EDUCATIONAL COMMENTARY – CRITERIA FOR AUTOMATED REFLEX TO MANUAL DIFFERENTIAL

Educational commentary is provided through our affiliation with the American Society for Clinical Pathology (ASCP). To obtain FREE CME/CMLE credits, click on Earn CE Credits under Continuing Education on the left side of the screen.

Learning Objectives

On completion of this exercise, the participant should be able to

- describe the appropriate application of the International Society for Laboratory Hematology guidelines for manual differential smear review;
- describe key considerations in developing specific local guidelines for differential review; and
- perform a validation study to document the criteria for performance of review with the local instrumentation and patient population.

“I am rarely happier than when spending an entire day programming my computer to perform automatically a task that it would otherwise take me a good ten seconds to do by hand.” Douglas Adams

Differential Review: The Historical Perspective

In the course of the last generation, technological innovations have swept over medical laboratories like a tsunami. Imagine if you will a time when complete blood cell counts (CBCs) were run on instruments that could not perform automated differential cell counts, and printed results with noisy sled printers that spit out multipart carbon paper report forms for distribution to physicians. That was the 1970s. (Apple was formed in 1976; the first IBM personal computer was sold in 1981.) Primitive laboratory information systems (LISs) had been developed in the late 1960s, but didn’t really invade laboratories in full force until the 1980s. Those early LIS products (e.g., Sunquest, Cerner, Rubicon) were rudimentary systems with non-graphical operating systems.

At this time, the first instrument interfaces evolved, crawling from primordial LIS departments into the laboratory. These were initially unidirectional (instruments to LIS); however, as analyzers became capable of performing multiple tests (CBC / differential / reticulocyte), bidirectional interfaces that performed a host query were developed, enabling clinicians to order only the tests they wanted. The next leap in interface development came with auto-verification. This typically operates by sending data from the instrument to a software package (middleware) that tests the results against a set of rules, and passes a Yes or No flag to the LIS, where results are automatically reported if appropriate. This has led to enormous gains in efficiency, with reduced turnaround time for normal results and more focused technical attention to abnormal cases.1-3 The downside is that if the interfaces are not working, life can become challenging. (Lab Tech: “Close the pod bay door, HAL.” Instrument: “I’m sorry, Dave, I can’t do that.” Mayhem ensues.) Hence, it is necessary to plan for computer downtime.
Before computerization, criteria for slide review had to be simple, because technicians reviewed the results for every case to decide whether slide review was required. Complicated rules were not easy to remember or to apply. In the modern world, however, middleware can apply rules based on patient age, sex, client, tests ordered, and previous results (delta checking). At the author’s laboratory, a set of CBC results must pass 84 rules before the results will auto-validate; in the middleware, those rules expand into 399 separate rules to accommodate the different combinations of orderable tests.

Applying Automated Intelligence to Review of the CBC Differential Count

In 2005, the International Society for Laboratory Hematology (ISLH) published recommended indications for slide review and/or manual white blood cell (WBC) differential count.\(^4\) This was an industry-funded initiative led by Berend Houwen (founder of both the ISLH and the journal Laboratory Hematology, employed at that time by Beckman Coulter). It was intended to meet a need in clinical laboratories using the new automated hematology instruments. Many laboratories were asking vendors for help developing criteria for the automated release of normal and near-normal results. The ISLH consensus group comprised 20 experts from multiple countries and used multiple instruments. Criteria at the participants’ laboratories were discussed and condensed into a set of 41 rules, which are available for review on the ISLH website (http://www.islh.org/2010/index.php?page=consensus_rules). The group’s goal was to minimize the number of significantly abnormal cases missed by auto-validation, while maximizing the number of patient results that could appropriately be allowed to report without review.

