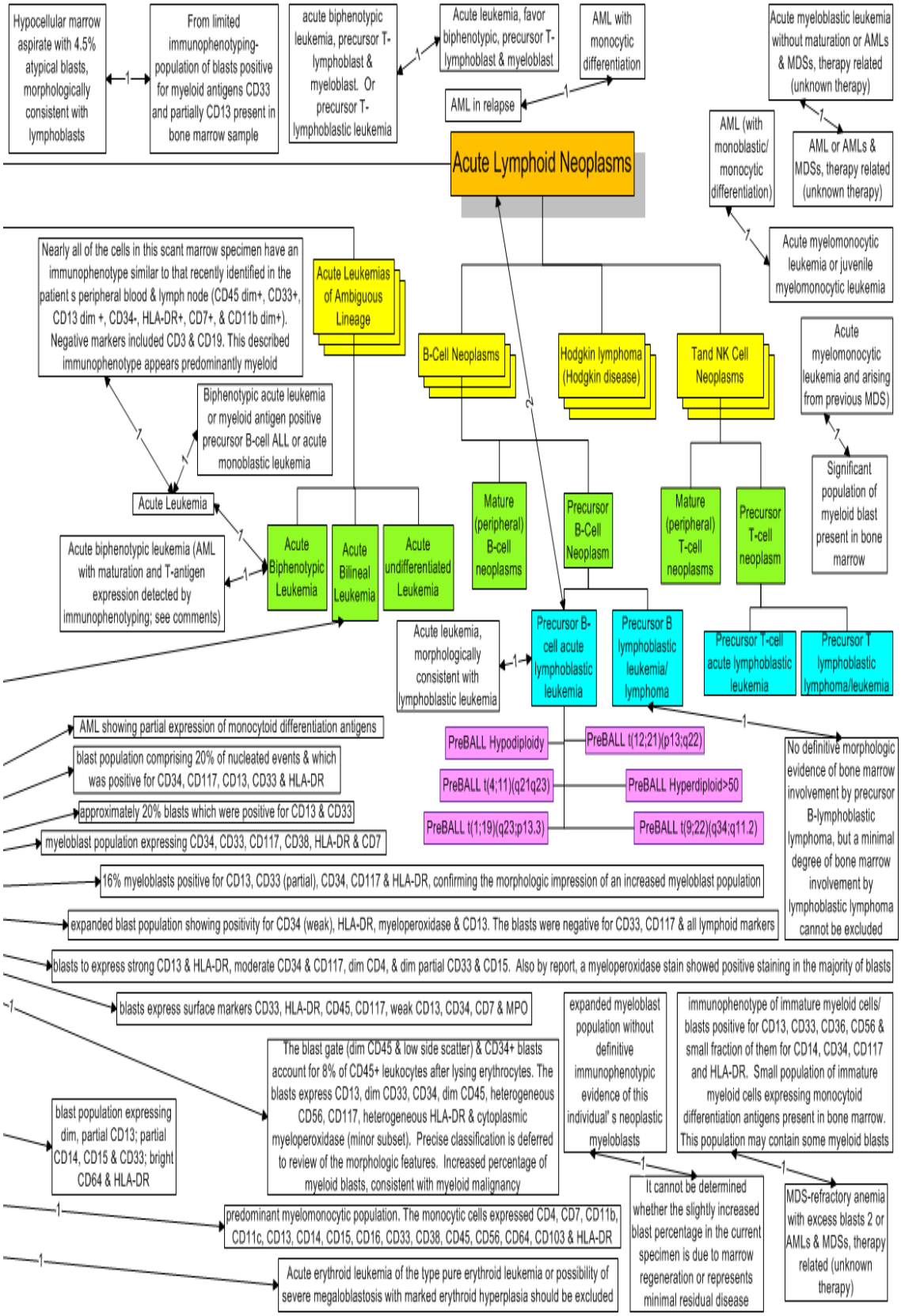
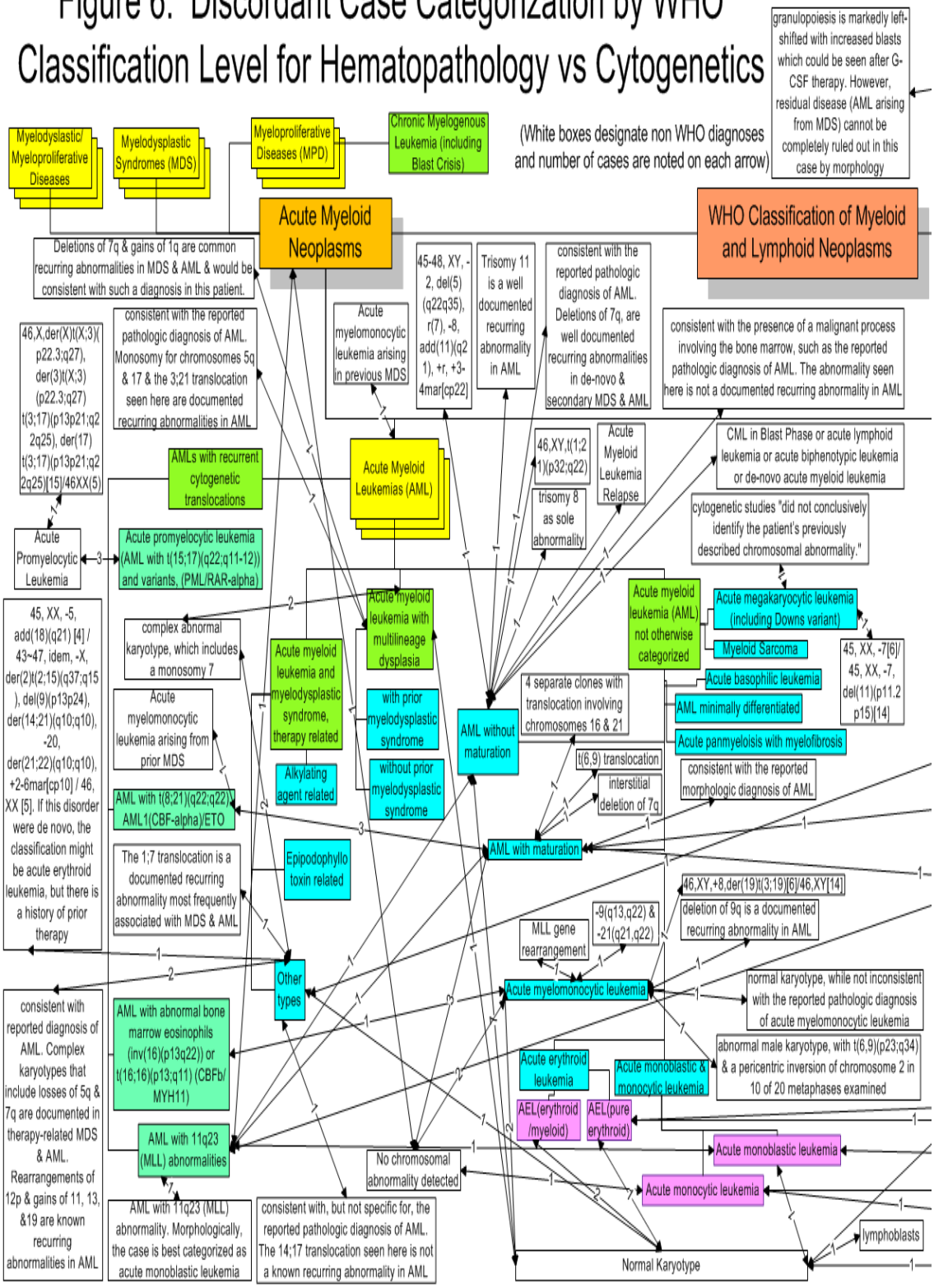

Acute Lymphoid Neoplasms
Hypocellular marrow aspirate with 4.5% atypical blasts, morphologically consistent with lymphoblasts
From limited immunophenotyping-population of blasts positive for myeloid antigens CD33 and partially CD13 present in bone marrow sample
acute biphenotypic leukemia, precursor T-lymphoblast & myeloblast. Or precursor T-lymphoblastic leukemia
Acute leukemia, favor biphenotypic, precursor T-lymphoblast & myeloblast
AML with monocytic differentiation
AML in relapse
Acute myeloblastic leukemia without maturation or AMLs & MDSs, therapy related (unknown therapy)
AML or AMLs & MDSs, therapy related (unknown therapy)
AML (with monoblastic/ monocytic differentiation)
Acute myelomonocytic leukemia or juvenile myelomonocytic leukemia
Nearly all of the cells in this scant marrow specimen have an immunophenotype similar to that recently identified in the patient s peripheral blood & lymph node (CD45 dim+, CD33+, CD13 dim +, CD34-, HLA-DR+, CD7+, & CD11b dim+). Negative markers included CD3 & CD19. This described immunophenotype appears predominantly myeloid
Acute Leukemias of Ambiguous Lineage
B-Cell Neoplasms
Hodgkin lymphoma (Hodgkin disease)
Tand NK Cell Neoplasms
Acute myelomonocytic leukemia and arising from previous MDS)
Biphenotypic acute leukemia or myeloid antigen positive precursor B-cell ALL or acute monoblastic leukemia
Significant population of myeloid blast present in bone marrow
Acute Leukemia
Acute biphenotypic leukemia (AML with maturation and T-antigen expression detected by immunophenotyping; see comments)
Acute Biphenotypic Leukemia
Acute Bilineal Leukemia
Acute undifferentiated Leukemia
Mature (peripheral) B-cell neoplasms
Precursor B-Cell Neoplasm
Mature (peripheral) T-cell neoplasms
Precursor T-cell neoplasm
Acute leukemia, morphologically consistent with lymphoblastic leukemia
Precursor B-cell acute lymphoblastic leukemia
Precursor B lymphoblastic leukemia/ lymphoma
Precursor T-cell acute lymphoblastic leukemia
Precursor T lymphoblastic lymphoma/leukemia
AML showing partial expression of monocytoid differentiation antigens
blast population comprising 20% of nucleated events & which was positive for CD34, CD117, CD13, CD33 & HLA-DR
approximately 20% blasts which were positive for CD13 & CD33
myeloblast population expressing CD34, CD33, CD117, CD38, HLA-DR & CD7
PreBALL Hypodiploidy
PreBALL t(12;21)(p13;q22)
PreBALL t(4;11)(q21q23)
PreBALL Hyperdiploid>50
PreBALL t(1;19)(q23;p13.3)
PreBALL t(9;22)(q34;q11.2)
No definitive morphologic evidence of bone marrow involvement by precursor B-lymphoblastic lymphoma, but a minimal degree of bone marrow involvement by lymphoblastic lymphoma cannot be excluded
16% myeloblasts positive for CD13, CD33 (partial), CD34, CD117 & HLA-DR, confirming the morphologic impression of an increased myeloblast population
expanded blast population showing positivity for CD34 (weak), HLA-DR, myeloperoxidase & CD13. The blasts were negative for CD33, CD117 & all lymphoid markers
blasts to express strong CD13 & HLA-DR, moderate CD34 & CD117, dim CD4, & dim partial CD33 & CD15. Also by report, a myeloperoxidase stain showed positive staining in the majority of blasts
blasts express surface markers CD33, HLA-DR, CD45, CD117, weak CD13, CD34, CD7 & MPO
The blast gate (dim CD45 & low side scatter) & CD34+ blasts account for 8% of CD45+ leukocytes after lysing erythrocytes. The blasts express CD13, dim CD33, CD34, dim CD45, heterogeneous CD56, CD117, heterogeneous HLA-DR & cytoplasmic myeloperoxidase (minor subset). Precise classification is deferred to review of the morphologic features. Increased percentage of myeloid blasts, consistent with myeloid malignancy
expanded myeloblast population without definitive immunophenotypic evidence of this individual' s neoplastic myeloblasts
immunophenotype of immature myeloid cells/ blasts positive for CD13, CD33, CD36, CD56 & small fraction of them for CD14, CD34, CD117 and HLA-DR. Small population of immature myeloid cells expressing monocytoid differentiation antigens present in bone marrow. This population may contain some myeloid blasts
blast population expressing dim, partial CD13; partial CD14, CD15 & CD33; bright CD64 & HLA-DR
predominant myelomonocytic population. The monocytic cells expressed CD4, CD7, CD11b, CD11c, CD13, CD14, CD15, CD16, CD33, CD38, CD45, CD56, CD64, CD103 & HLA-DR
Acute erythroid leukemia of the type pure erythroid leukemia or possibility of severe megaloblastosis with marked erythroid hyperplasia should be excluded
It cannot be determined whether the slightly increased blast percentage in the current specimen is due to marrow regeneration or represents minimal residual disease
MDS-refractory anemia with excess blasts 2 or AMLs & MDSs, therapy related (unknown therapy)

Figure 6: Discordant Case Categorization by WHO Classification Level for Hematopathology vs Cytogenetics



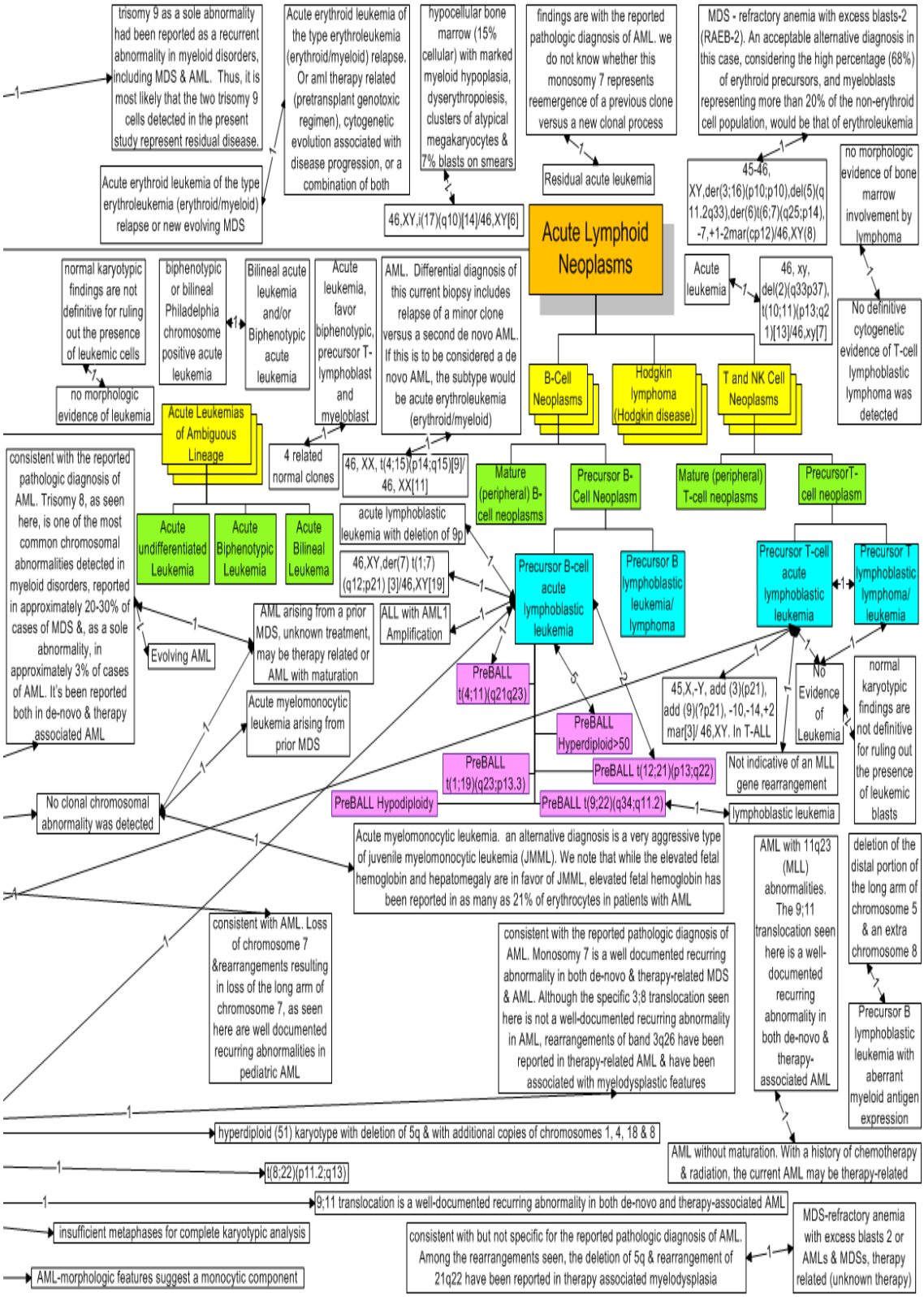
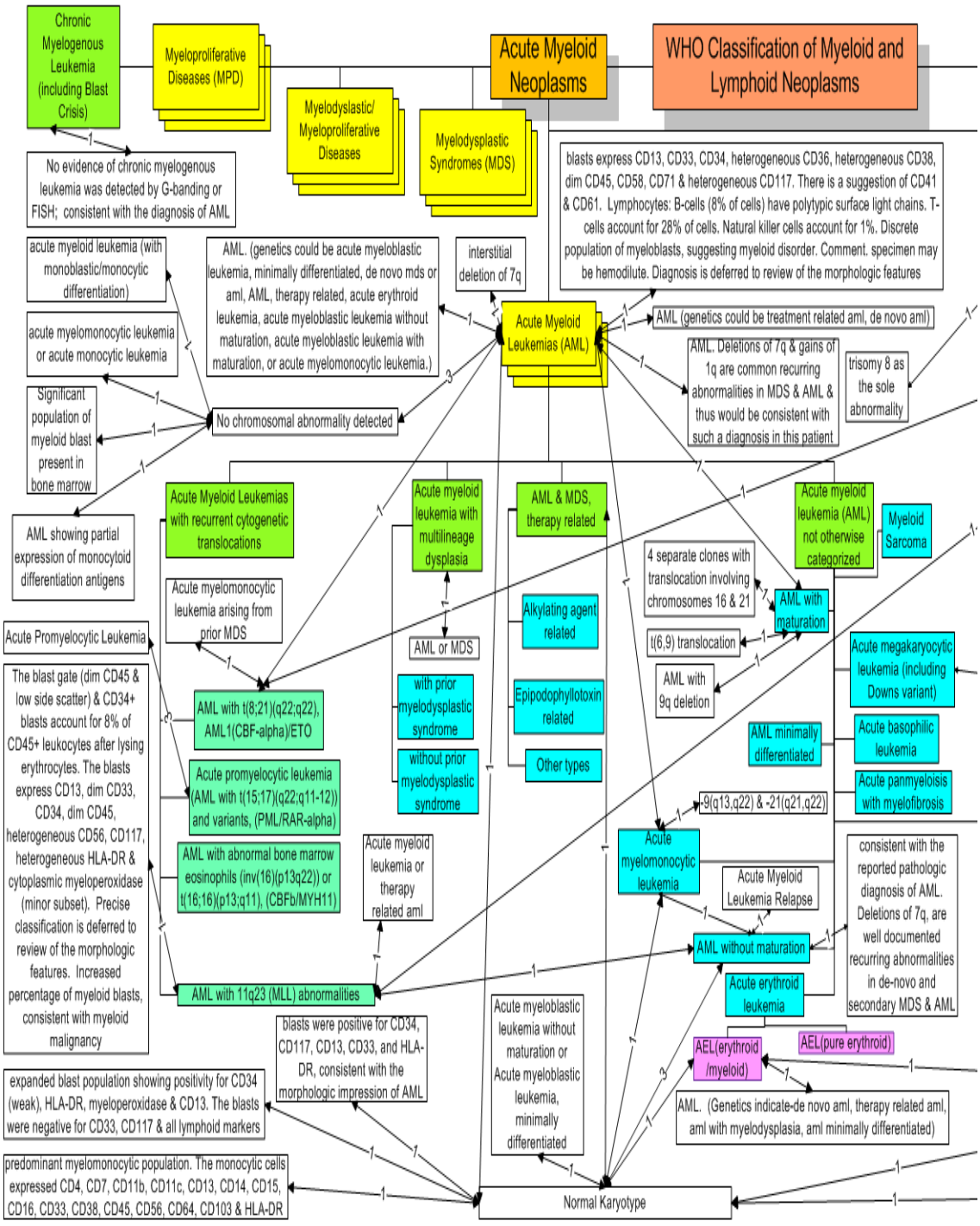
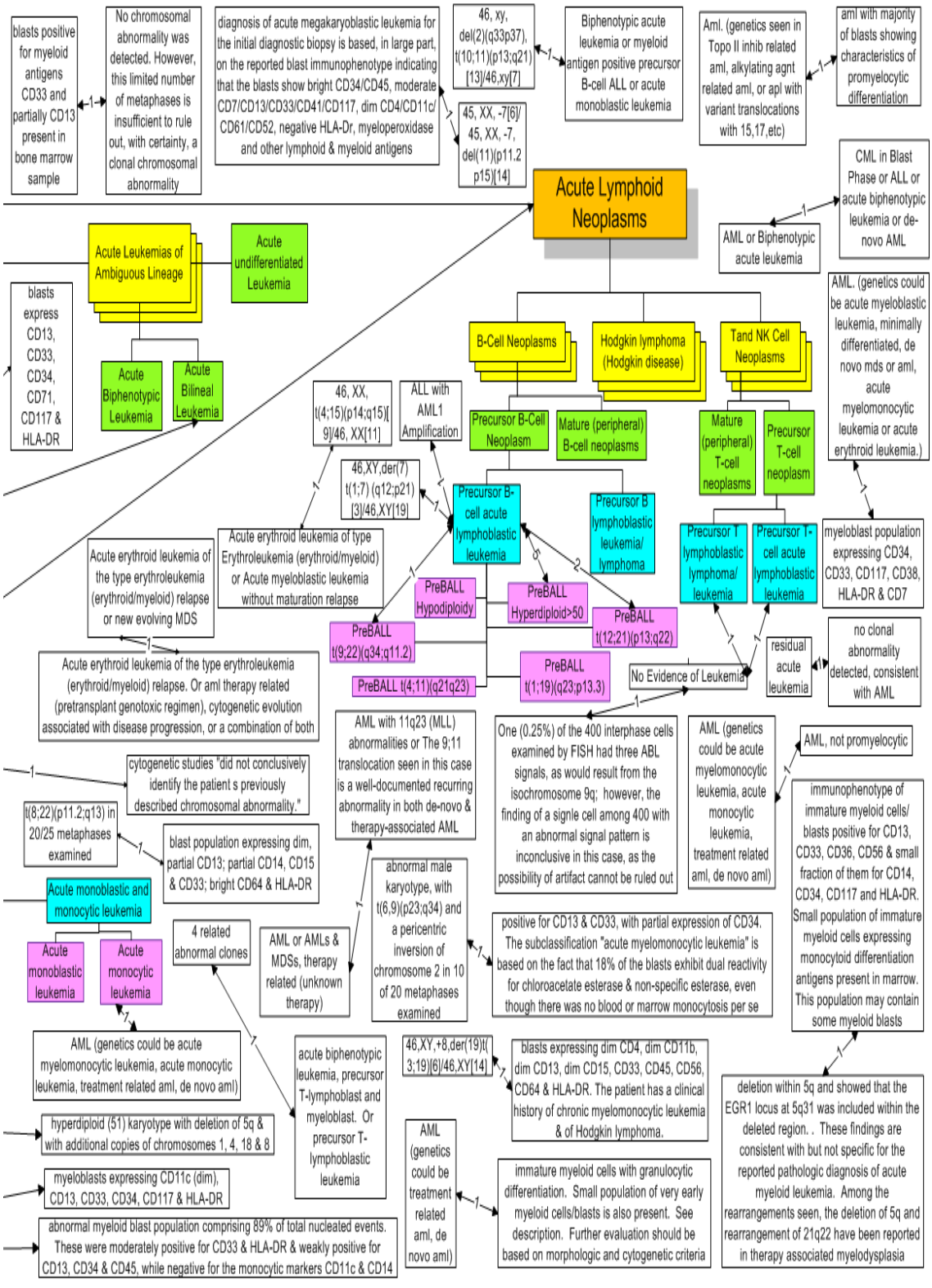