Differential Review Consensus Guidelines: One Size Does Not Fit All

The ISLH Steering Committee for the International Consensus Group has emphasized that the consensus guidelines are recommendations, not rules. There was great diversity of opinion developing many of the recommendations, and it is unlikely that any of the participants apply the guidelines exactly as written. For a number of legitimate reasons, a uniform set of review criteria cannot be applied universally. These include the following:

- Review criteria are designed to detect and verify clinically significant abnormalities. The definition of *significant* can vary dramatically between practices. For example, criteria for review tend to be more liberal in oncology practices, where cytopenias are common. In outpatient laboratories where most patient results are normal, more conservative criteria are often used. Laboratories serving neonatal intensive care units (NICUs) may need infant-specific policies.
- Test volume can affect the laboratory’s capacity to review CBCs. Observation suggests that small laboratories that see few abnormal samples often choose to review mildly abnormal cases that laboratories doing thousands of CBCs per day will allow to auto-validate.
Many criteria are triggered by instrument flags (e.g., Suspect Blasts) or by numeric results exceeding linearity. The flags reported, flagging sensitivity, and linearity limits may vary by instrument make/model. Unique modes on a given instrument (e.g., a low WBC mode) may affect how low the automated results remain reliable.

Likewise, many criteria are designed to detect interferences, which will vary depending on the instrument. The mean corpuscular hemoglobin concentration level at which lipemia should be investigated, for instance, may vary between platforms.

There are regional and international differences in slide review criteria. For instance, the reporting of bands, or left shift, was described by one ISLH participant as “an American parochialism”; the ISLH guideline deliberately sidesteps this issue by making the recommendation to follow lab SOP for automated left-shift flags.

Because of such factors, it is common practice for laboratories to develop customized differential review criteria via deliberation involving pathologists and senior hematology technologists. Local interested physicians may also provide input.

**Guiding Principles in Evaluating Criteria Used to Prevent Auto-Validation**

A guiding principle for differential review criteria is to reduce review to the greatest extent possible. However:

- **Common interferences** should be intercepted before results are released (e.g., cold agglutinins, lipemia).
- **Sample collection problems** should be detected when possible (e.g., platelet clumping/clotting, hemolysis, dilution, mislabeled specimens).
- **Very abnormal results** may require confirmation of a new finding by rerunning a sample or by slide review. With improvements in instrument reliability and precision, the necessity for rerunning samples may be debated. However, many laboratories elect to rerun all critical values on the same sample before reporting. Slide review of very abnormal cell count results to detect WBC/RBC/PLT morphologic abnormalities is recommended.
- **WBC abnormalities**: Consider slide review for cases that may indicate any of the following:
  - Leukemia or other hematopoietic neoplasms
  - Reactive processes (to detect reactive lymphocytes; significant neutrophilic left shift as defined at the testing facility)
- **RBC abnormalities**: Consider slide review for cases that may indicate any of the following:
  - Hemolytic anemias (spherocytes, schistocytes, polychromasia, nucleated red blood cells [NRBCs], very high RBC distribution width and/or reticulocyte count, RBC Fragments flag)
• Hereditary conditions (e.g., high RBC count and low mean corpuscular volume [MCV] in thalassemia; hereditary spherocytosis)

• Leukoerythroblastosis (NRBC + immature granulocytes) in adults for possible marrow-occupying (myelophthistic) disorders such as myeloma, myelofibrosis, and metastatic carcinoma

• Moderate/marked low MCV for iron deficiency/thalassemia

• Moderate/marked high MCV for oval macrocytes, hypersegmented neutrophils, dysplastic atypia such as pelgeroid neutrophils, etc.

• **Platelet Abnormalities**

  o Review low counts for clumping/clotting; schistocytes typical of thrombotic thrombocytopenic purpura (TTP) / disseminated intravascular coagulation (DIC); large platelets typical of idiopathic thrombocytopenic purpura (ITP)

  o Very high counts for abnormal platelet (PLT) morphology, as seen with essential thrombocytopoiesis; teardrop cells, as seen with myelofibrosis; or other indications of myeloproliferative disorders

• **Instrument limitations:** Analyzer weaknesses may require rules to catch spurious results. For example, at the author’s laboratory system, the NRBC review level was raised to 5/100 WBC in adults. It was subsequently discovered that one instrument (of 21 in the system) was intermittently reporting increased NRBCs on normal samples, which were auto-releasing, and quality control (QC) samples were not detecting this problem. Letters recommending retesting at no charge were sent to physicians of affected patients. The instrument problem was corrected with a service call, and the decision was made to reduce the review level back to 1/100 WBC in adults.