Acute Lymphoid Neoplasms
trisomy 9 as a sole abnormality had been reported as a recurrent abnormality in myeloid disorders, including MDS & AML. Thus, it is most likely that the two trisomy 9 cells detected in the present study represent residual disease.
Acute erythroid leukemia of the type erythroleukemia (erythroid/myeloid) relapse or new evolving MDS
Acute erythroid leukemia of the type erythroleukemia (erythroid/myeloid) relapse. Or aml therapy related (pretransplant genotoxic regimen), cytogenetic evolution associated with disease progression, or a combination of both
hypocellular bone marrow (15% cellular) with marked myeloid hypoplasia, dyserythropoiesis, clusters of atypical megakaryocytes & 7% blasts on smears
46,XY,i(17)(q10)[14]/46,XY[6]
findings are with the reported pathologic diagnosis of AML. we do not know whether this monosomy 7 represents reemergence of a previous clone versus a new clonal process
Residual acute leukemia
MDS - refractory anemia with excess blasts-2 (RAEB-2). An acceptable alternative diagnosis in this case, considering the high percentage (68%) of erythroid precursors, and myeloblasts representing more than 20% of the non-erythroid cell population, would be that of erythroleukemia
45-46, XY,der(3;16)(p10;p10),del(5)(q11.2q33),der(6)t(6;7)(q25;p14),-7,+1-2mar(cp12)/46,XY(8)
no morphologic evidence of bone marrow involvement by lymphoma
normal karyotypic findings are not definitive for ruling out the presence of leukemic cells
no morphologic evidence of leukemia
biphenotypic or bilineal Philadelphia chromosome positive acute leukemia
Bilineal acute leukemia and/or Biphenotypic acute leukemia
Acute leukemia, favor biphenotypic, precursor T-lymphoblast and myeloblast
AML. Differential diagnosis of this current biopsy includes relapse of a minor clone versus a second de novo AML. If this is to be considered a de novo AML, the subtype would be acute erythroleukemia (erythroid/myeloid)
Acute leukemia
46, xy, del(2)(q33p37), t(10;11)(p13;q21)[13]/46,xy[7]
No definitive cytogenetic evidence of T-cell lymphoblastic lymphoma was detected
B-Cell Neoplasms
Hodgkin lymphoma (Hodgkin disease)
T and NK Cell Neoplasms
Acute Leukemias of Ambiguous Lineage
4 related normal clones
46, XX, t(4;15)(p14;q15)[9]/46, XX[11]
Mature (peripheral) B-cell neoplasms
Precursor B-Cell Neoplasm
Mature (peripheral) T-cell neoplasms
PrecursorT-cell neoplasm
consistent with the reported pathologic diagnosis of AML. Trisomy 8, as seen here, is one of the most common chromosomal abnormalities detected in myeloid disorders, reported in approximately 20-30% of cases of MDS &, as a sole abnormality, in approximately 3% of cases of AML. It's been reported both in de-novo & therapy associated AML
Acute undifferentiated Leukemia
Acute Biphenotypic Leukemia
Acute Bilineal Leukema
acute lymphoblastic leukemia with deletion of 9p
46,XY,der(7) t(1;7)(q12;p21) [3]/46,XY[19]
ALL with AML1 Amplification
Precursor B-cell acute lymphoblastic leukemia
Precursor B lymphoblastic leukemia/ lymphoma
Precursor T-cell acute lymphoblastic leukemia
Precursor T lymphoblastic lymphoma/ leukemia
Evolving AML
AML arising from a prior MDS, unknown treatment, may be therapy related or AML with maturation
PreBALL t(4;11)(q21q23)
PreBALL Hyperdiploid>50
PreBALL t(12;21)(p13;q22)
PreBALL t(1;19)(q23;p13.3)
PreBALL Hypodiploidy
PreBALL t(9;22)(q34;q11.2)
45,X,-Y, add (3)(p21), add (9)(?p21), -10,-14,+2 mar[3]/ 46,XY. In T-ALL
No Evidence of Leukemia
normal karyotypic findings are not definitive for ruling out the presence of leukemic blasts
Not indicative of an MLL gene rearrangement
lymphoblastic leukemia
Acute myelomonocytic leukemia arising from prior MDS
No clonal chromosomal abnormality was detected
Acute myelomonocytic leukemia. an alternative diagnosis is a very aggressive type of juvenile myelomonocytic leukemia (JMML). We note that while the elevated fetal hemoglobin and hepatomegaly are in favor of JMML, elevated fetal hemoglobin has been reported in as many as 21% of erythrocytes in patients with AML
AML with 11q23 (MLL) abnormalities. The 9;11 translocation seen here is a well-documented recurring abnormality in both de-novo & therapy-associated AML
deletion of the distal portion of the long arm of chromosome 5 & an extra chromosome 8
Precursor B lymphoblastic leukemia with aberrant myeloid antigen expression
consistent with AML. Loss of chromosome 7 &rearrangements resulting in loss of the long arm of chromosome 7, as seen here are well documented recurring abnormalities in pediatric AML
consistent with the reported pathologic diagnosis of AML. Monosomy 7 is a well documented recurring abnormality in both de-novo & therapy-related MDS & AML. Although the specific 3;8 translocation seen here is not a well-documented recurring abnormality in AML, rearrangements of band 3q26 have been reported in therapy-related AML & have been associated with myelodysplastic features
hyperdiploid (51) karyotype with deletion of 5q & with additional copies of chromosomes 1, 4, 18 & 8
AML without maturation. With a history of chemotherapy & radiation, the current AML may be therapy-related
t(8;22)(p11.2;q13)
9;11 translocation is a well-documented recurring abnormality in both de-novo and therapy-associated AML
insufficient metaphases for complete karyotypic analysis
consistent with but not specific for the reported pathologic diagnosis of AML. Among the rearrangements seen, the deletion of 5q & rearrangement of 21q22 have been reported in therapy associated myelodysplasia
MDS-refractory anemia with excess blasts 2 or AMLs & MDSs, therapy related (unknown therapy)
AML-morphologic features suggest a monocytic component

Figure 7: Discordant Case Categorization by WHO Classification Level for Flow Cytometry vs Cytogenetics

(White boxes designate non WHO diagnoses and number of cases are noted on each arrow)




Acute Lymphoid Neoplasms
blasts positive for myeloid antigens CD33 and partially CD13 present in bone marrow sample
No chromosomal abnormality was detected. However, this limited number of metaphases is insufficient to rule out, with certainty, a clonal chromosomal abnormality
diagnosis of acute megakaryoblastic leukemia for the initial diagnostic biopsy is based, in large part, on the reported blast immunophenotype indicating that the blasts show bright CD34/CD45, moderate CD7/CD13/CD33/CD41/CD117, dim CD4/CD11c/CD61/CD52, negative HLA-Dr, myeloperoxidase and other lymphoid & myeloid antigens
46, xy, del(2)(q33p37), t(10;11)(p13;q21)[13]/46,xy[7]
45, XX, -7[6]/45, XX, -7, del(11)(p11.2p15)[14]
Biphenotypic acute leukemia or myeloid antigen positive precursor B-cell ALL or acute monoblastic leukemia
Aml. (genetics seen in Topo II inhib related aml, alkylating agnt related aml, or apl with variant translocations with 15,17,etc)
aml with majority of blasts showing characteristics of promyelocytic differentiation
CML in Blast Phase or ALL or acute biphenotypic leukemia or de-novo AML
AML or Biphenotypic acute leukemia
Acute Leukemias of Ambiguous Lineage
Acute undifferentiated Leukemia
blasts express CD13, CD33, CD34, CD71, CD117 & HLA-DR
Acute Biphenotypic Leukemia
Acute Bilineal Leukemia
B-Cell Neoplasms
Hodgkin lymphoma (Hodgkin disease)
Tand NK Cell Neoplasms
AML. (genetics could be acute myeloblastic leukemia, minimally differentiated, de novo mds or aml, acute myelomonocytic leukemia or acute erythroid leukemia.)
46, XX, t(4;15)(p14;q15)[9]/46, XX[11]
ALL with AML1 Amplification
Precursor B-Cell Neoplasm
Mature (peripheral) B-cell neoplasms
Mature (peripheral) T-cell neoplasms
Precursor T-cell neoplasm
46,XY,der(7)t(1;7)(q12;p21)[3]/46,XY[19]
Precursor B-cell acute lymphoblastic leukemia
Precursor B lymphoblastic leukemia/lymphoma
Precursor T lymphoblastic lymphoma/leukemia
Precursor T-cell acute lymphoblastic leukemia
myeloblast population expressing CD34, CD33, CD117, CD38, HLA-DR & CD7
Acute erythroid leukemia of the type erythroleukemia (erythroid/myeloid) relapse or new evolving MDS
Acute erythroid leukemia of type Erythroleukemia (erythroid/myeloid) or Acute myeloblastic leukemia without maturation relapse
PreBALL Hypodiploidy
PreBALL Hyperdiploid>50
PreBALL t(12;21)(p13;q22)
PreBALL t(9;22)(q34;q11.2)
PreBALL t(4;11)(q21q23)
PreBALL t(1;19)(q23;p13.3)
Acute erythroid leukemia of the type erythroleukemia (erythroid/myeloid) relapse. Or aml therapy related (pretransplant genotoxic regimen), cytogenetic evolution associated with disease progression, or a combination of both
No Evidence of Leukemia
residual acute leukemia
no clonal abnormality detected, consistent with AML
AML with 11q23 (MLL) abnormalities or The 9;11 translocation seen in this case is a well-documented recurring abnormality in both de-novo & therapy-associated AML
One (0.25%) of the 400 interphase cells examined by FISH had three ABL signals, as would result from the isochromosome 9q; however, the finding of a signle cell among 400 with an abnormal signal pattern is inconclusive in this case, as the possibility of artifact cannot be ruled out
AML (genetics could be acute myelomonocytic leukemia, acute monocytic leukemia, treatment related aml, de novo aml)
AML, not promyelocytic
cytogenetic studies "did not conclusively identify the patient s previously described chromosomal abnormality."
t(8;22)(p11.2;q13) in 20/25 metaphases examined
blast population expressing dim, partial CD13; partial CD14, CD15 & CD33; bright CD64 & HLA-DR
immunophenotype of immature myeloid cells/blasts positive for CD13, CD33, CD36, CD56 & small fraction of them for CD14, CD34, CD117 and HLA-DR. Small population of immature myeloid cells expressing monocytoid differentiation antigens present in marrow. This population may contain some myeloid blasts
Acute monoblastic and monocytic leukemia
Acute monoblastic leukemia
Acute monocytic leukemia
4 related abnormal clones
AML or AMLs & MDSs, therapy related (unknown therapy)
abnormal male karyotype, with t(6,9)(p23;q34) and a pericentric inversion of chromosome 2 in 10 of 20 metaphases examined
positive for CD13 & CD33, with partial expression of CD34. The subclassification "acute myelomonocytic leukemia" is based on the fact that 18% of the blasts exhibit dual reactivity for chloroacetate esterase & non-specific esterase, even though there was no blood or marrow monocytosis per se
AML (genetics could be acute myelomonocytic leukemia, acute monocytic leukemia, treatment related aml, de novo aml)
acute biphenotypic leukemia, precursor T-lymphoblast and myeloblast. Or precursor T-lymphoblastic leukemia
46,XY,+8,der(19)t(3;19)(6)/46,XY[14]
blasts expressing dim CD4, dim CD11b, dim CD13, dim CD15, CD33, CD45, CD56, CD64 & HLA-DR. The patient has a clinical history of chronic myelomonocytic leukemia & of Hodgkin lymphoma.
deletion within 5q and showed that the EGR1 locus at 5q31 was included within the deleted region. . These findings are consistent with but not specific for the reported pathologic diagnosis of acute myeloid leukemia. Among the rearrangements seen, the deletion of 5q and rearrangement of 21q22 have been reported in therapy associated myelodysplasia
hyperdiploid (51) karyotype with deletion of 5q & with additional copies of chromosomes 1, 4, 18 & 8
myeloblasts expressing CD11c (dim), CD13, CD33, CD34, CD117 & HLA-DR
abnormal myeloid blast population comprising 89% of total nucleated events. These were moderately positive for CD33 & HLA-DR & weakly positive for CD13, CD34 & CD45, while negative for the monocytic markers CD11c & CD14
AML (genetics could be treatment related aml, de novo aml)
immature myeloid cells with granulocytic differentiation. Small population of very early myeloid cells/blasts is also present. See description. Further evaluation should be based on morphologic and cytogenetic criteria

In Figure 5, discordant cases between hematopathology and flow cytometry laboratories are diagrammed. The pattern occurring most often with four cases was when one laboratory diagnosis was acute myeloid leukemia, while the other laboratory diagnosis was acute myeloblastic leukemia with maturation. Otherwise, the results indicate a wide variety in diagnoses reported by each laboratory including many unique or non-WHO categories. Only a few cases involved discordances between WHO categorizations.

Figure 6 depicts similar results in comparing the diagnoses between hematopathology and cytogenetics laboratories. The most frequent discordance with five cases occurred between a diagnosis of Precursor B-cell acute lymphoblastic leukemia from the hematopathology laboratory and a diagnosis of Pre B acute lymphoblastic leukemia of the hyperdiploid type > 50, from the cytogenetics laboratory. This relationship also occurred with five cases between the flow cytometry and cytogenetics laboratories in Figure 7. Results are similar to those in Figures 5 and 6.

Next, cases were assessed by what "level" or type of relationship was noted in each pair. Figures 8 and 9 display the data and their relationships. Total percentage of cases for each pair is 100% since each case was classified as a single type of relationship. On average there were eight times as many discordant cases with vertical or parent-child relationships between diagnoses as opposed to horizontal or sibling-to-sibling relationships. In hematopathology and flow cytometry comparison, 28% of discordant cases diagnoses were vertically related from a more general to a more specific relationship. Many cases exhibited an "other relationship," neither a strictly horizontal nor a strictly vertical relationship, among all laboratory comparisons. Even more cases were unable to be classified using the methodology. The average number of cases from all laboratories compared which exhibited an "other relationship," was 47%. In short, for each laboratory pair, the majority of discordant case relationships were designated as unable to be classified, while only a small number of cases were designated as having a vertical or horizontal relationship. A contributing factor to these findings is the variety of non-WHO diagnoses reported.

Figure 8: Discordant Cases Classified by Category Levels for each Laboratory Pair

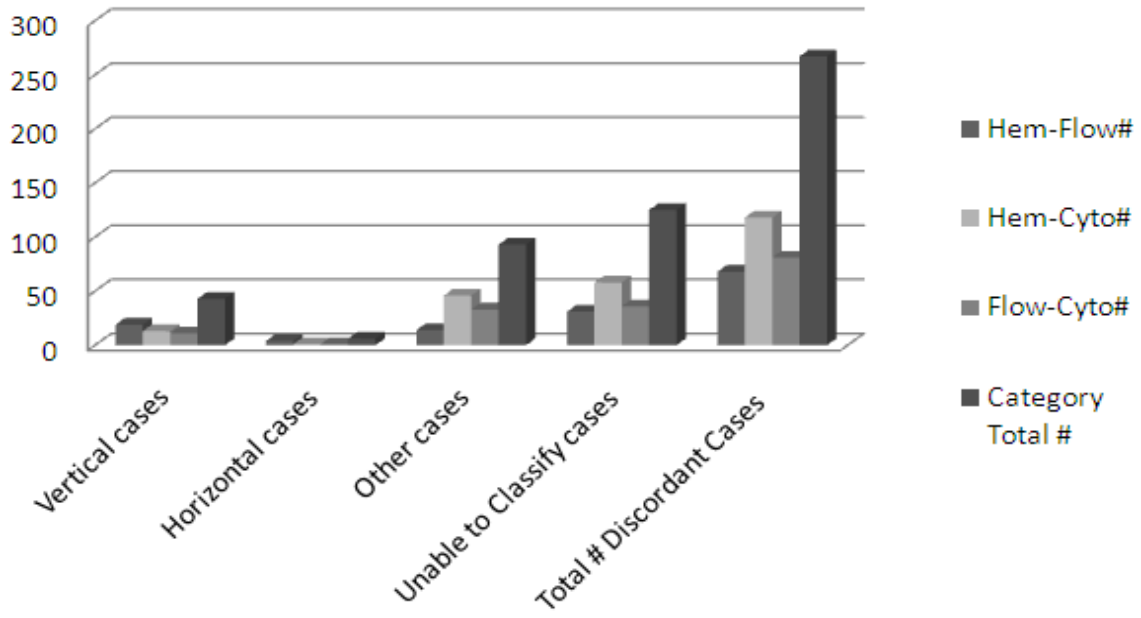
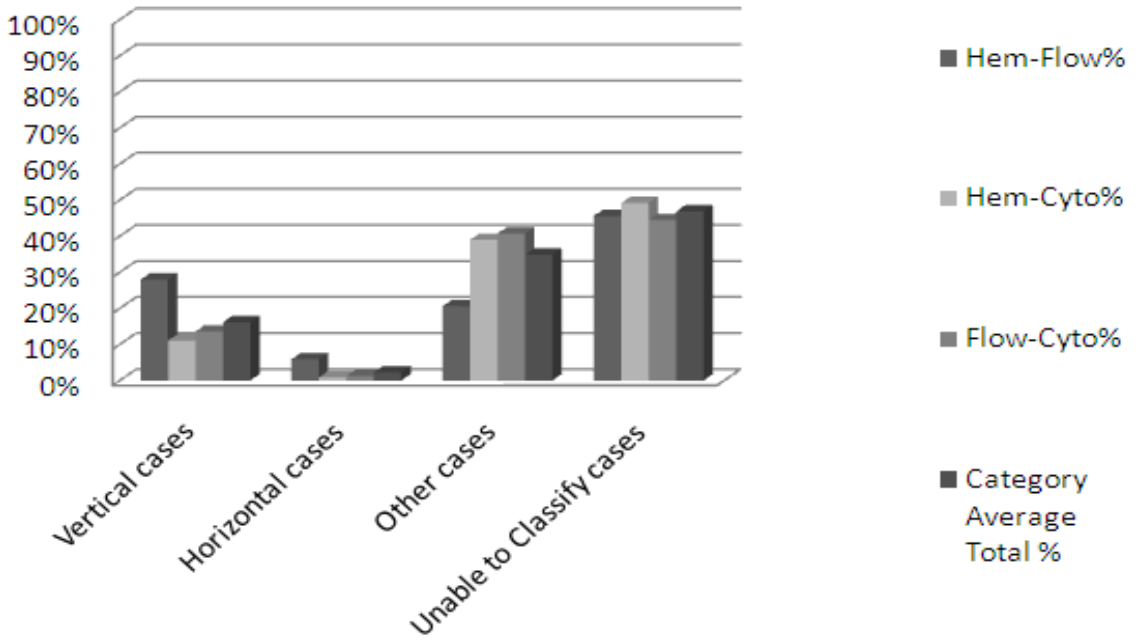


Figure 9: Discordant Case Percentages Classified by Category Levels for each Laboratory Pair



Phase II: Categorization By Limiting Factors

"Categorization By Limiting Factors" methodology classifies discordant cases by limiting factors contributing to discordance. Factors may be laboratory based or due to other sources as in Figures 10 and 11.

Figure 10: Limiting Factor Counts for Discordant Cases

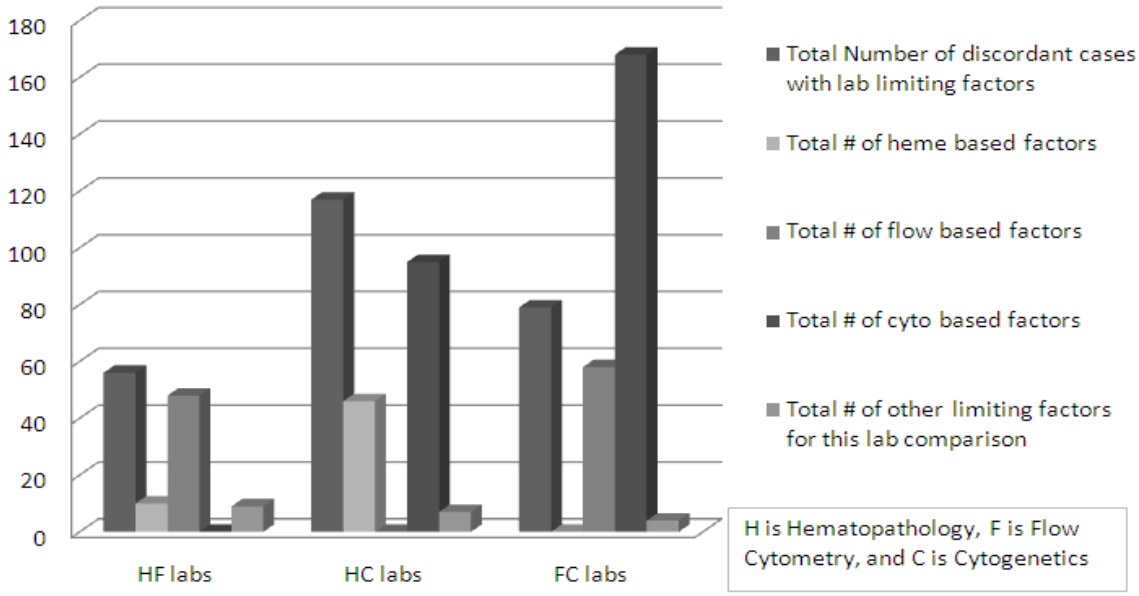
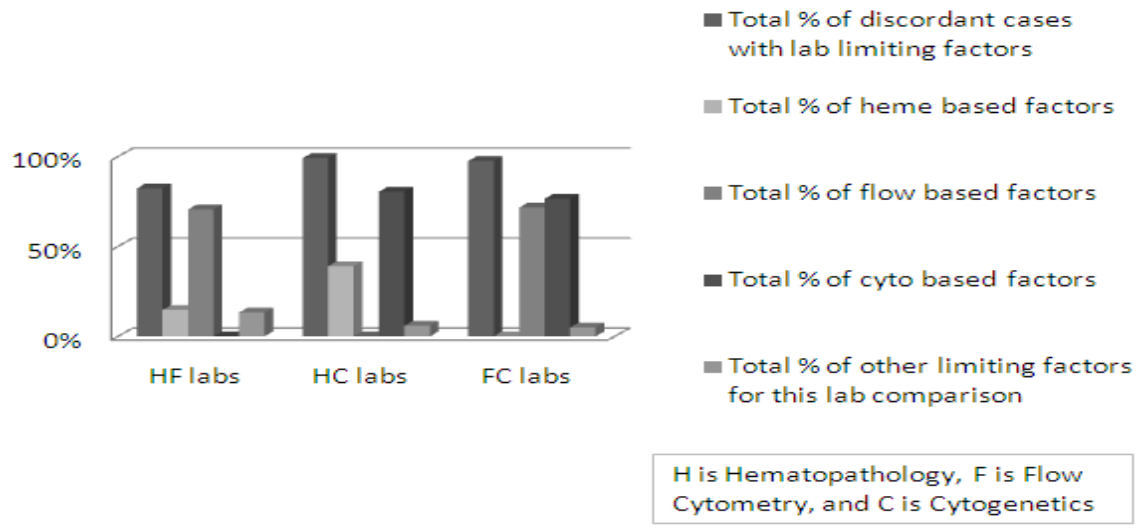


Figure 11: Limiting Factor Percentages for Discordant Cases



Figures 10 and 11 delineate the counts and percentage of limiting factors from each laboratory, other sources, and the total number of factors contributing to discordance. Total percentages may not add up to 100% since each case may have more than one limiting factor. Results indicate most factors are laboratory based. Among laboratories, cytogenetics factors account for over 75% of limiting factors contributing to discordance as shown in Figure 11. Results indicate over 99% of hematopathology and cytogenetics discordant cases feature a laboratory based limiting factor. These findings suggest factors within the laboratory environment contribute to nearly all of the diagnostic discordances noted in this research.

Discordances were also classified by specific limiting factors. The top five factors for each laboratory pair, which laboratory diagnosis they impact, and number of cases exhibiting these factors are listed in Table 4. For hematopathology and flow cytometry comparison, all discordances were attributed to flow cytometry laboratory factors. The top limiting factor contributing to discordances occurs when hematopathology issues a specific diagnosis and flow cytometry reports a more general diagnosis.

Between hematopathology and cytogenetics laboratories, most limiting factors are attributed to the cytogenetics laboratory. Cytogenetics factors are commonly attributed to a normal karyotype, which may or may not be consistent with a reported leukemia by the second laboratory. Cytogenetics may also report an abnormal karyotype consistent with an acute leukemia, but the report does not indicate a leukemia subtype. The top hematopathology based factor is limitation of hematopathology microscopy in detecting cytogenetics translocations.

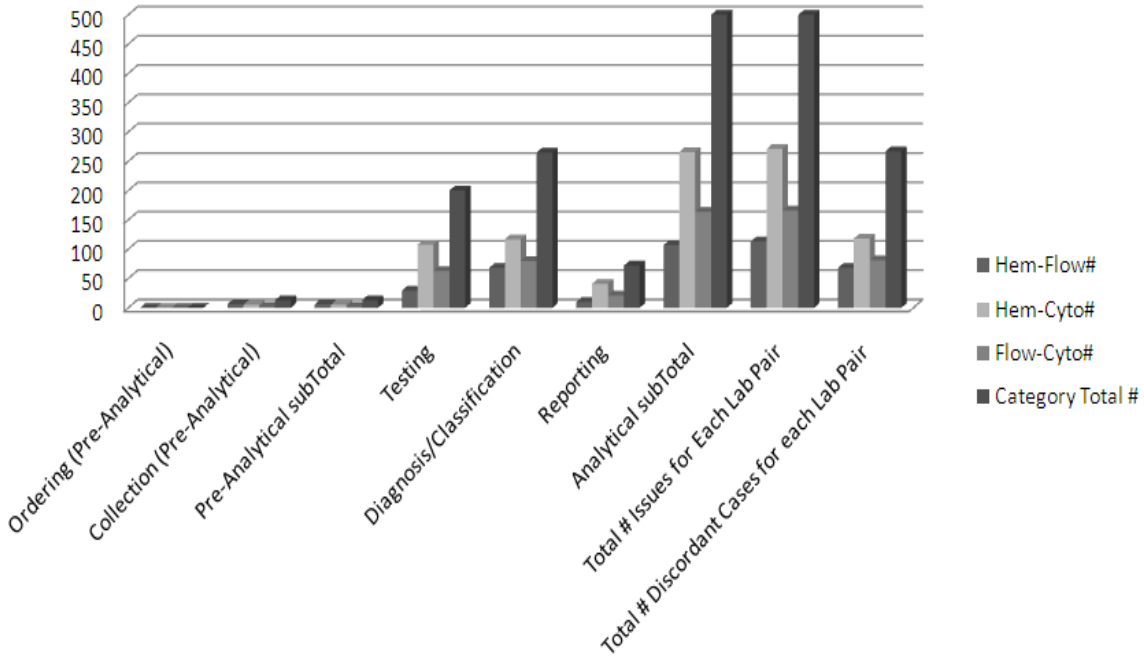
Lastly, in comparing flow cytometry and cytogenetics diagnoses, contributing factors are dispersed between the two laboratories for each case. The most common limiting factors are the same above for flow cytometry and cytogenetics. Overall, flow cytometry factors are featured most often among top limiting factors, followed by those in cytogenetics. Hematopathology factors are noted the least often.

Table 4: Top Limiting Factors of Discordant Cases by Number and Percentages								
Top Limiting Factors of Discordant Cases by Number and Percentages	# Heme based factors	% Heme based factors of total discordant cases for lab pair	# Flow factors	% Flow based factors of total discordant cases for lab pair	# Cyto factors	% Cyto based factors of total discordant cases for lab pair	# Cases per factor	Labs Compared
Acute myeloblastic leukemia with maturation from heme versus a blast population from flow that is not further differentiated as to whether it is leukemia or not or as to which subtype	0	0.00%	5	7.35%	0	0.00%	5	HF
Hematopathololgy calls this acute myeloblastic leukemia with maturation, while flow cytometry can only call this acute myeloid leukemia without further subclassifying which subtype	0	0.00%	3	4.41%	0	0.00%	3	HF
Acute myeloblastic leukemia without maturation from heme versus acute myeloid leukemia from flow, but it is not differentiated as to which type	0	0.00%	4	5.88%	0	0.00%	4	HF
Acute myeloblastic leukemia without maturation from heme versus myeloid blasts from flow which are not differentiated as to whether this is leukemia or not and to which subtype	0	0.00%	3	4.41%	0	0.00%	3	HF
Flow wasn't able to subcategorize the acute myeloid leukemia	0	0.00%	4	5.88%	0	0.00%	4	HF
Hematopathology detected an acute myeloid leukemia without maturation, while cytogenetics did not detect any disease, just a normal karyotype	0	0.00%	0	0.00%	6	5.08%	6	HC
Hematopathology classified this as Acute myelomonocytic leukemia, whle cytogenetics found a normal karyotype which is not helpful in further characterizing this patient's disease	0	0.00%	0	0.00%	5	4.24%	5	HC
Hematopathology determined this disease to be AML and MDS therapy related (unknown treatment), while cytogenetics determined that the chromosomal abnormality is consistent with AML including AML and MDS therapy related	0	0.00%	0	0.00%	5	4.24%	5	HC
Hematopathology is unable to see the cytogenetic translocation to further subclassify this leukemia	5	4.24%	0	0.00%	0	0.00%	5	HC
Flow cytometry categorized this disease as precursor B lymphoblastic leukemia and was unable to see the translocations as seen in cytogenetics when they classified this as Precursor B-cell ALL (cytogenetic subgroups) hyperdiploid >50	0	0.00%	5	6.17%	0	0.00%	5	FC
Flow cytometry results indicate acute myeloid leukemia showing partial expression of monocytoid differentiation antigens without further subclassification, while the cytogenetics indicated a normal karyotype but failed to indicate whether or not it is consistent with any leukemia type	0	0.00%	3	3.70%	3	3.70%	3	FC
Flow did not classify this leukemia, cyto did not issue a diagnosis with the karyotype to let us know whether or not it is consistent with a leukemia or not)	0	0.00%	3	3.70%	3	3.70%	3	FC
Flow cytometry was unable to subclassify this acute myeloid leukemia, while cytogenetics detected a normal karyotype and failed to indicate if it is consistent with a leukemia type or not	0	0.00%	5	6.17%	5	6.17%	5	FC
Flow cytometry was unable to subclassify this acute myeloid leukemia, while the cytogenetics results are consistent with an acute myeloid leukemia but did not indicate which type	0	0.00%	4	4.94%	4	4.94%	4	FC
Flow indicates myeloid blast population, while cyto indicates normal karyotype-unknown if leukemia or not	0	0.00%	3	3.70%	3	3.70%	3	FC

Phase II: Categorization By Where Issues Occur in the Testing Process

This classification method utilizes limiting factors from the previous sections to determine where in the testing process limiting factors occur. The limiting factors are classified into pre-analytical and analytical phases of testing and subcategorized by ordering, collection, testing, diagnosis/classification and reporting issues as in Figure 12. A limiting factor may initially only affect a portion of the testing process, but can also impact subsequent testing processes. In some cases, multiple effects result in multiple categorizations or total categorizations less than 100%.

Figure 12: Discordant Cases Classified by Where in the Testing Process Issues Occurred for each Laboratory Pair



The data in Figure 12 indicate minimal pre-analytical issues for all lab pairs with no ordering issues, and only thirteen collection issues. Analytical testing issues accounted for 98% of the issues contributing to discordances, with about half of these issues from the diagnosis and classification process. Hematopathology and cytogenetics contribute most issues to the diagnosis and classification sub-categorization. Of note, reporting issues were defined as information from another report needed by the pathologist at the

time of diagnosis. Data from this categorization indicate most issues contributing to diagnostic discordance occurred in the analytical testing phase, specifically in the diagnosis and classification process. Specific etiologies from each testing phase are detailed in the next section.

Phase II: Categorization By Etiology

Categorization by etiology is another method utilized to provide information about diagnostic discordance. Figure 13 shows the number of cases classified by their general etiology as human, disease, both or unknown. Each category is utilized once for each case comparison so etiology categorizations for each laboratory pair total 100%.

Results show a combination of human and disease etiologies contributed most to diagnostic discordances. Human etiologies contributed to discordances the second most frequently, while disease etiologies contributed the least. Among laboratory pairs, hematopathology and cytogenetics contributed over half of the etiologies from both humans and disease, while hematopathology and flow cytometry contributed most to the human etiology classification.

Figure 13: Categorization of Diagnostic Discordance by Etiology Counts

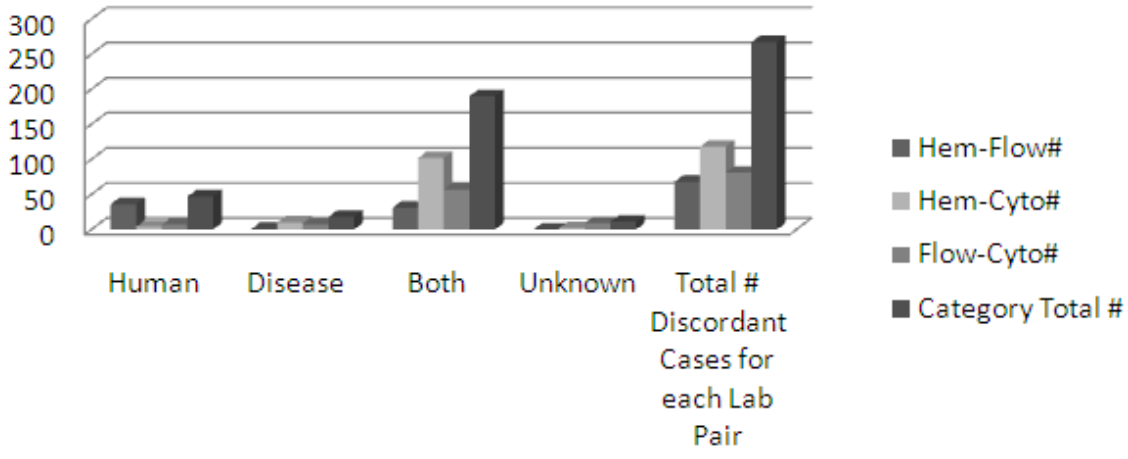


Table 5: Discordant Case Classification by Specific Etiologies					
Testing Phase	Specific Etiology	Heme-Flow	Heme-Cyto	Flow-Cyto	Category Total #
Pre-Analytical: Order		0	0	0	0
Pre-Analytical: Collection	1. poor specimen quality	1	5	1	7
	a. Fibrotic specimen (Disease)	1	0	0	1
	b. Inadequate volume	2	2	1	5
	b. Inadequate volume (Human)	0	2	1	3
	b. Inadequate volume (Disease)	0	2	1	3
	c. Diluted with blood	4	3	1	8
	c. Diluted with blood (Human)	3	3	1	7
	c. Diluted with blood (Disease)	3	3	1	7
	d. Uneven disease distribution at sample site (Disease)	0	33	22	55
	2. Sample Preparation (crush vs slides)	0	1	0	1
	a. Not enough preps of each type	1	1	0	2
	a. Not enough preps of each type (Human)	0	1	0	1
	a. Not enough preps of each type (Disease)	0	1	0	1
	b. Each sample type shows different numbers or morphology	0	0	0	0
	b. Each sample type shows different numbers or morphology (Human)	0	0	0	0
	b. Each sample type shows different numbers or morphology (Disease)	0	0	0	0
Analytical: Testing	1. Rare staining of other lineage cells*	3	0	2	5
	1. Rare staining of other lineage cells* (Human)	0	0	0	0
	1. Rare staining of other lineage cells* (Disease)	0	0	0	0
	2. Loss of expression with processing*	1	1	0	2
	2. Loss of expression with processing* (Human)	0	0	0	0
	2. Loss of expression with processing* (Disease)	0	0	0	0
	3. Too few cells available for adequate analysis*	2	3	1	6
	3. Too few cells available for adequate analysis* (Human)	0	2	0	2
	3. Too few cells available for adequate analysis* (Disease)	0	2	0	2
	4. Loss of architecture in process, damaged specimen	2	0	0	2
	4. Loss of architecture in process, damaged specimen (Human)	0	0	0	0
	4. Loss of architecture in process, damaged specimen (Disease)	0	0	0	0
	5. Detection of chromosomal abnormalities in normal patients is of unknown signficance* (Disease)	0	0	1	1
	6. Inappropriate utilization of testing and quality assurance practices* (Human)	0	0	0	0
Analytical: Diagnosis/ Classification	1. Variation in morphological diagnosis* (Disease)	9	13	2	24
	2. Categorizations prone to subjective variability* (Human)	9	10	6	25
	3. Indistinct category boundaries (ie MDS evolving to AML) (Disease)	6	12	5	23
	4. Variability in where a case fits in a categorization bin (ie cytogenetically like one disease, morphologically like another)*	18	41	62	121
	4. Variability in where a case fits in a categorization bin (ie cytogenetically like one disease, morphologically like another)* (Human)	0	0	0	0
	4. Variability in where a case fits in a categorization bin (ie cytogenetically like one disease, morphologically like another)* (Disease)	0	0	0	0
	5. Cases fitting into multiple categories*	27	96	45	168
	5. Cases fitting into multiple categories* (Human)	0	0	0	0
	5. Cases fitting into multiple categories* (Disease)	0	0	0	0
	6. Differences in opinion about which results most influence categorization* (Human)	12	24	28	64
	7. Utilization of subjective decision making in diagnosis* (Human)	48	55	61	164
	8. Categories influencing the diagnostic process* (Human)	6	0	3	9
	9. "Changing treatment options, which will affect classification to th extent that measures of "clinical relevance" influence classification."* (Human)	1	1	1	3
	10. Impact of new diagnostic testing with old catgories* (Human)	0	2	1	3
	11. Political and social factors in each department (acceptance of classifications)* (Human)	0	0	0	0
	12. Unexpected variability in disease manifestations (antigen expression, reponse to treatment, etc)* (Disease)	13	70	43	126
	13. "Tactic knowledge influences"* (Human)	0	3	1	4
	14. "Reliance on surrogates" instead of real disease markers	0	0	0	0
	14. "Reliance on surrogates" instead of real disease markers (Human)	0	0	0	0
	14. "Reliance on surrogates" instead of real disease markers (Disease)	0	0	0	0
Analytical: Reporting	10 general reporting issues were noted, but no specifics were counted.				0
	Although the data indicate the following specific issues that were noted, but not counted:				0
	1. Report does not indicate if patient had a previous MDS for diagnoses arising from a previous MDS				0
	2. Treatment regimen not reported, but required to subclassify AML and MDS, treatment related				0
	3. Report does not indicate a previous leukemia type in order to classify the current disease as a new leaukemia versus a				0

Discordant cases were also classified by specific etiologies described in Appendix C. Specific etiologies and their case counts were analyzed for each laboratory comparison as indicated in Table 5. Reporting issues were not part of the scope of this project since they occur in the post-analytical testing phase. However, there were instances where previous diagnostic information was necessary at the time of diagnosis. The most prevalent pre-analytical issue contributing to discordance for all three laboratory comparisons is an uneven disease distribution at the sample site comprising 6% of total issues. Testing issues were quite minimal as shown in Table 5. Most issues were of the diagnosis/classification type, found in the analytical testing phase. The top four issues, by percentage of total issues, were cases fitting into multiple categories (20%), utilization of subjective decision making in diagnosis (19%), unexpected variability in disease manifestations (15%) and variability in where cases fit in a categorization bin (14%). The next chapter discusses these results further.

The hematopathology and flow cytometry column in Table 5 illustrates issues by specific etiology for this laboratory pair. Of 15 collection issues, the factor contributing to the most cases is poor specimen quality due to it being diluted with blood. Rare staining of other lineage cells was the most frequently noted factor contributing to testing issues in this laboratory comparison, while "utilization of subjective decision making in diagnosis," was the most frequently noted diagnosis/classification issue with 48 cases.

No reporting or ordering issues were noted in the column for the hematopathology and cytogenetics laboratories comparison found in Table 5. However, the most common collection issue with 33 cases noted is an uneven disease distribution at the sample site. This issue is attributed to a disease process when one specimen shows disease and another does not. Under testing, having too few cells available for adequate analysis occurred in 3 cases, while 96 cases in the diagnosis/classification categorization featured cases fitting into multiple categories. The latter commonly occurred when it was reported that more than one WHO diagnosis was possible.

Similarly in the column for flow cytometry and cytogenetics laboratory pair in Table 5, no cases with reporting or ordering issues were noted. The most common collection issue with 22 cases noted was an uneven disease distribution at the sample site with one specimen indicating disease and the other specimen lacking disease. The testing issue most frequently noted is rare staining of other lineage cells with only two cases. The diagnosis/classification issue noted most often with 62 cases is variability in where a case fits in a categorization bin. These cases are cytogenetically like one disease, but flow cytometry is reporting them as a different disease.

Phase II: Categorization By Reporting Issues

The last method of categorization of discordant cases is by "Reporting Issues." Figures 14, 15 and 16 show the percentage of reporting issues with each laboratory pair.

Figure 14: Discordant Cases Classified by Reporting Issue Total Percentages

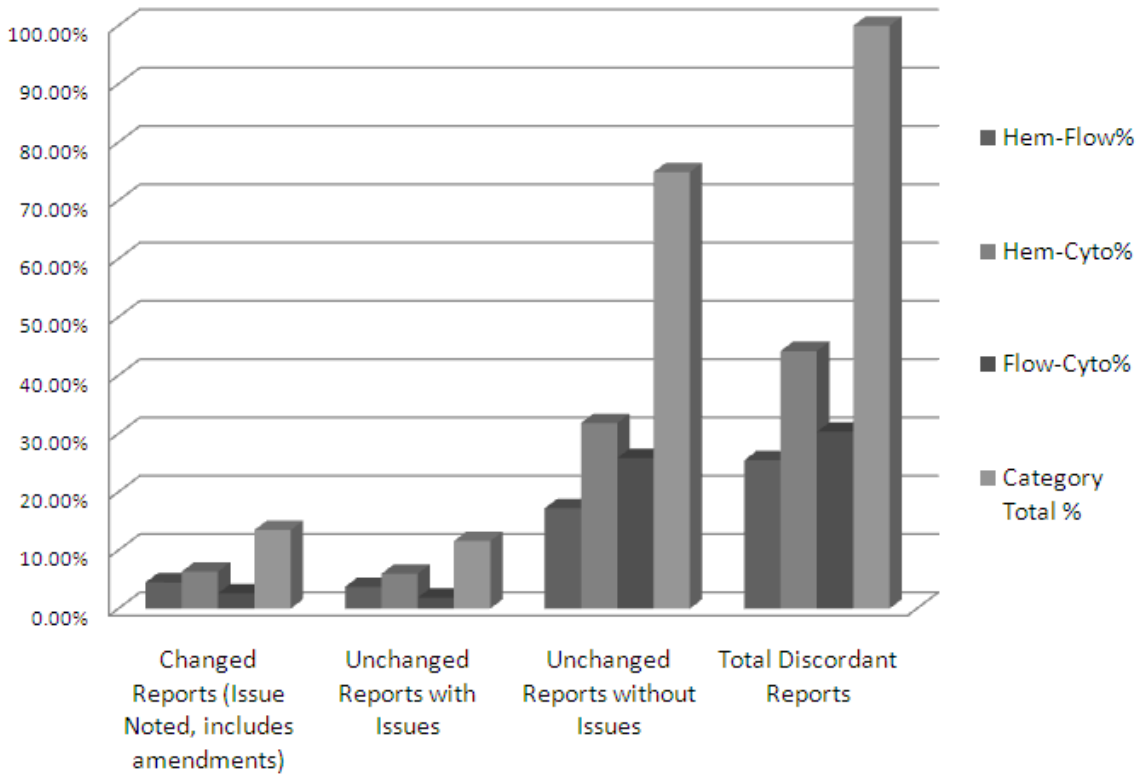


Figure 14 indicates 13% of reports were changed, 12% were unchanged with issues and 75% were unchanged without issues, accounting for 100% of discordant reports. Changed or amended reports indicate reporting issues have been noted and corrected. "Unchanged reports without issues," indicates normal reports lacking reporting issues, while "unchanged reports with issues," indicates missed reporting issues. Hematopathology and cytogenetics laboratories had the highest rate of changed reports (6%), unchanged reports with issues (6%), and normal or unchanged reports without issues (32%). Results indicate this laboratory pair contains the highest rate of normal reports, amended and changed reports, and reports with undetected issues. The amendment rate is consistent with the practice of amending hematopathology reports to add immunohistochemical stains or other testing results. However, these findings do not account for the rate of unchanged reports with issues indicating undetected issues.

Figure 15: Discordant Case Classified by Typographical Reporting Issue Percentages

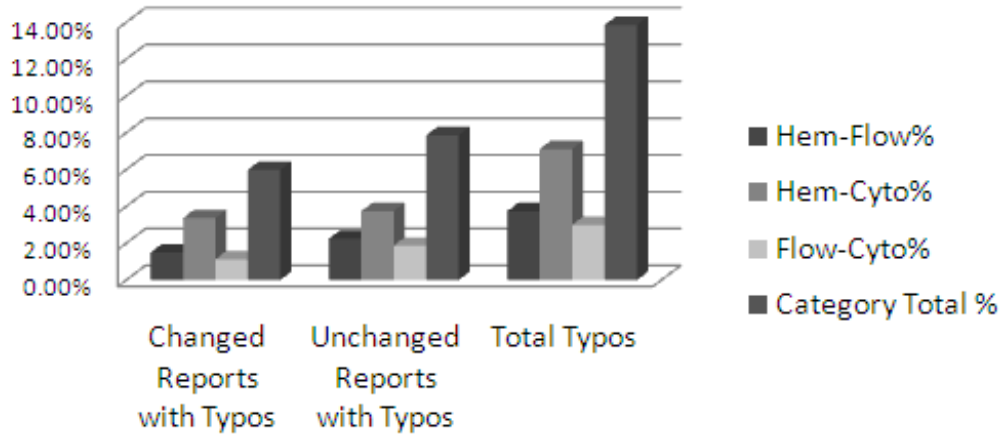


Figure 15 depicts classification of discordant cases by typographical issues among both changed and unchanged reports. The changed reports designation indicates typographical errors were noted and corrected, while unchanged reports indicates typographical errors were present, but not addressed. Typographical errors were found in 14% of discordant reports with 6% of them in changed reports and 8% of them in

unchanged reports. Among laboratory comparisons, hematology and cytogenetics had the highest number of typographical errors, both in changed (3%) and unchanged reports (4%).

Figure 16: Discordant Case Classified by Contradictory Reporting Issue Percentages

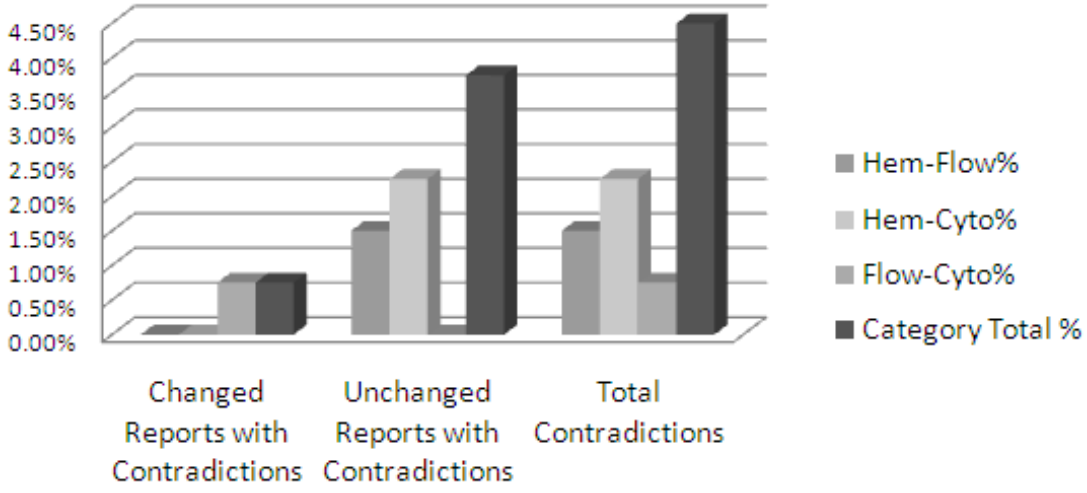


Figure 16 illustrates classification of discordant report issues by contradictions found in changed and unchanged reports. Total contradictions were low at 4.5%, but most contradictions were in unchanged reports indicating they were undetected. Hematopathology and cytogenetics led labs with 2% total contradictions, while flow cytometry and cytogenetics was the only pair with a changed report with a contradiction. Overall, few reports had contradictions.

Chapter 5

Discussion: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Phase I: Determination of Lexical Diagnostic Discordance

Lexical diagnostic discordance among the cytogenetics, hematopathology, and flow cytometry laboratories was successfully detected and measured in Phase I of the research. The first hypothesis is supported by these findings.

The research methods used here were designed to replicate how some decision support tools function. Therefore, detection of discordance was performed manually, while measurement of discordance was computerized. Data show lexical percent agreement can be successfully measured retrospectively among each type of paired laboratory comparison. This method revealed pathologists reported a wide variety of diagnoses. This variety prevented successful Kappa determination of lexical diagnostic concordance. Therefore the original study design was modified to include detection and measurement of semantic diagnostic agreement.

Lexical discordance analyses revealed some interesting findings. First, bone marrow diagnostic discordance results have not previously been reported in the literature. A zero discordance rate or complete agreement was not expected since healthcare is not perfect. The findings indicate about 95% lexical discordance for each laboratory comparison, which is much higher than anticipated.

Raab reported discrepancy rates indicating disagreement between diagnoses from a primary and secondary reviewer ranged from 6.7% in anatomic pathology, to 6.8% in surgical pathology, and 6.5% in cytology (76). Raab's experiments were based upon measurement of semantic discordances, not lexical discordances. Raab's studies do provide valuable information about the rates of discordance in other areas of pathology. Organ specific discrepancy frequency was reported to be 1.7% in bone marrow specimens

(76). However Raab's study focused on histological analysis of bone marrow specimens in the anatomic pathology laboratory and not analyses by the cytogenetics, flow cytometry or hematopathology laboratories.

These findings provide the first known data regarding bone marrow biopsy diagnostic discordance rates. Data indicate lexical disagreement ranges from 92% to 95% among laboratory pairs. Data also indicate successful measurement of diagnostic discordance with the methodology described. From an informatics perspective, detection of diagnostic discordance can be realized by the automation of the methodology utilized in this study. One means of automatic discordance detection is natural language processing, but its application in pathology is still in the research stage (77).

The research also revealed pathologists employ a variety of practices in reporting bone marrow diagnoses. Some of these practices are problematic as they contributed to the high levels of diagnostic disagreement found with lexical methods. Table 1 demonstrates this variety with case data. Example cases one and six illustrate the observed phenomenon. Pathologists report out multiple diagnoses for a specimen, add terminology to WHO diagnostic categories, use non-WHO diagnoses, fail to subcategorize a case or do not fully interpret test results when issuing a diagnosis. All these practices contribute to non-standard terminologies and diagnoses being reported, despite recognition of the WHO diagnoses as the gold standard. Standard terminologies are utilized to ensure physicians are communicating about the same disease or medical terms. Use of non-standard terminologies may not clearly communicate the patient's disease process and cause confusion amongst pathologists or clinicians. These practices also explain why lexical diagnoses issued by each laboratory were not in agreement.

Multiple diagnoses reported by one laboratory may indicate uncertainty about either diagnosis. When compared to a certain diagnosis issued by the other laboratory, it is not clear as to which diagnosis prevails assuming at least one of the two diagnoses are in agreement with the other. In some cases, reporting additional information about the findings or disease may aid clinician decision making. The disease may take on

characteristics from more than one WHO categorization. For example, an "acute monocytic leukemia" occurring after a treated myelodysplastic syndrome, could be reported as "acute myeloid leukemia and myelodysplastic syndrome, therapy related," or "acute monocytic leukemia," or both. It is unclear how reporting multiple diagnoses impacts physician decision making with regards to treatment and patient care. When laboratories typically issue a single diagnosis, and a clinical decision support tool is designed with that premise, reporting multiple diagnoses could yield unexpected results.

Similarly when the pathologist adds terminology to the WHO diagnostic category, it may be unclear what disease state is described. For example from a diagnosis of "acute myeloid leukemia with maturation with aberrant T-lymphoid antigen expression," it is unclear whether this represents a myeloid disorder, a lymphoid disorder or some combination such as "acute biphenotypic leukemia." Each diagnosis requires a different treatment regimen. Therefore it is imperative to know which diagnosis prevails.

The previous diagnosis describes not only reporting of WHO diagnoses with additional terminology, but also utilization of non-WHO diagnoses. This aberration must be considered in decision support tool development as it may impact the function of such a tool.

Lastly, when pathologists fail to subcategorize a diagnosis, discordance rates are impacted. Several cases reported flow cytometry data without a diagnostic interpretation. Without a diagnosis, it is unclear if the clinician is able to interpret the results. How does this impact clinical decision making and provision of patient care? Several other cases contained a general diagnosis such as "acute myeloid leukemia." It is unclear whether this general diagnosis is in agreement with a more specific "acute myeloid leukemia" diagnosis.

Although this research does not attempt to determine why these aberrations occur in pathology practice, it would be an interesting study. This research does categorize discordances to provide information about factors contributing to these discordances.

Considering the informatics perspective, these lexical diagnostic discordances would have been flagged by a clinical decision support tool utilizing the methodology in this study. Many flagged discordances would be false positive since reported diagnoses are describing the same disease concept, even if they differ lexically. Pathologists would concur that these flags would be false positives. If pathologists encountered high false positive alerts, the alerts would be quickly ignored. Effective clinical decision support needs to accommodate lexical differences and to rely instead on semantic differences.

Another solution in overcoming issues posed by lexical diagnostic comparisons is to utilize a standardized terminology. The data have shown reporting variations of WHO diagnoses leads to discordances. By integrating standard WHO diagnoses or a standard terminology such as the Systemized Nomenclature of Medicine-Clinical Terms, SNOMED CT®, pathology diagnostic reporting variability would likely be reduced. Standard terminologies can be easily integrated into pick lists and other decision support tools. Furthermore, since SNOMED CT® features a fully specified name with an associated numeric code, the numeric code can be deployed within a decision support tool to compare diagnoses for agreement. However, this can only be achieved by standardization of diagnostic reporting processes.

In conclusion, although diagnostic discordance can be successfully detected and measured by lexical comparison, lexical discordance alerts are not an effective means of alerting pathologists due to their high false positive rates. Furthermore, automatic detection of lexical discordance by natural language processing is still in the research phase and not currently applicable for clinical use.

Phase I: Determination of Semantic Diagnostic Discordance

Semantic diagnostic discordance was successfully determined and measured via percent semantic agreement and Cohen's Kappa statistic among the laboratories reporting bone marrow diagnoses. The data also support the second hypothesis that diagnostic discordance can be detected and measured with these methods.

Semantic comparisons revealed some interesting findings. First, percent agreement was successfully calculated as a measure of agreement in performing semantic diagnostic comparisons. Figures 2, 3, and 4 illustrate more cases are in agreement semantically, than lexically, despite a number of unclassifiable cases. A number of cases are deemed "unclassifiable" utilizing semantic comparison. Semantic comparison methodology accommodates for variances in diagnostic terminology regarding a disease process. These cases would be flagged as discordant via lexical methods, but be considered in agreement via semantic methods. These findings indicate semantic comparison better measures diagnostic agreement and discordance since it mirrors actual pathology practice.

The second interesting finding in determination of semantic discordance was the need to designate cases as unclassifiable when it was unclear whether the diagnoses were in agreement or not due to a lack of information. The unclassifiable categorization features several types of cases. The first occurs when one laboratory issues a diagnosis that agrees with one of two diagnoses issued by a second laboratory. Unclear agreement and disagreement in the diagnoses results in the unclassifiable designation.

The second type of case occurs when one laboratory issues a general diagnosis and a second laboratory issues a consistent, but more specific diagnosis. Without additional information, it is unclear if the diagnoses are in agreement or not. This relationship is often seen between the hematopathology or flow cytometry diagnosis and that of the cytogenetics laboratory when the first two labs cannot determine cytogenetics findings with their testing methodologies. Therefore, this is not a true discordance, but rather a limitation of the testing process.

The third type of unclassifiable case is similar to the second type. It occurs when one laboratory reports test results without a diagnostic interpretation. Frequently flow cytometry reported positive flow markers, but failed to indicate with which leukemia the results were most consistent. Reports often have sufficient information in the findings for pathologists to issue a general diagnosis, if not a specific WHO classification. It is

unclear why the pathologists are not providing diagnostic interpretations. In some cases, there is insufficient information available to the pathologist at the time of diagnosis.

Ideally, one would expect diagnoses issued by all three laboratories would always be in complete agreement. This was rarely observed. Given usage of rigorous measures to ensure testing process quality, it is reasonable to assume that any testing process issues would be noted. Appropriate measures would be taken to resolve such issues by laboratory staff and questionable testing results would not be reported. Therefore finding unclassifiable cases indicates a problem in the information process, most likely in pre-analytical or post-analytical testing consistent with literature reports.

Phase I: Semantic Diagnostic Discordance Determination by Kappa

Kappa results indicate it can be utilized as a measure of agreement in assessing diagnostic discordance for bone marrow biopsies, thus supporting the third hypothesis. Cohen's Kappa statistic tested the third hypothesis, Kappa=0, indicating diagnoses in agreement are not significantly different than by chance alone (70). Data for the determination of semantic diagnostic discordance show Kappa is greater than zero in all three laboratory comparisons. Diagnostic agreement measured by Kappa is statistically significant or more than by chance alone. Although Kappa and agreement levels were unknown prior to analysis, Kappa was anticipated to be between zero, indicating chance diagnostic agreement, and one, indicating complete agreement.

Several aspects of experimental design need to be considered with the analyses. Kappa's outcome can only be "agree" or "disagree," whether calculated manually or automatically with SAS's "proc freq." Therefore, the unclassifiable category was not considered in calculating Kappa. Instead, most unclassifiable cases by semantic diagnostic comparison were designated as disagree by Kappa. This limitation indicates that Kappa is dependent on the data input.

Dependence on the data input for calculating Kappa was also evident in detection and assessment of lexical diagnostic discordance. Great variability in lexical diagnoses

prevented SAS's "proc freq" from successfully calculating the Kappa statistic. Data are required to be in a square contingency table for Kappa to be successfully calculated. Diagnostic variability prevented both form occurring. Kappa was successfully calculated from semantic diagnostic comparison since there was less variability and fewer unique diagnoses.

Semantic Kappa measurements indicated moderate and substantial agreement levels in Figures 2, 3, and 4, and Table 3. Cases with a laboratory issuing multiple diagnoses or that were otherwise unclassifiable were designated as "disagree," by Kappa methodology.

Another key assumption in experimental design with calculation of Cohen's Kappa statistic is each laboratory diagnosis is rendered independently from the other. In clinical practice this may not be the case as pathologists may consult with each other. However, if consulting occurred, few or no disagreements and perfect agreement would be expected. Despite these considerations with Kappa, data indicate Kappa can successfully be calculated in the comparison of bone marrow diagnoses as a measure of agreement.

Furthermore, data support the third hypothesis that bone marrow diagnostic discordance can be detected and measured semantically by Cohen's Kappa statistic. However, data do not support this hypothesis for lexical determination of Kappa. Kappa utilized in combination with WHO categorizations or standardized terminology would be expected to be an excellent measure of agreement. Furthermore, automation of Kappa calculation with this semantic diagnostic comparison could be easily incorporated into a clinical decision support tool to alert pathologists to diagnostic discordances.

Phase I Discussion Summary

Results from this phase of research have demonstrated diagnostic discordance, as typically detected and measured in anatomic pathology, can be successfully applied and measured in hematopathology, despite limited literature reports. Three methods were

utilized to detect diagnostic discordance: lexical comparison, semantic comparison and comparison via Kappa statistic. The first two employed manual or human detection of discordance, followed by automatic calculation of percent agreement. The last method featured automatic detection and calculation of discordance from appropriately formatted data by the "proq freq" function in SAS resulting in Cohen's Kappa statistic. Data indicate that semantic and Kappa methods are superior to lexical methodology and both can be integrated into decision support tools in aiding the pathologist in providing quality patient care.

Semantic methods would be a significant aid to the pathologist when utilized in combination with a standardized terminology. One problem revealed is that pathologists utilize different diagnostic text in their reports to describe the same disease process. By providing standardized text choices in the form of the WHO hematological classifications to the pathologist at the time they generate their report, variability in the text of the diagnoses can be eliminated. Furthermore, use of a standard terminology such as SNOMED CT®, which uses standard text, synonyms and encoding for computerization, would further enrich decision support to the diagnostic process. Standard terminologies permit computerization of decision support and automatic semantic encoding at the back end of the report. They also permit measurement, comparisons, and analyses to be performed with encoded data. Each LIS is capable of providing standardized terminology, decision support and semantic interoperability functionality. However, these functions may be implemented differently in each LIS.

One current limitation of pathology reporting decision support tools is they retrospectively provide pathologists with a list of diagnoses on the same specimen issued by each laboratory (66). Pathologists utilizing these contemporary tools often detect discordance themselves manually, well after final reports have been issued. This process is sufficient for quality assurance monitoring, but not ideal for real-time pathology reporting and patient care. Ideally, a pathologist would be alerted to a potential diagnostic discordance prior to the report being sent to the clinician.

For example, one LIS may provide the pathologist with a drop down menu from which to choose standard WHO diagnostic categories from within their report writing program. Behind the scenes each WHO diagnosis can be semantically encoded. When the report is complete and saved, before it is finalized with an electronic signature, a rule may be triggered which performs a search for other bone marrow specimens on this patient. If it is determined that cytogenetics or flow cytometry were ordered, the diagnosis issued by the hematopathology laboratory could be compared to that of the flow cytometry laboratory. This could be done by simple automation of our semantic comparison methodology or it may be done via calculation of the Kappa statistic. Either way, if there is not agreement in semantic codes, an alert would fire letting both pathologists know of each diagnosis for further resolution. This may occur just prior to the pathologist finalizing the pathology report to ensure no discrepant results are reported to the clinician.

Another LIS may integrate standard terminology via a checklist format, such as the CAP Cancer Protocols. The computerized version of the checklist permits encoding with standardized terminologies. The pathologist selects the button next to the appropriate WHO diagnosis listed. Once the checklist is complete as indicated by the pathologist selecting a button, a transformation process is triggered that encodes the selected items with a standardized terminology. During this process, a rule may fire providing the capability to search for other bone marrow specimens and their diagnoses on this patient. This process can be performed with automated Kappa statistic or semantic methodologies, to determine if an alert should be generated for the pathologist if discordance is found.

In addition to real-time detection of diagnostic discordance, percent semantic agreement and Kappa can be utilized for quality assurance reporting. They could be integrated into a quality assurance module to automatically retrospectively compare diagnoses over a time for discordances among different laboratories within an institution. Discordance monitoring may also be utilized for noting the response to interventions (training, decision support, or process redesigns), performing diagnostic correlations to

comply with regulatory requirements, for comparing institutions to each other and other similar quality comparisons. One limitation to this process is there is no standard as to what level of Kappa is considered significantly discordant, especially clinically. Additional research and consensus from the medical community are needed to determine appropriate Kappa decision standards. Kappa is already utilized in some anatomic pathology laboratories so data exists to determine Kappa standards in this realm, but further data is needed in hematopathology. There are many potential uses for this methodology and corresponding data collected.

Although Kappa and semantic discordance methods can be integrated into prospective decision support programs, assessment of discordances is performed by the pathologist. One drawback with rules based comparisons is any discordance will result in pathologist alerting. Discordances occurring due to limitations in the testing process will result in false alerting, which pathologists learn to ignore, sometimes referred to as alert fatigue. Alert fatigue can also result in the ignoring of truly positive alerts. Additional functionality is needed in the discordance assessment process. Phase I discordances are categorized in Phase II to gain information about their contributing factors. This information permits the assessment of which discordances are really true discordances and not due to laboratory testing limitations.

Phase II: Categorization of Discordant Cases

The second phase of research tests the fourth hypothesis that categorization of factors contributing to diagnostic discordance allows one to distinguish between discordances due to limitations in the testing process and those due to other causes. These other discordances require alerting a pathologist. Categorization of discordances from the first phase provides additional information about discordances themselves and factors contributing to these discordances.

Phase II: Categorization By Level

This first categorization utilized semantic comparison discordances from the first phase to diagram the relationships between diagnoses in Figures 5, 6, and 7. These diagrams further highlight the wide variety of diagnoses reported by pathologists, especially where they vary from the WHO diagnostic categorizations. Figures 8 and 9 also show the impact the variety of diagnoses have on categorization by relationship level as most of the cases are unable to be classified due to this variety.

Focusing on relationship diagrams first, data indicate in each laboratory comparison, diagnoses drift from the WHO categorizations. These differences may be due to two diagnoses being reported by a lab, on laboratory reporting only the results without a diagnostic interpretation, or one laboratory reporting a more specific WHO diagnosis than the other. By looking at these data by level, more information is learned concerning the diagnostic relationship.

The vertical relationship or parent–child relationship shows relationships in which one diagnosis is more general at the parent level and one diagnosis is more specific at the child level. This categorization accounted for 27% of the discordances between the hematopathology and flow cytometry diagnoses and overall accounted for the third most frequent type of discordance by level. This finding suggests that there is a lack of sub categorization occurring with one of these laboratories in reporting leukemia diagnoses. Case 2 in Table 1 is an example of one laboratory reporting a specific diagnosis, while the other reports a general diagnosis exhibiting a vertical relationship. Further analysis of the pathology report is needed to determine if discordances of this type are due to a limitation in the laboratory testing process as often occurs when hematopathology is unable to detect cytogenetic translocations. It is unclear why a more specific diagnosis is not being reported, especially if the pathology report contains sufficient information for one to be made.

Vertical relationship issues correspond to what Raab calls an overcall or undercall error as he considers all discordances as a type of error (12). However, this author and many in the pathology community would agree that not all discordances are errors. Limitations in

laboratory testing are not usually considered errors. This research shows that the categorization of discordances is vital in making this distinction.

Horizontal or sibling relationships were least frequently noted with the data. These relationships usually feature a characteristic making them distinct from each other, even though they are closely related. The data indicate that although these discordances occur most often between the hematopathology and flow cytometry diagnoses, they are still far less frequent than the "other cases," and "Unable to classify cases," categorizations.

"Other cases," categorization by level occurred when there was both a horizontal and vertical component of the relationship, such as those on the diagonal. These types of relationships occurred second most frequently comprising about one third of the discordances notes. They featured diagnostic characteristics that are less closely related. Examples of this categorization include cases where one laboratory issues a specific diagnosis while the other lab issues two diagnoses where one of the diagnoses agrees with the first lab and the other does not. It is unclear which of the two diagnoses best describes the patient's disease, as the disease is usually described by a single diagnosis.

Lastly, when the relationship between the diagnoses was unknown and they were unable to be categorized, they were designated as the "unable to classify cases." Cases in this category occurred most frequently, comprising about one half of the noted discordances. This categorization also includes cases where one laboratory diagnosis is for a specific leukemia, but the other laboratory was not able to find evidence of leukemia. Therefore it is unclear if the patient has leukemia or not. Cases where laboratory findings were not interpreted are also included in this categorization. Raab indicates that any discordance more than a minor categorization change is more likely to be more clinically significant (12). Although this study does not look at the clinical impact discordances have upon the patient, these findings imply that those cases in the "other" and "unable to categorize" categories could have a significant impact on patient care.

The lower frequencies of horizontal and vertical relationships indicate that most diagnostic discordances are not due to discordances that are closely rated vertically or

horizontally within the WHO categorizations. This is supported by the larger number of discordances falling into the "other cases" and "unable to classify cases," categories. That is, the diagnoses are more different than they are alike.

Another explanation for the findings may be due to limitations in the laboratory testing methodology. Pathologists would agree such limitations are not errors despite the presence of a diagnostic discordance. A limitation of laboratory testing occurs when one laboratory reports a diagnosis that is more general that another laboratories diagnosis due to differences in the detection limits and sensitivities of the testing methodologies employed. This phenomenon is commonly seen when the hematopathology laboratory is unable to detect cytogenetic translocations, while the cytogenetics laboratory is able to do so. Hematopathology may report a general diagnosis of acute myeloid leukemia, but cytogenetics issues the specific cytogenetic based WHO category.

This problem is expected to increase in frequency as more molecular diagnostics and cytogenetics testing is introduced. There will be an increase in the reporting of specific genetic based diagnostic categorizations which are unable to be seen utilizing historic gold standard methodologies such as microscopy in hematopathology. It remains to be seen how the pathology community will respond to the increase and complexity of information that must be considered in the future.

In conclusion, categorization of discordances by level provides information concerning the type of relationship between the diagnoses and indicates how closely the diagnoses are related. Raab indicates that discordances more than a minor categorization difference can be of clinical significance. Categorization by level allows for the distinction to be made between minor categorization differences such as vertically or horizontally from those that are likely to be more significant when they are in the "other" category or are unable to be categorized.

Phase II: Categorization By Limiting Factors

Categorization by limiting factors is the second method utilized in this research to categorize diagnostic discordances. Results from this methodology indicate whether or not laboratory based issues or non laboratory issues contribute to the diagnostic discordances found in Phase I. Furthermore, this analysis indicates where information is lacking, where a clinical decision support tool can best be implemented and how other factors impact bone marrow diagnostic discordance.

Results indicate most factors contributing to diagnostic discordance occur in the clinical laboratory, as evident in each laboratory comparison. Despite many laboratories having more than one factor contributing to the discordances, some trends emerge in the data. In the hematopathology and flow cytometry comparison, over 70% of the limiting factors were flow cytometry based factors, while about 15% were from hematopathology. In comparing hematopathology and cytogenetics, the hematopathology limiting factors approached 40%, while those from cytogenetics were about 80%. In comparing flow cytometry and cytogenetics, flow and cytogenetics contributing factors were 72% and 77% respectively. Data indicate cytogenetics has the most factors, followed by flow cytometry and hematology. Furthermore, hematopathology based factors, involved in about 15% of discordances, are not a major contributing factor to discordances, unlike cytogenetics and flow cytometry laboratories. These findings support the hypothesis that categorization of discordances provides additional information in determining which laboratory based factors impact clinical decision making the most.

Further analyses determine the top specific factors which contribute to discordances for each laboratory as shown in Table 4. Of the few hematopathology factors, the top one contributing to diagnostic discordances is the inability of the hematopathology tests to determine cytogenetic translocations which results in their classifying leukemias into non-cytogenetic WHO diagnostic categories. Cytogenetics can issue a specific cytogenetics based WHO diagnostic classification with their testing methodology. In hematopathology, this factor is a limitation of the laboratory testing process.

Flow cytometry methodology is also unable to determine cytogenetic translocations, and consequently reports diagnoses from non-cytogenetic WHO categories. Another flow cytometry factor is a failure to subclassify a diagnosis by a laboratory. Many times the reported diagnosis was simply "acute myeloid leukemia" with no indication of type. Other times, the diagnosis did not indicate if the results support a leukemia diagnosis of "myeloid blast." It is unclear as to why reported diagnoses are not being further subclassified, especially when there is enough flow marker information to further focus the differential diagnoses. Percentage of blasts is another piece of information determined by flow cytometry. It is also a criterion for leukemia diagnosis that is being reported, but not reflected in the diagnosis. Results again confirm sub-categorization is not occurring in the flow cytometry diagnostic process. This factor is not a limitation of the laboratory testing process since test results are being reported.

Cytogenetics also featured limiting factors contributing to diagnostic discordances. One factor is the finding of a normal karyotype by the cytogenetics laboratory while another laboratory finds evidence of disease. In some cases, the cytogenetics laboratory indicates whether or not the normal karyotype is consistent with leukemia or not, while in other cases, this consistency is not reported. The second factor occurs when cytogenetics reports out results that are consistent with an acute myeloid leukemia, but does not specify which one, even though the other laboratory issued a diagnosis with a specific leukemia subtype. Again both of these are sub-classification issues, warranting an alert to be made to the pathologist. This study does not look at why sub-classification is not routinely occurring or why further interpretive reporting is not occurring although they would both be interesting studies. A future aspect of this research would be to study the impact diagnostic discordances have on clinician decision making and patient outcomes.

In other cases, other non laboratory factors contribute to discordances. These include insufficient information available to the pathologist at the time of diagnosis, not attributed to aspects of the laboratory testing process. One example occurs when the patient's previous diagnosis or treatment regimen is unavailable in considering a diagnosis of "acute myeloid leukemia and myelodysplastic syndrome, therapy related." Decision

support can be built into a LIS or pathology reporting system to provide pathologist access to information critical to a diagnosis. Information may be in a pharmacy information system, in the clinical history, or even in a radiology report. Having the right information at the right time and in the right format is essential for an accurate diagnosis.

Having the right information is essential not only for the pathologist, but also other end users of information generated by the pathology laboratory. These include the clinician, public health entities, and cancer registries. The College of American Pathologists recognizes the need for pathology reports to include essential information in reporting cancers such as leukemia. Their cancer checklists, including the bone marrow checklist, include essential elements mandated by the American College of Surgeons Commission on Cancer (78) for cancer reporting. The bone marrow checklist features WHO categorizations including the sub-classification of leukemias by cytogenetic subtype or therapy. The checklists also incorporate information from various laboratory areas such as flow cytometry, cytogenetics and cancer biomarkers.

Other non-laboratory factors involve the specimen or disease process. Insufficient specimen contributes to diagnosis discordance when it limits the number of tests performed on a specimen thereby limiting the information reported. This human factor impacts decision making, but can be rectified by initially obtaining additional specimens. Another disease related factor occurs when a disease fits into two categorizations upon review of the test results by the pathologist. A pathologist needs to be alerted to discordances in this case to decide which categorization is most clinically relevant for reporting.

The data in Figures 10 and 11, as well as Table 4 support the hypothesis that categorization of discordances helps delineate problematic discordances from those which are limitations of the lab testing process. Limitations of the testing process may not warrant an alert to the pathologist, while other factors which occur should result in an alert to the pathologist. These are important considerations in the design of a clinical decision support tool.

Determination of factors contributing to diagnostic discordance from an informatics perspective is similar to root cause analysis and human factors research methods (79). However, with root cause analysis, the investigation typically is initiated by a medical error that has occurred. This study analyzes diagnostic discrepancies retrospectively and proactively prior to a mitigating error precipitating the investigative process. Diagnostic discordances may occur in situations not due to error as with laboratory limitations. Regardless, contributing factors or latent conditions may contribute to error or affect the patient outcome if not addressed (79). It is important to determine which contributing factors are involved in discordances so that interventions such as decision support tools may be implemented to prevent errors.

This research also confirms that current clinical decision support systems that do not distinguish diagnostic discordances due to laboratory limitations versus other etiologies are inadequate for pathology needs. As previously mentioned, pathologists will not adopt a clinical decision support alerting system that results in many false positive alerts. The pathology community views alerts due to laboratory testing limitations as one example of a false positive alert. Therefore, future clinical decision support tools need to incorporate mechanisms that make this distinction. The first line mechanism of detection of diagnostic discordance could be achieved with Kappa or other initial comparison of the diagnoses. On an individual or small dataset basis, Kappa values of zero indicate no agreement, while a Kappa of one indicates complete agreement. Next, proper assessment of the discordance would need to be performed. This can occur via rules-based methodology or other artificial intelligence techniques as exemplified by the expert system developed by Nguyen and Diamond which interprets flow cytometry data (80). Unfortunately, the right combination of artificial intelligence techniques with the pathological assessment which differentiates true discordances from laboratory limitations has yet to be fully developed.

In conclusion, diagnostic discordances can be successfully classified by limiting factors methodology. This methodology distinguishes between laboratory and non-laboratory issues, and indicates which contributing factors occurs most with each laboratory. This

categorization also confirms that laboratory limitations need to be distinguished from other contributing factors in designing future clinical decision support tools in order to foster the widespread acceptance and adoption of such tools by the pathology community.

Phase II: Categorization By Where Issues Occur in the Testing Process

The third method categorizes diagnostic discordances by where in the testing process issues occur. Again the discordant cases from the first phase of research were utilized for this analysis. The focus of this method is on aspects of the testing process in the pre-analytical phase, including ordering and collection, and the analytical phase, which involves testing, diagnosis/classification, and reporting. Ideally, there should be no issues in the bone marrow testing process or information produced as a result due to the stringent quality control processes laboratories employ. However, since this process is not perfect, issues may arise. Information about where in the testing process issues occur is helpful in determining where interventions such as decision support are best utilized.

Beginning with the pre-analytical issues, no ordering issues were noted and collection issues comprised on average 2.36% of total issues contributing to diagnostic discordances among all three of the laboratory comparisons. This is important to note as studies have shown that one of the most common pre-analytical issues is patient misidentification errors at the time of collection. Although more common in other areas of laboratory testing, it did not seem to be a factor in our research. Perhaps it is due to bone marrow biopsy testing being a more invasive procedure that is given more forethought than the more routine peripheral blood draw. Thus, longer patient interaction may allow patient identification issues to be resolved prior to bone marrow collection.

Analytical issues comprised 98% of total issues on average for all three laboratory comparisons contributing to diagnostic discordances. Within the analytical phase subcategories, most issues occurred in the diagnosis and classification subcategory. Issues were fewer in the highly regulated testing subcategory and reporting subcategory.

In all three laboratory areas, diagnosis and classification issues greatly outpaced those from testing and reporting. One explanation is diagnosis and classification contains a number of subjective and objective components which may vary with individual pathologist practice. It appears that the pathologist does not have the right information at the right place and at the right time when the diagnosis and classification of the disease occurs.

These findings correlate with the findings in the previous section whereby decision making is affected by insufficient information at the time of diagnosis in some cases. The findings support the hypothesis that categorization distinguishes those factors or issues impacting patient care the most. In this case, those from the analytical phase of testing, specifically in diagnosis and classification point to the need for an intervention such as a decision support tool depending on the specific etiology. Further details about the specific issues from each phase of testing contributing to discordances are discussed in the section, "categorization by etiology."

Furthermore, if better communication was occurring among laboratories prior to issuance of their final reports, one would expect to see better levels of diagnostic agreement. This is especially true with the hematopathology and flow cytometry comparisons. The diagnosis and classification issues contributing to diagnostic discordances are much higher in this laboratory pair than those seen in the other two laboratory comparisons. The data indicate that the hematopathology and flow cytometry laboratories would benefit from a clinical decision support tool.

Phase II: Categorization By General Etiologies

Categorization by General Etiologies involves categorizing diagnostic discordances by human, disease, both or unknown etiologies. This method provides information concerning whether a human or disease issue contributed to the discordances noted. It also addresses the research question, "Can diagnostic discordances be categorized to further characterize the issues contributing to diagnostic discordances?" Knowledge of

the etiology type guides the choice of intervention needed such as clinical decision support or education and training.

Data in Figure 13 indicate for each laboratory comparison, a combination of human factors and disease factors predominantly contributed to discordances. The exception occurred in the hematopathology and flow cytometry comparison, where etiologies were noted to be mostly human. The findings support the hypothesis that the classification of discordances distinguishes which factors impact patient care and clinical decision making the most.

It is not surprising that human etiologies contribute to most of the discordances between hematopathology and flow cytometry, as well as factoring into discordances found in the other laboratory pairs. These findings suggest that interventions such as education or training may best prevent discordances depending on the specific etiology. One common human etiology, patient misidentification, was not supported with the data. Other human etiologies prone to diagnostic discordance issues would benefit from a clinical decision support tool alerting the pathologist prior to report finalization.

In considering human etiologies, pathology workflow and testing processes needs to be reviewed from an informatics perspective. Information generated by the testing process will only increase in complexity and volume especially as more genomic testing is translated into clinical testing. Test findings are emerging with unknown meaning clinically. The pathologist will need to interpret test results both for the clinician to understand their meaning and to select the appropriate treatment for patient care.

Another factor impacting discordances is the delay in reporting cytogenetics results relative to flow cytometry and hematopathology results. What happens when these results do not correlate and diagnostic discordances are found? Should pathologists be reviewing the results from the other laboratories? Should there be changes in the testing, diagnosis or reporting processes to address this issue sooner than weeks after the initial testing, especially if there are potential effects on patient care? There are

many more questions to be addressed in the future when new testing methods and information are integrated with current testing processes and reporting methodologies.

Answers to these questions may explain why results indicate a combination of human and disease issues are contributing to diagnostic discordances in this research. If there are results that are not as anticipated, how is the pathologist responding to this information in reporting about the patient's disease status to the clinician? Some laboratories correlate their information as best as they can with the other laboratory findings and reports by issuing an interpretation that is consistent with these other laboratory findings. However, not all laboratories are reporting results in this manner. Some are just listing the raw test data and leaving the interpretation up to another laboratory or the clinician. Hence combinations of factors are contributing to the discordances seen.

Another human issue is lack of standardized terminology such as WHO diagnoses or SNOMED CT® when issuing a diagnosis. In some cases, the disease may behave like several diseases depending on which test results predominate. The pathologist may report more than one diagnosis leading to diagnostic discordance in providing the clinician with information about the disease prognosis, how it may respond to treatment and other factors related to patient care. Pathologists may also prefer different diagnostic terminologies and thus deviate from reporting standards. It is unclear why there is this variation, but this question could form the foundation for a future study. It was not the intent of this study to look at individual pathologist variations or individual pathologist practice.

Use of standard terminologies and practices provides for better communication of the patient's disease status especially among different healthcare institutions in the event that a patient receives treatment from or their testing is performed by several facilities. It is imperative that the clinician knows which disease is described in bone marrow reports. In conclusion there are a variety of both human and disease factors contributing to diagnostic discordances in this study.

Phase II: Categorization By Specific Etiologies

Diagnostic discordances can be categorized by their specific etiology as listed in Appendix C. Categorization by etiology provides information on the specific cause or etiology contributing to the discordance based upon information found in each laboratory report. Secondly, etiology also indicates where in the testing process issues contributing to diagnostic discordances occur including those as a result of an earlier issue. Consequently, those designing clinical decision support tools can target those identified areas impacting discordances and patient care the most.

Categorization of specific etiologies was performed for each laboratory pair as seen in Table 5. The table also indicates where each etiology impacted the testing process. The number of cases for each specific etiology is noted in each row, separated by columns for each laboratory pair and the category total. A case may have several different specific etiologies contributing to its discordance, so the sum of cases in parentheses may not add up to the case total.

Results from pre-analytical testing indicate no order issues with any laboratory pairs, but there were some collection issues. Specific etiologies with collection include poor specimen quality, inadequate volume, dilution with blood, uneven disease distribution at sample site, and sample preparation. The most commonly cited etiologies in the collection process for each laboratory comparison are: dilution with blood (hematopathology and flow cytometry), and uneven disease distribution at the sample site with one specimen showing disease while another does not (hematopathology and cytogenetics, flow cytometry and cytogenetics). The latter was the most prevalent specific etiology noted amongst all pre-analytical issues and the sixth most prevalent etiology overall.

Pre-analytical phase data indicate two main issues from this testing process section contributing to diagnostic discordances. The first, dilution with blood, may be seen as a result of poor sampling technique and be considered a human issue or it may be seen as a result of the disease process and be considered a disease issue. Human etiologies can

be remedied with additional training, but disease etiologies may have no intervention to prevent diagnostic discordances that are a facet of the disease process.

The second issue in this phase of testing is "uneven disease distribution." It is unclear if "uneven disease distribution" is a result of the disease process itself due to heterogeneous presentation of the disease, or due to a limitation in laboratory testing. As noted above, disease etiologies may not be remedied by an intervention. Many times, pre-analytical phase issues remain undetected until specimens are analyzed in the analytical phase of testing. If a specimen is unable to be obtained or of such poor quality at this stage, often the procedure and tests are canceled and reordered at a later time, thus preventing a discordance from happening later in the testing process. Otherwise, specimens are sent to laboratories for any analyses that can be performed on the specimens and for further assessment by each laboratory and ultimately, the pathologist. Therefore, detection of issues leading to potential discordance is left up to the clinician's judgment and a decision support tool would not provide any additional assistance in detection of issues during the pre-analytical phase of testing.

In the analytical phase of testing, specific etiologies with both the testing process and diagnosis/classification process contributing to the diagnostic discordances are seen. Data indicate most common etiologies for each laboratory pair include, "rare staining of other lineage cells" (hematopathology and flow cytometry, flow cytometry and cytogenetics), and "too few cells available for adequate analysis" (hematopathology and cytogenetics).

Each etiology can be a result of human processes or disease processes. Human etiologies may be affected by techniques in performing testing. In addition, having too few cells available for adequate analyses may be a consequence of pre-analytical collection issues, whether due to a human not collecting enough sample or due to the disease process resulting in a "dry tap." Disease manifestations may also impact the staining of other lineage cells. All of these can explain the data.

In detecting information process issues resulting in diagnostic discordances, one assumption of this research is highly regulated testing process and quality assurance measures prevent most testing process issues from progressing to discordances. This assumption is due to the regulations preventing the reporting of results which fail quality control measures. Consequently the data support this assumption, since this portion of the testing process indicated the fewest number of specific etiologies contributing to discordances. Detection of both, "rare staining of other lineage cells," and "too few cells to analyze," should be detected by human and quality control processes. At the point of detection of issues within the testing process, two immediate decisions could be made. One decision is results and information may be compromised resulting in the canceling of the test and thus preventing later discordances. Second is to let results and information flow to the pathologist for their consideration in the diagnosis and classification process. For those tests passing quality control processes for either etiology, this information is also allowed to flow to the pathologist for consideration and detection of any issues as part of the diagnosis and classification process.

Once issues are detected, assessment is made concerning their impact on the information process and if any interventions, such as clinical decision support tools, need to be implemented. For human etiologies, an appropriate intervention may be implementation of a training and education. This would address both collection and testing issues with specific human etiologies indicated by the data. For disease etiologies, assessment may be an intervention may not provide a remedy prior to the information reaching the pathologist in the diagnosis and classification process.

Data from the diagnosis and classification portion of the analytical phase of testing point to several specific etiologies which contribute to the diagnostic discordances seen. Among each laboratory comparison, the most common cited etiologies are "variability in where a case fits in a categorization bin," "cases fitting into multiple categories," "differences in opinion about which results most influence categorization," and "utilization of subjective decision making in the diagnosis." In addition, "unexpected

variability in disease manifestations," was noted as a common etiology in both hematopathology and cytogenetics, and flow cytometry and cytogenetics comparisons. Discussion of each etiology, examples, why they may occur, their detection, assessment and possible interventions follows.

The most prevalent etiology involves, "cases fitting into multiple categories." This etiology involves a disease process displaying features of more than one categorization or multiple categorizations that can be utilized to describe the disease process and the pathologist needing to choose one to report. With several laboratories performing testing, one of the pathologists may choose one diagnosis and another pathologist may choose another diagnosis even though they both adequately describe the disease process observed. This may be a limitation of the WHO categorizations. This etiology may also be noted when hematopathology categorizes an AML or ALL morphologically, while cytogenetics issues a more specific WHO cytogenetic categorization based upon the cytogenetic profile. This etiology may be seen in the case where morphologically, the disease is an "acute monocytic leukemia," but if a treatment based categorization is selected, the disease may be reported as an "acute myeloid leukemia and myelodysplastic syndrome, therapy related." The question becomes which diagnosis is most clinically relevant? How does the pathologist choose which diagnosis to report and which not to report? These cases may have a human etiology, such as pathologist decision making, or a disease etiology, such as how the disease expresses itself.

"Utilization of subjective decision making in the diagnosis," is the second most frequent etiology found with the data. This etiology is human based since it is based upon the subjective judgment of each pathologist. The pathologist interprets the laboratory data and reports the interpretation and supporting data to the clinician. However, several times flow cytometry or cytogenetics data were reported without a corresponding interpretation. This occurred despite enough information being available from the karyotype or markers to issue at least a general leukemia diagnosis, if not a specific WHO categorization. It is unclear why these test results are not being interpreted and a diagnosis issued in the report to the clinician. One may speculate that the clinician can

interpret these results. Questions arise in how these considerations impact their clinical decision making, especially if clinicians are unable to determine the specific type of leukemia presenting in their patient or in determining the appropriate latest treatment protocols for that specific diagnosis or findings.

"Unexpected variability in disease manifestations," is the third most common etiology noted with the data. This etiology is disease based as the disease does not behave as expected. One example of this etiology is noted when different laboratories report different diagnoses on the same specimen. Similar to "variability in where a case fits into a categorization bin," a disease may not behave or produce results as expected. Another example of this etiology can occur with a disease featuring characteristics of one or more diagnostics categories instead of behaving like a single disease process. A third example occurs with unexpected or unexplained test results that are not consistent with a disease category such as lymphoid markers expressed in certain acute myeloid leukemias or cytogenetics findings in which the clinical significance is unknown. This etiology may be considered a gray area, especially as cytogenetics and molecular diagnostics testing progresses. The expansion of information produced by these new testing methods may result in findings that are unclear as to their clinical significance. In these cases, pathologists make their best judgment in categorizing the disease process according to the WHO criteria. As new information is learned about these test results, they can be incorporated into pathological and clinical practice in providing patient care.

The fourth most prevalent etiology of "variability in where a case fits into a categorization bin," was noted in cases whereby one laboratory reported one WHO diagnostic category, while another laboratory's results favored another WHO diagnostic category. Another example of cases of discordance attributed to this etiology occurs when one laboratory issued a diagnosis and the other laboratory indicates no evidence of leukemia or disease. Cases where hematopathology or flow cytometry issued a diagnosis, but cytogenetics was able to issue a more specific diagnosis based upon a translocation are a third manifestation of this etiology. In all the cases categorized with

this etiology, these findings can be seen with human etiologies, disease etiologies or both.

In some cases, the exact cause explaining the differences in diagnoses leading to "variability in where a case fits into a categorization bin," is unclear without further research and root-cause analysis. Analytical testing issues such as the "rare staining of other lineage cells," may result in a laboratory in reporting an incorrect WHO categorization. "Too few cells available for analysis," could also result in laboratory not being able to detect disease, thus reporting no evidence of leukemia. In other cases, "too few cells," may result in the detection of disease, but it may be a more general categorization due to insufficient sample permitting additional testing or staining to be performed. When compared to a laboratory issuing a specific diagnosis, diagnostic discordance may be the effect seen. Lastly, limitations in laboratory testing process may result in a laboratory utilizing a method that is not sensitive enough to detect small quantities of disease such as minimal residual disease. When compared to a laboratory that does utilize more sensitive techniques, diagnostic discordance may arise between the diagnoses. Either way one laboratory is reporting the presence of disease while another is not. Further research needs to be done to determine the exact etiology of this phenomenon as well as determine its clinical significance and effect on patient care.

Lastly, the fifth most common etiology is differences in opinion about which results most influence categorization and thus is also a human etiology. It is also featured in cases whereby some laboratory findings indicate one WHO diagnostic category, while another laboratory's findings favor another WHO categorization. In deciding which categorization best describes the patient's disease process, a pathologist may choose to report more than one diagnosis to the clinician in some cases. In other cases, the pathologist may utilize a combination of two categorizations. For example, when a patient previously treated for a myelodysplastic syndrome presents with a new "de novo leukemia with monocytic features," does the hematopathologist designate the disease as "acute monocytic leukemia," "acute myeloid leukemia and myelodysplastic syndrome, therapy related, alkylating agent related," or "acute myeloid leukemia with multilineage

dysplasia with prior myelodysplastic syndrome?" These diagnostic considerations do not even include cytogenetics abnormalities and their subsequent categorizations. Which of these designations is most important to the clinician for patient care? How is the pathologist's diagnostic decision making affected when equivocal results are noted impacting the diagnostic categorization reported? In short, this etiology tends to add more uncertainty, rather than certainty to a diagnosis making it more likely for a discordance to occur.

Detection of issues contributing to discordances is often left to the pathologist in determining where in a multi-step decision making process, issues occurred and how they should be addressed. First, the test results produced by a laboratory may need to be assessed by the pathologist in the context of the suspected disease process. Often, the testing performed by the three laboratories is complex, requiring a pathologist to interpret the findings. If the test results and interpretation are consistent with and clearly indicate a diagnostic category, the pathologist communicates this diagnosis on the report with the findings.

Many times, where this diagnostic process is clear-cut, the laboratory report is issued independently of the other laboratories performing testing on specimens collected at the same time. Often a pathologist may fail to detect an issue contributing to discordance, especially if it occurs when they themselves are unaware of any discordance with their reported diagnosis and those reported by other laboratories analyzing specimens from the same initial bone marrow procedure. The first laboratory reporting results often does not have the advantage of looking up the other laboratories reporting on these specimens since their testing is often still in progress. However, with subsequent reports, the pathologist should be reviewing the reported findings and diagnoses from the other laboratories as they assimilate information available concerning a patient at the time of their diagnosis. If this occurred with the cases analyzed in this research, it would be expected that few to no diagnostic discordances would be noted. The data indicate otherwise, that there are issues with the information process.

Furthermore, the data indicate that the five most prevalent etiologies in this research project occur in the diagnosis and classification portion of the analytical phase of testing. Clearly, these etiologies are impacting the diagnostic discordances as noted with the data. These findings support the hypothesis that classification of diagnostic discordances provides information distinguishing which issues or etiologies impact patient care the most and diagnostic decision making. The data indicate that there is a need to detect issues with the information process, assess which ones are truly problematic or that can be prevented by an intervention, and implement an intervention when issues are detected.

For disease based etiologies, an intervention aimed at preventing diagnostic discordances may not be possible in cases where the discordances are a result of variations in the disease presentation. These etiologies may include cases fitting into multiple categories, unexpected variability in disease manifestations, and variability in where a case fits into a categorization bin. Despite the latest knowledge about the disease and testing processes, a disease may behave as one categorization when analyzed by one laboratory's methods and behave as another categorization during another laboratory's analyses. Test results may also be unexpected for a particular disease whether due to heterogeneous presentation or the exhibition of false positives or false negatives with testing methods.

For human etiologies, interventions may take the form of training and education depending on the etiologies contributing to the discordances seen. Etiologies of cases fitting into multiple categories, utilization of subjective decision making, variability in where a case fits into a categorization bin, or differences in opinion about which results most influence categorization, could all be due in part to pathologist decision making. In these cases, additional training or continuing education may be needed, such as when the 2008 WHO diagnostic categorizations began to be utilized in diagnostic decision making. Training may also address bone marrow collection issues or specimen processing techniques, resulting in insufficient or poor specimen quality, which in turn

contributes to not having enough information necessary to adequately diagnosis or subcategorize the patient's disease

As with the addition of any new laboratory test to the test menu, training and education will be needed for laboratory staff and pathologists on various aspects impacting the testing process. This training includes appropriate and inappropriate specimens and containers for the collection of samples, limitations and interferences of the testing process, proper interpretation of test results, exceptions in the testing process, and impact the results have on the diagnostic process. It is critical that all involved in the testing and information process be current on the latest changes, especially in light of all the variations with genetic and personalized medicine testing. It may no longer be enough to communicate a general diagnosis of "acute myeloid leukemia," to the clinician. With the information explosion, the clinician may need to know more specific details in order to provide the most appropriate treatment to the patient.

Another intervention to reduce discordance caused by human or disease etiologies is secondary review of a case by another pathologist. However, this intervention may not be possible at institutions with only one pathologist on staff. Informatics tools such as telepathology may be feasible in these cases. Disadvantages of this intervention include a more timely diagnosis and an increased cost in the diagnostic process due to secondary review labor costs.

Limitations of the testing process can occur quite frequently. Each testing methodology has a different sensitivity and specificity in detecting disease. One laboratory may not detect any disease with their methodology, while another detects disease and yet another laboratory detects even more specific markers of that disease, resulting in a specific disease categorization. Such results may not be an error per se, but may be how a particular disease presents itself. Often when discordances such as these are noted, the pathologist will correlate the findings with an interpretation or explanation in the pathology report so that the patient's disease status is clearly communicated to the clinician.

Despite the implementation of many of these interventions, discordances may still occur, especially in cases where an intervention is unable to prevent discordances from occurring. There is a need to alert pathologists of discordant results from other laboratories so that any errors are detected, pathologists can address them. The top etiologies revealed from the categorization of the discordances in this study should be the target of a clinical decision support alerting mechanism.

The assessment of discordances can be achieved via simple decision support rules-based decision support tools or more complex artificial intelligence based mechanisms. Rules could be built to check for known limitations of laboratory testing which contribute to diagnostic discordance. One known limitation is the diagnosis of "acute promyelocytic leukemia," by hematopathology compared to the diagnosis of acute promyelocytic leukemia with a t(15;17) translocation," by cytogenetics. A rule could exclude these two diagnoses from these two laboratories from being flagged as discordant. Similarly, if a standard encoded terminology were utilized, both of these diagnoses may be mapped to the same underlying disease process and consequently not be flagged as discordant as well.

Design of an appropriate alerting system should incorporate the appropriate level of physician alerting for optimal adoption and utilization. The decision support system needs to aid the pathologist's workflow and not disrupt it. Those cases that are indeed discordant, needing further assessment by a pathologist should be flagged as such so that the pathologist can correlate the final diagnostic report with any previous reports issued. Actual design of a clinical decision support tool is out of the scope of this research project. Further research would be needed with a larger number of cases from a number of medical centers to detect further patterns in discordances.

In conclusion, this method of categorization highlights the top etiologies contributing to diagnostic discordances in the dataset. The data indicate that most of these etiologies occur in the diagnosis and classification portion of the analytical phase of testing. Therefore, the analytical phase of testing should be targeted for intervention, rather than

the pre-analytical phase of testing. In some cases, especially those where the etiology is disease-based, an intervention that prevents discordance may not be possible. In cases with a human etiology, training and education may be the best intervention in preventing discordance. However with both etiologies, for the discordances which remain, it is critical that pathologists are alerted to them. Clinical decision support alerts are a key mechanism to provide that awareness so that pathologists can appropriately address them in their reports to the clinician.

Phase II: Categorization By Reporting Issues

The last method of diagnostic discordance categorization is categorization by reporting issues. Discordant cases are categorized by each laboratory pair to characterize how reporting issues contribute to diagnostic discordance. Both amended and unchanged reports were analyzed as they indicate which issues were noted and consequently changed, and which ones were missed, respectfully. It was anticipated some reports would be amended as some test results, such as special stains, are added to the report after it has been finalized.

Figure 14 indicates that of the reports with diagnostic discordances, 75% of reports were normal, lacking issues and amendments. The remaining 25% of reports were either amended (13%) or had undetected issues (12%). The latter percentage indicates that a number of issues were missed by pathologist review. 6% of the amended reports and 8% of the unchanged reports had typographical errors. These results indicate that a decision support tool alerting the pathologists to issues prior to a report being finalized is needed in preventing undetected issues. The hematology and cytogenetics laboratory pair may benefit the most from such a decision support tool since they had the largest number of detected and undetected issues. Further analysis such as root-cause analysis is needed to determine if this rate is simply due to limitations in laboratory testing or due to true discordances, as well as the impact on patient care.

All reports were further analyzed for typographical errors and subdivided by amended or changed reports, and unchanged reports. A typographical error is considered an issue

so these subdivisions can also be considered changed reports with typographical errors and unchanged reports with typographical issues to distinguish them from changed reports without typographical errors and unchanged reports without typographical errors, or normal reports. Results show that 14% total typographical errors evenly split between changed and unchanged reports. However, data show more unchanged reports with typographical errors than changed reports with typographical errors indicating many are undetected. Again, hematopathology and cytogenetics comparison led the other comparisons with the most typographical errors in the changed and unchanged reports. The findings indicate typographical errors are caught about half of the time. This could be easily remedied by utilizing an intervention such as a spell checker, a type of decision support tool, within report writing software**.**

Reports were also analyzed for contradictory information within the report. An example would include one laboratory indicating the presence of leukemia, while another indicating no presence of leukemia or normal findings. Less than 5% of all cases had contradictions, which is an encouraging finding. In fact, of those reports that were changed or amended, less than 1% had contradictory information and that was with the flow cytometry and cytogenetics pair. The unchanged reports with contradictions predominated with the hematopathology and cytogenetics comparison comprising the most cases again.

Detailed analyses were not performed on the exact etiology for each contradiction. Given contradictions were most common with hematopathology and cytogenetics, one explanation is limitations in laboratory testing where one lab detects disease and the other does not. Detection of discordant cases can be performed by a clinical decision support alert, but pathologists should assess contradictions because of their complexity. Alternately, report review by another pathologist after a discordance alert is another means to address contradiction. Nahkleh indicates although secondary review is favorable it is not always cost effective and delays the final report, which may affect patient care (48). However, single pathologist practices may not be able to use this method and clinical decision support tools would be more feasible.

To distinguish reports amended or changed due to additional test information from those due to other issues, reports were further subcategorized. The first sub-categorization is additional or delayed information added in the amendment indicating completed test results. Often in these cases, data indicate the final diagnosis remains unchanged. The second sub categorization is an amendment to make a correction in the report, indicating a true error. Data show that half of the amended reports were changed due to additional information, while the other half of changes were due to corrections indicating an error. A similar CAP Q-Probes study of anatomic pathology reports focused on studying only amendments with error, but found review of reports prior to their finalization reduced the number of amended reports (48). This suggests that report review prior to finalization is the best place for implementation of a clinical decision support tool in preventing discordances.

Categorization of reporting issues indicates that issues remain undetected through the phases of laboratory testing and are consequently found in the final report. These findings further emphasize the need for clinical decision support alerting pathologists to issues before the report is finalized. The data have shown that typographical errors, contradictions and other issues occur in addition to diagnostic discordance. The reporting categorization method supports the hypothesis that categorization of factors contributing to discordance allows one to distinguish those discordances requiring physician notification from those which do not. Amended or changed reports without issues are a normal part of pathology workflow and do not require alerting a pathologist of their occurrence. However, contradictions and some typographical errors, especially those which remain undetected require alerting of a pathologist for either correction or clarification in the final report. One alerting mechanism which addresses the common typographical error is use of a spell checker, while more complex issues can benefit from an alert detecting a discordance followed by pathologist assessment. Each of these aspects need to be considered in designing a clinical decision support tool designed to optimally provide alerts of diagnostic discordance.

Chapter 6

Conclusions and Future Work: Bone Marrow Biopsy Diagnostic Discordance Determination and Categorization

Conclusions

The first phase of research demonstrates diagnostic discordance, an indicator of potential information issues, can be successfully detected and measured among hematopathology, flow cytometry and cytogenetics laboratories. It is believed the first data about bone marrow biopsy diagnostic discordance rates, both lexically and semantically, are provided by this research. Cohen's Kappa statistical methodology provides the means for automatic detection and measurement of semantic discordance as well, an important consideration for clinical decision support tool design. Utilization of these methods of detecting and measuring diagnostic discordance is also valuable in quality improvement initiatives, determining what levels of discordance are significant, standards development, clinician decision making, and ultimately, patient care.

In the second phase, categorization of discordances provided additional information necessary in distinguishing those discordances which are limitations in the testing process from those which are truly problematic and have a higher impact on patient care. Assessment is also made in delineating which discordances warrant pathologist alerting. Furthermore, each categorization method provided vital information about which interventions are best utilized in preventing discordances, how discordances are best addressed, and where interventions should be optimally positioned in the testing process.

Categorization by level or how diagnoses are related provided an indicator of which diagnoses are more closely related and likely due to a laboratory limitation versus those which are not so related and more likely an indicator of a more significant discordance. Categorization by limiting factors specifies both laboratory and non laboratory factors which contribute to discordances in prospectively addressing the origin of issues

resulting in discordances. Furthermore, categorization by etiology, human or disease, as well as specific causes also provide guidance as to what interventions may be best utilized in preventing or alerting pathologists of discordances. Categorization by where issues occur in the testing process point to where interventions are best placed into practice. Often, human origins of discordance are best addressed with training or education, while discordances rooted in disease etiologies may not be prevented but can be addressed by a pathologist who is appropriately alerted to their presence. Although each categorization provides a perspective of what contributes to discordances, the resulting information is valuable in tailoring a clinical decision support alert or other intervention to the complexity of pathology practice in ensuring quality information is produced for patient care. Categorization by reporting further indicates that clinical decision support alerts to diagnostic discordance and other reporting issues are needed prior to report finalization to prevent undetected issues from potentially impacting patient care.

The research also revealed several problematic areas impacting the information process. One is use of non-standard WHO diagnostic terminologies, which was shown to be a barrier in achieving semantic agreement among reported diagnoses. Use of standardized terminologies also can form the backbone of a clinical decision support tool comparing diagnoses based upon the disease concept. The data also indicate that information complexity is growing in laboratory medicine and pathology. This is only expected to increase as more genetic and personalized testing occurs. Decision support is needed not only by pathologists in performing the appropriate interpretation of test results, but also clinicians in selecting the appropriate patient care.

Future Work

Much more can be done to expand this research. A third phase could use discordant case data to compose case vignettes presented to clinicians to study the impact discordant cases have on clinical decision making. Survey questions could probe the information and sufficiency of information reported in determining their effect on

clinical diagnosis and patient treatment. In cases where insufficient information is reported, questions could be asked about what information is needed and what sources are consulted in practice.

Another avenue related to this research is to study the impact diagnostic discordance has on patient care. Raab and others have performed studies on classifying discordant cases by their clinical significance, patient impact and outcomes. In addition it may be interesting to study the financial impact on both providers and patients due to diagnostic discordances. Such results could shape financial decision support tools utilized by hospital administration.

Application of the diagnostic discordance measurement methods at other institutions would provide metrics applicable to pathology and healthcare in general. Data may indicate how robust these methods are, how applicable they are to other healthcare settings, the levels of discordance at other institutions, and what the categorizations reveal about discordance patterns at other institutions. Widespread studies may provide the data necessary for standards development in measuring discordance, categorizing discordance, decision tool development, and determination of appropriate clinical, statistical or financial decision points. Once standards are established and adopted by their respective professional communities, these and other metrics may be incorporated into quality measures such as CAP surveys or Q-Probes studies. Standards may also be established in the informatics community as part of national standards organizations or even via governance.

In conclusion, the diagnosis of acute myeloid and acute lymphoid leukemia via bone marrow biopsy testing by the cytogenetics, flow cytometry and hematopathology laboratories is a complex, information laden process. When issues in this process arise, diagnostic discordance is an indicator which can be utilized to alert the pathologist to those issues that are problematic. The research has shown that diagnostic discordance can be successfully detected and measured with Cohen's Kappa statistic. However, effective clinical decision support systems need to incorporate a means to delineate true

discordant cases from those that result from a limitation of laboratory testing, as shown by categorization of discordances. These considerations form the foundation for the development of a clinical decision support tool that can appropriately alert the pathologist and clinician to potential issues with the bone marrow report so that they can be addressed and rectified.

Bibliography

1. Kohn Linda T, CJM, Donaldson Molla S., editors. To Err is Human: Building A Safer Health System. Institute of Medicine HCS, editor. Washington, DC: National Academy Press; 2000. 312 p.

2. Raab SS, Nakhleh RE, Ruby SG. Patient safety in anatomic pathology: Measuring discrepancy frequencies and causes. Arch Pathol Lab Med. 2005 Apr;129(4):459-66.

3. Jacobs P. Discrepant bone marrow aspirations in leukaemia. The Lancet. 1977 August 13, 1977;2:355-6.

4. Hann IM, Jones PHM, Evans DIK. Discrepancy of bone marrow aspirations in acute lymphoblastic leukaemia in relapse. The Lancet. 1977 June 4, 1977;1:1215-6.

5. Golembe B, Ramsay NKC, McKenna R, Nesbit ME, Krivit W. Localized bone marrow relapse in acute lymphoblastic leukemia. Med Pediatr Oncol. 1979;6:229-34.

6. Trask BJ. Human cytogenetics: 46 chromosomes, 46 years and counting. Nat Rev Genet. 2002 Oct;3(10):769-78.

7. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292-302.

8. Bennett JM, Catovsky D, Daniel MT. Proposals for the classification of the acute leukaemias: French-American-British cooperative group. Br J Haematol. 1976;33:451.

9. Foucar, K, American Society of Clinical Pathologists. Bone marrow pathology. 2nd ed. Chicago: ASCP Press; 2001. 704 p.

10. Medicare, Medicaid and CLIA programs; regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA)--HCFA. Final rule with comment period. UNITED STATES: ; 1992. 7002 p.

11. Jones BA, Novis DA. Follow-up of abnormal gynecologic cytology: A College of American Pathologists Q-probes study of 16132 cases from 306 laboratories. Arch Pathol Lab Med. 2000 May;124(5):665-71.

12. Raab SS, Grzybicki DM, Janosky JE, Zarbo RJ, Meier FA, Jensen C, Geyer SJ. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. Cancer. 2005 Nov 15;104(10):2205-13.

13. Clary KM, Silverman JF, Liu Y, Sturgis CD, Grzybicki DM, Mahood LK, Raab SS. Cytohistologic discrepancies: A means to improve pathology practice and patient outcomes. Am J Clin Pathol. 2002 Apr;117(4):567-73.

14. Laboratory Accreditation Checklists, Hematology & Coagulation Checklist. Northfield, IL: College of American Pathologists' Laboratory Accreditation Program; 2006 October 31, 2006

15. Laboratory Accreditation Checklists, Flow Cytometry Checklist. Northfield, IL: College of American Pathologists' Laboratory Accreditation Program; 2006 October 31, 2006

16. Laboratory Accreditation Checklists, Cytogenetics Checklist. Northfield, IL: College of American Pathologists' Laboratory Accreditation Program; 2006 December 12, 2006

17. Laboratory Accreditation Checklists, Anatomic Pathology Checklist. Northfield, IL: College of American Pathologists' Laboratory Accreditation Program; 2006 December 12, 2006

18. Riley R, Ben-Ezra J, Pavot D, Forysthe R, Massey D, Wright L, Smith E, Hogan T. An Illustrated Guide to Performing the Bone Marrow Aspiration and Biopsy. 09/25

19. Bain BJ. Bone marrow trephine biopsy. J Clin Pathol. 2001;54(10):737-42.

20. Lofsness Karen G. SEM. Chapter 29, Preparation and Evaluation of Bone Marrow In: Lotspeich-Steininger Cheryl A., S.-M.E.A.,Koepke John A., editor. Clinical Hematology: Principles, Procedures, and Correlations. New York: JB Lippincott Company; 1992; p. 366-78.

21. Orfao A, Schmitz G, Brando B, Ruiz-Arguelles A, Basso G, Braylan R, Rothe G, Lacombe F, Lanza F, Papa S, Lucio P, San Miguel JF. Clinically useful information provided by the flow cytometric immunophenotyping of hematological malignancies: Current status and future directions. Clin Chem. 1999;45(10):1708-17.

22. Braylan RC, Atwater SK, Diamond L, Hassett JM, Johnson M, Kidd PG, Leith C, Nguyen D. U.S.-Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: Data reporting. Cytometry. 1997;30(5):245-8.

23. Cornacoff R. Direct and Short-Term Procedure for Harvesting Bone Marrow Chromosomes. 07/30

24. Frizzera G, Wu CD, Inghirami G. The usefulness of immunophenotypic and genotypic studies in the diagnosis and classification of hematopoietic and lymphoid neoplasms. an update. Am J Clin Pathol. 1999;111(1):S13-39.

25. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA, Bloomfield CD. The World Health Organization classification of neoplastic

diseases of the hematopoietic and lymphoid tissues. report of the clinical advisory committee meeting, airlie house, virginia, november, 1997. Ann Oncol. 1999;10(12):1419-32.

26. Jaffe ES, Harris NL, Diebold J, Muller-Hermelink HK. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues. A progress report. Am J Clin Pathol. 1999;111(1):S8-12.

27. LeClair SJ, Rodak BF. The new WHO nomenclature: Lymphoid neoplasms. Clin Lab Sci. 2002 Winter 2002;15(1):55-9.

28. Rodak BF, LeClair SJ. The new WHO nomenclature: Introduction and myeloid neoplasms. Clin Lab Sci. 2002 Winter 2002;15(1):44-54.

29. Chan JK. The new World Health Organization classification of lymphomas: The past, the present and the future. Hematol Oncol. 2001;19(4):129-50.

30. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA, Bloomfield CD. The World Health Organization classification of neoplasms of the hematopoietic and lymphoid tissues: Report of the clinical advisory committee meeting--airlie house, virginia, november, 1997. Hematol J. 2000;1(1):53-66.

31. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA, Bloomfield CD. The World Health Organization classification of neoplastic diseases of the haematopoietic and lymphoid tissues: Report of the clinical advisory committee meeting, airlie house, virginia, november 1997. Histopathology. 2000;36(1):69-86.

32. Laboratory Accreditation Checklists, Laboratory General Checklist. Northfield, IL: College of American Pathologists' Laboratory Accreditation Program; 2006 October 31, 2006

33. Comprehensive Accreditation Manual for Laboratory and Point of Care Testing (CAMLAB) [Internet]. Joint Commission on the Accreditation of Healthcare Organizations

34. Hertenstein B, Heil G, Heimpel H. Allogenic bone marrow transplantation or chemotherapy for patients with acute myeloid leukemia in first complete remission: A decision analysis approach. Ann Hematol. 1996;72:223-30.

35. Jones BA, Novis DA. Cervical biopsy-cytology correlation. A College of American Pathologists Q-probes study of 22 439 correlations in 348 laboratories. Arch Pathol Lab Med. 1996;120(696255576):523-31.

36. Kremer M, Spitzer M, Mandl-Weber S, Stecker K, Schmidt B, Hofler H, Quintanilla-Martinez L, Fend F. Discordant bone marrow involvement in diffuse large B-cell lymphoma: Comparative molecular analysis reveals a heterogeneous group of disorders. Lab Invest

37. Onciu M, Berrak SG, Medeiros LJ, Katz RL, Huh YO. Discrepancies in the immunophenotype of lymphoma cells in samples obtained simultaneously from different anatomic sites. Am J Clin Pathol. 2002 Apr;117(4):644-50.

38. Cramer S. Judging mistakes in pathology--res ipse non loquitur. Arch Pathol Lab Med. 2006 October 2006;130:1430.

39. Zarbo RJ, Meier FA, Raab SS. Error detection in anatomic pathology. Arch Pathol Lab Med. 2005 Oct;129(10):1237-45.

40. Foucar K. Bone marrow examination in the diagnosis of acute and chronic leukemias. Hematol Oncol Clin North Am. 1988 Dec;2(4):567-84.

41. Peterson LC, Agosti SJ, Hoyer JD. Protocol for the examination of specimens from patients with hematopoietic neoplasms of the bone marrow: A basis for checklists. Arch Pathol Lab Med. 2002;126(9):1050-6.

42. Medscape General Medicine [Internet]Vol. 8, No. 4,

43. Coffin C. Pediatric surgical pathology, pitfalls and strategies for error prevention. Arch Pathol Lab Med. 2006 May 2006;130:610.

44. Boone DJ. Is it safe to have a laboratory test? CMPT Connections Newsletter. 2003 Spring 2003;7(1):4.

45. Renshaw AA. Improving consensus on the cause of cytologic-histologic discrepancy: Searching for quality assurance methods that work. Am J Clin Pathol. 2006 Dec;126(6):831-2.

46. Vrbin CM, Grzybicki DM, Zaleski MS, Raab SS. Variability in cytologic-histologic correlation practices and implications for patient safety. Arch Pathol Lab Med. 2005 Jul;129(7):893-8.

47. Raab SS, Tworek JA, Souers R, Zarbo RJ. The value of monitoring frozen section-permanent section correlation data over time. Arch Pathol Lab Med. 2006 Mar;130(3):337-42.

48. 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 Apr;122(4):303-9.

49. Cioc AM, Julius CJ, Proca DM, Tranovich VL, Keyhani-Rofagha S. Cervical biopsy/cytology correlation data can be collected prospectively and shared clinically. Diagn Cytopathol. 2002 Jan;26(1):49-52.

50. 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 Nov;126(5):736-9.

51. Renshaw AA, Gould EW. Measuring errors in surgical pathology in real-life practice: Defining what does and does not matter. Am J Clin Pathol. 2007 Jan;127(1):144-52.

52. Nakhleh RE. Error reduction in surgical pathology. Arch Pathol Lab Med. 2006 May;130(5):630-2.

53. Check W. Going with the flow in hematopathology. CAP Today [Internet]. 2004 November 2004 [cited October 14, 2007]

54. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. J Biomed Inform. 2004 Jun;37(3):193-204.

55. Bagg A. Lineage ambiguity, infidelity, and promiscuity in immunophenotypically complex acute leukemias: Genetic and morphologic correlates. Am J Clin Pathol. 2007 Oct;128(4):545-8.

56. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007 Mar-Apr;14(2):141-5.

57. Paxton A. Playing by the rules with expert software. CAP Today [Internet]. 2004 November 2004 [cited October 14, 2007]

58. Diamond LW, Mishka VG, Seal AH, Nguyen DT. Are normative expert systems appropriate for diagnostic pathology? J Am Med Inform Assoc. 1995;2(2):85-93.

59. Ratei R, Karawajew L, Lacombe F, Jagoda K, Del Poeta G, Kraan J, De Santiago M, Kappelmayer J, Bjorklund E, Ludwig WD, Gratama JW, Orfao A, European Working Group of Clinical Cell Analysis. Discriminant function analysis as decision support system for the diagnosis of acute leukemia with a minimal four color screening panel and multiparameter flow cytometry immunophenotyping. Leukemia. 2007 Jun;21(6):1204-11.

60. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. BMJ. 2005 Apr 2;330(7494):765.

61. Nykaenen P. Decision Support Systems from a Health Informatics Perspective]. Tampere, Finland: University of Tampere; 2000

62. Schwarzinger I. The hama-COM laboratory data management system (LDS): Tailor-made computerization of the individual working procedures arising in a hematology laboratory with a high throughput of pathologic samples. Ann Hematol. 1997 Jul-Aug;75(1-2):47-54.

63. Asare AL, Ellis JS, Caldwell CW. A decision-support system for flow cytometry immunophenotyping. Am J Clin Pathol. 2002;118(4):567-73.

64. van Oosterhout EM, Talmon JL, de Clercq PA, Schouten HC, Tange HJ, Hasman A. Three-layer model for the design of a protocol support system. Int J Med Inform. 2005 Mar;74(2-4):101-10.

65. Cerner Millenium Support Guide-PathNet Anatomic Pathology Product Overview. Cerner Corporation; 2003 April 2003

66. Ohori NP, Schoedel KE, Rajendiran S. Cytologic-histologic correlation of nongynecologic cytopathology cases: Separation of determinate from indeterminate cytologic diagnoses for analysis and monitoring of laboratory performance. Diagn Cytopathol. 2003 Jan;28(1):28-34.

67. Leukemia: Introduction. University of Virginia, Department of Pathology

68. Connelly D. University of Minnesota Cancer Center Nucleus Database System Users Guide. ; 2002.

69. Lewis RJ. Reliability and Validity: Meaning and Measurement. May 17, 1999

70. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 March 1977;33(1):159-74.

71. Birkhead BG, Salt D, Jackson RRP. Reliability and consistency of bone marrow assessments of complete remission in acute myelogenous leukaemia. Hematol Oncol. 1987;5:139-45.

72. Olmsted A. Calculating kappa measures of agreement and standard errors using SAS software: some tricks and tips [Internet].

73. Liu H, Hays R. Measurement of interrater agreement: A SAS/IML macro kappa procedure for handling incomplete data. In: Francis-Minor, S. L. S. K. a. C., editor. SAS Users Group International Conference ; 1999 April 11-14, 1999Miami Beach, Florida, USA: SAS Institute, Inc; 1999. p. 1620-5.

74. SAS OnlineDoc (Registered Trademark), Version 8 [Internet]. Cary, NC: SAS Institute, Inc.

75. Sandhaus LM, Meyer P. How useful are CBC and reticulocyte reports to clinicians? Am J Clin Pathol. 2002 Nov;118(5):787-93.

76. Raab SS, Grzybicki DM, Zarbo RJ, Meier FA, Geyer SJ, Jensen C. Anatomic pathology databases and patient safety. Arch Pathol Lab Med. 2005 Oct;129(10):1246-51.

77. McCowan IA, Moore DC, Nguyen AN, Bowman RV, Clarke BE, Duhig EE, Fry MJ. Collection of cancer stage data by classifying free-text medical reports. J Am Med Inform Assoc. 2007 Nov-Dec;14(6):736-45.

78. Cancer Protocols and Checklists: Bone Marrow. Northfield, IL: College of American Pathologists; 2005 [cited May 16, 2008]. Available from: http://www.cap.org/apps/docs/committees/cancer/cancer_protocols/2005/bonemarrow05_ckw.pdf

79. Reason J. Human error: Models and management. BMJ. 2000 Mar 18;320(7237):768-70.

80. Diamond LW, Nguyen DT, Jouault H, Imbert M, Sultan C. An expert system for the interpretation of flow cytometric immunophenotyping data. J Clin Comput. 1994;22(2):50-8.

Appendices

Appendix A

Selected WHO Hematological Neoplasm Classifications [Harris, 1999 #192]

1. Myeloid Neoplasms
- a. Myeloproliferative diseases
- b. Myelodysplastic/myeloproliferative diseases
  - i. Chronic myelomonocytic leukemia
  - ii. Atypical chronic myelogenous leukemia
  - iii. Juvenile myelomonocytic leukemia
- c. Myelodysplastic syndromes
- d. Acute myeloid leukemias*
  - i. AMLs with recurrent cytogenetic translocations
  - ii. AML with t(8;21)(q22;q22), AML1(CBF-alpha)/ETO
  - iii. Acute promyelocytic leukemia (AML with t(15;17)(q22;q11-12) and variants, PML/RAR-alpha)
  - iv. AML with abnormal bone marrow eosinophils (inv(16)(p13q22) or t(16;16)(p13;q11), CBFß/MYH11X)
  - v. AML with 11q23 (MLL) abnormalities
  - vi. AML with multilineage dysplasia
    1. With prior myelodysplastic syndrome
    2. Without prior myelodysplastic syndrome
  - vii. AML and myelodysplastic syndromes, therapy-related
    1. Alkylating agent–related
    2. Epipodophyllotoxin-related (some may be lymphoid)
    3. Other types
  - viii. AML not otherwise categorized
    1. AML minimally differentiated
    2. AML without maturation
    3. AML with maturation
    4. Acute myelomonocytic leukemia
    5. Acute monocytic leukemia
    6. Acute erythroid leukemia
    7. Acute megakaryocytic leukemia
    8. Acute basophilic leukemia
    9. Acute panmyelosis with myelofibrosis
    10. Acute biphenotypic leukemias

2. Lymphoid Neoplasms
- a. B-Cell neoplasms
  - i. Precursor B-cell neoplasm
  - ii. Precursor B-lymphoblastic leukemia/lymphoma (precursor B-cell acute lymphoblastic leukemia)
- b. T-cell and NK-cell neoplasms

i. Precursor T-cell neoplasm
ii. Precursor T-lymphoblastic lymphoma/leukemia (precursor T-cell acute lymphoblastic leukemia)

3. Acute Lymphoid Leukemias

i. Precursor B-cell acute lymphoblastic leukemia (cytogenetic subgroups)
    1. t(9;22)(a34;q11); BCR/ABL
    2. t(v;11q23); MLL rearranged
    3. t(1;19)(q23;p13) E2A/PBX1
    4. t(12;21)(p12;q22) ETV/CBF-alpha
ii. Precursor T-cell acute lymphoblastic leukemia
iii. Burkitt-cell leukemia

Appendix B

Cancer Case Counts for Hematological Neoplasia with the following ICD-O Codes (University of Minnesota Cancer Center Nucleus Database (NDB) query performed on December 19, 2003)

Cancer Case Counts

Hematological Neoplasia	ICD-O Code	2000	2001	2002	2003
Ac. myelomonocytic leuk. w abn. mar. eosinophils	9871/3			1	1
Acute biphenotypic leukemia	9805/3		2	2	3
Acute leukemia, NOS	9801/3	1	1		
Acute megakaryoblastic leukemia	9910/3	1		1	
Acute monocytic leukemia	9891/3	10	3	3	4
Acute myeloid leuk. with multilineage dysplasia	9895/3			5	13
Acute myeloid leukemia with maturation	9874/3	3	2	2	7
Acute myeloid leukemia without maturation	9873/3	5	3	1	4
Acute myeloid leukemia, M6 type	9840/3		1	2	
Acute myelomonocytic leukemia	9867/3	7	3	3	8
Acute promyelocytic leuk.,t(15;17)(q22;q11-12)	9866/3	1	1	2	
Precursor cell lymphoblastic leukemia, NOS	9835/3	24	6		2
Precursor T-cell lymphoblastic leukemia	9837/3		2	5	1
Precursor T-cell lymphoblastic lymphoma	9729/3		1	1	3
Therapy-related acute myeloid leukemia, NOS	9920/3		3	1	1

Appendix C

Diagnostic Discordance Classification by Specific Etiology

Cases are analyzed to determine the etiology of the discordance or contributing issues to the discordance. Issues are divided by where they occur in the testing process, followed by the specific issue and then some issues are subdivided by their human or disease etiology. Explanations of each issue are listed. "*" designated issues are from Kathy Foucar's Bone Marrow text, while the other issues originate from the author.

1. Pre-analytical issues (occurring during the ordering and/or collection processes)
    a. Poor sample quality
        i. Fibrotic specimen (disease)
            1. e.g. The patient's marrow is so fibrotic that an insufficient portion is aspirated, the portion aspirated is fibrotic, and may not be representative of the marrow.
        ii. Inadequate volume
            1. Human (not enough sample collected by doctor due to judgment or technique issues)
                a. e.g. The person collecting the marrow may have poor technique and not collect enough marrow for all testing. The sample may not be representative of the marrow if sampled too deep or shallow or filled with blood.
            2. Disease (not enough sample to be collected/not enough bone marrow being made or unable to collect enough sample since marrow too fibrotic)
                a. e.g. The patient's marrow is so fibrotic that an insufficient portion is aspirated.
        iii. Diluted with blood*
            1. Human (technique issues-sampled bloody area/wrong site sampled/not sampled deep enough, etc)
                a. e.g. The person collecting the marrow may have poor technique and collect more blood than marrow for testing. The sample may not be representative of the marrow if sampled too deep or shallow or filled with blood.
            2. Disease (marrow is bloody, dilute, and is truly representative of patient's state)
                a. e.g. The marrow may be hypocellular or diluted with blood and be truly representative of the patient's state.
    b. Uneven disease distribution at sample site
        i. Disease (one specimen shows disease where another specimen does not)

a. e.g. The specimens collected may have been from different portions of the marrow and one portion may contain disease, while another does not.

c. Sample Preparation (crush prep versus slides)
    i. Not enough sample preparations of each type
        1. Human (technologist does not make enough slides)
            a. e.g. The person collecting the marrow may not have collected enough to make enough slides or crush preps.
        2. Disease (technologist can't make enough slides due to insufficient specimen, poor specimen quality, clotted specimen, etc)
            a. e.g. The specimen collected may be insufficient in volume or quality to make enough crush preps or slides. This may occur with fibrotic specimens or hypocellular marrow.
    ii. Each sample type (crush/slide) shows different numbers or morphology
        1. Human (technologist is unable to make a good slide or preparation due to poor technique (slide pulling), not using enough specimen, etc)
            a. e.g. The person making slides or crush preps may have poor technique and make a slide that is too thin or too thick or pull a slide that has an uneven distribution of cells.
        2. Disease (slides made with good technique still are not good due to clumping/uneven distribution of sample, fibrotic marrow, or no marrow to work with. May have too few cells or too many-leukocytosis, with cells overlapping, etc, too thick or too thin of a slide)
            a. e.g. The person making slides or crush preps may have good technique, but the slide may be too thin, too thick, have an uneven distribution of cells, or clumping due to the patient's disease status.

2. <u>Analytical Issues</u>
    a. During Testing
        i. Rare staining of other lineage cells* (disease)
            1. e.g. The disease may cause antigens to be abnormally expressed on cell lines in which they normally are not.
        ii. Loss of expression with processing*
            1. Disease
                a. e.g. The disease may cause antigens to have different expressions than normal. Cells may be

more fragile or samples more fibrotic leading to a loss of antigens with the testing process.

2. Human
    a. e.g. The testing process may cause a loss of antigens during processing of samples. There also may be human error or technique that causes cells to be washed off slides or out of tubes.

iii. Too few cells available for adequate analysis*
1. Disease
    a. e.g. In concert with a poor quality specimen or diluted specimen as described above, or just in collection of hypocellular marrow, there may be too few of a number of cells collected for analysis. This includes not enough cells to achieve an adequate culture for cytogenetics, not enough cells to test all the requested antibodies via flow cytometry or not enough slides pulled for hematopathology.
2. Human
    a. e.g. In concert with a poor quality specimen or diluted specimen as described above, or just in not collecting enough marrow, there may be too few of a number of cells collected for analysis. This includes not enough cells to achieve an adequate culture for cytogenetics, not enough cells to test all the requested antibodies via flow cytometry or not enough slides pulled for hematopathology.

iv. Loss of architecture in process, damaged specimen
1. Disease
    a. e.g. In a hypocellular marrow or fibrotic marrow, the specimen itself could be damaged by the collection process or specimen processing as part of the testing process.
2. Human
    a. e.g. A specimen could be damaged manually as part of the processing in the testing process, such as when reagents or stains interact with cells or when slides are pulled.

v. Detection of chromosomal abnormalities in normal patients which is of unknown significance* (disease)
1. e.g. This occurred with specimens that were negative for disease via hematopathology or flow cytometry, but had cytogenetics abnormalities. Cytogenetic abnormalities are being discovered daily, as well as new methods of detecting them.

vi. Inappropriate utilization of testing and quality assurance practices* (human)
    1. e.g. This would occur if controls were not used or used incorrectly, and results were reported when controls were out of acceptable range. It is assumed that this did not occur since it is unethical practice and in violation of regulatory requirements.

b. During Diagnosis/Classification of Disease Process
    i. Variation in morphological diagnosis* (disease)
        1. e.g. This occurs in cases where the morphological findings aren't what is typically seen for a particular diagnostic category. This may occur when an AML and MDS therapy related also has features of an Acute monoblastic leukemia.
    ii. Categorizations prone to subjective variability* (human)
        1. e.g. This was utilized mostly to describe cases where the information was present in the report, but the diagnosis was not subcategorized. It was most commonly seen with flow cytometry where the markers were listed without a categorization noted.
    iii. Indistinct category boundaries* (Disease-ie MDS evolving to AML)
        1. e.g. This occurs in cases where it is nearly impossible to tell the difference between related diagnoses morphologically. It is difficult to tell the difference between a MDS with a blast count of 19% and an AML with a blast count of 20%.
    iv. Variability in where a case fits in a categorization bin (ie cytogenetically like one disease, morphologically like another)*
        1. Disease
            a. e.g. This occurs in cases where there is a mismatch of diagnoses issued by each lab and the cause is due to the disease. The disease may look more like one type of leukemia via morphology and express differing flow markers.
        2. Human
            a. e.g. This occurs in cases where there is a mismatch of diagnoses issued by each lab and the cause is due to a human issue like categorization. This may occur when morphology issues a diagnosis and can't see the cytogenetics, when cytogenetics issues a more specific diagnosis based upon their findings.
    v. Cases fitting into multiple categories*
        1. Disease

        a. e.g. This occurs in cases where the disease presents itself and the findings can be one of several leukemia types.
    2. Human e.g.
        a. e.g. This occurs in cases where the findings listed in the report could occur in more than one leukemia type. A common case of this is when flow cytometry did not subcategorize or issue a diagnosis, but listed just the flow markers.

vi. Differences in opinion about which results most influence categorization* (human)
    1. e.g. This occurs when a disease can be categorized into one of several leukemia types and the type chosen by one lab does not match that of the other lab. Do the cytogenetics results trump the morphological findings?

vii. Utilization of subjective decision making in diagnosis* (human)
    1. e.g. This occurs when the pathologists decides not to issue a more specific diagnosis even though the information is present in the report.

viii. Categories influencing diagnostic process* (human)
    1. e.g. This occurs when a case requires additional information to categorize it that is not a result of the laboratory testing (ie previous diagnosis or treatment information)

ix. "Changing treatment options, which will affect classification to the extent that measures of "clinical relevance" influence classification."* (human)
    1. e.g. This occurs when a case is categorized as one disease, but after treatment it behaves more like another disease process (ie ALL treatment working on what appears to be an AML).

x. Impact of new diagnostic testing with old categories* (human)
    1. e.g. This occurs when a new cytogenetic marker arises that is specific for leukemia, but no category exists to reflect that diagnosis. This can also occur when FISH or molecular diagnostics testing indicate the presence of markers or proteins associated with one disease, but morphologically the disease looks like something else.

xi. Political and social factors in each department (acceptance of classifications)* (human)
    1. e.g. This occurs when the pathologists utilize older classification methods like the FAB, even though the gold standard is the WHO.

xii. Unexpected variability in disease manifestations (antigen expression, response to treatment, etc)* (disease)

xiii. e.g. This occurs when the case exhibits characteristics for more than one categorization, exhibits characteristics that place it in one category, but does not exhibit all the characteristics of that category or otherwise has atypical findings. This also occurs when laboratory indicates there is disease and another does not find evidence of the disease with their methodology.

xiv. "Tacit knowledge influences"* (human)
    1. e.g. This occurs when the pathologists cannot decide on a diagnosis or modify the diagnostic categorization to include other findings.

xv. "Reliance on surrogates" instead of real disease markers*
    1. Disease
        a. e.g. This occurs when the testing for a disease is dependent on a protein, staining or other indicator that is indicative of a disease, but may not always reflect the true disease state.
    2. Human
        a. e.g. This occurs when the pathologists diagnose based on descriptors or characteristics of the disease, even though they may not have definitive indicators of the disease.