• **Instrument flags indicating analytic uncertainty:** Instruments may flag results as “unreliable” using different flagging systems. These include flags for abnormal scattergram or abnormal distribution.

• **Linearity:** Samples above linearity may require dilution, or manual counting in the case of automated NRBC or reticulocyte counts exceeding linearity.

• **Delta checking:** Dramatic differences in results (particularly RBC parameters) over a short period of time should be investigated. Platelet delta checks (particularly in a downward direction) may pick up platelet clumping/clotting, but false-positive flags are common. The WBC count is inherently volatile (neutrophil half-life in circulation is 6 to 8 hours), so is not recommended for use in delta checking. Delta checks are usually applied over a short span of time (<72 h in the ISLH recommendations). A limitation of delta checking is that it is only useful on patients undergoing serial testing (e.g., inpatients); it will not detect a problem with the first sample drawn.
in a series. However, it can be useful in reducing repeated nonproductive review for chronic conditions such as anemia.

- **Neonatal CBCs:** Some institutions decide to perform a slide review or manual differential count on all infants (<2 months), to confirm the platelet count (due to frequency of platelet clumping with these patients) and to look for reactive atypia due to the common concern for sepsis.

### Special Considerations with Slide Review / Manual Differential Criteria

- **Neutrophilic Left Shift:** Opinions about band neutrophil reporting are strongly held, and arguments tend to take on a religious fervor; there are “band atheists,” who deny their meaningful existence, and “band fundamentalists,” who insist on their recognition. Most CBC instruments do flag for left shift when bands are increased, but they do not enumerate the bands separately. The College of American Pathologists (CAP) and the Clinical Laboratory Standards Institute (CLSI) recommend against reporting the band percentage.

  Conversely, a number of studies report that “bandemia” can be useful in detecting appendicitis, and there are studies showing the utility of bands in evaluation of neonatal sepsis and occult infections such as *Clostridium difficile* in adults. Band counts greater than 10% are one of the American College of Chest Physicians systemic inflammatory response syndrome (SIRS) criteria used to diagnose sepsis.

In aggregate, the medical literature is contradictory on the utility of bands. As a result, many laboratories (in the United States, at least) are faced with the difficult decision of how to deal with persistent requests from the emergency department and NICU for band counts. Current practice in North America varies highly among laboratories, ranging from never reporting bands to always reporting them when a manual differential count is performed. An intermediate approach (used at the author’s laboratory) is to report immature granulocyte percentage (IG % - which includes metamyelocytes, myelocytes, and promyelocytes), but not band percentage, when an auto CBC is ordered. However, if a Left Shift flag (or any other differential flag) is reported by the instrument, a 100-cell differential count is performed with the bands tallied. A left shift result is then reported as 1 to 4+, with the semiquantitative value derived from the manual count as follows:

<table>
<thead>
<tr>
<th>Bands</th>
<th>IG % Left Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤10%</td>
<td>Not reported</td>
</tr>
<tr>
<td>11-19%</td>
<td>1+ Left Shift</td>
</tr>
<tr>
<td>21-29%</td>
<td>2+ Left Shift</td>
</tr>
<tr>
<td>30-39%</td>
<td>3+ Left Shift</td>
</tr>
<tr>
<td>≥40%</td>
<td>4+ Left Shift</td>
</tr>
</tbody>
</table>

This approach provides an indication of left shift in a semiquantitative way, which reflects the
inherent variability of manual band counts better than a raw band percentage. This also provides a path for commenting on the occasional cases that have a normal WBC count but give an instrument Left Shift flag and show dramatic increases in bands on slide review (a so-called degenerative left shift).

- **Lymphocytosis:** For all differential criteria, the absolute count (diff #) rather than percentage (diff %) should be used. A common cutoff for review of lymphocytosis is lymphocyte count $\geq 5.0 \times 10^3/\mu L$, which represents the minimum diagnostic criteria level for chronic lymphocytic leukemia. It is quite common, however, for the laboratory to see CBCs from the emergency department in patients who have transient stress lymphocytosis related to seizure, trauma, asthma attacks, or resuscitation attempts; these counts almost invariably normalize by the next day. There have been two good studies examining the predictive value of lymphocyte count in adults, which show that nonspecific lymphocytosis is common, and the optimum cutoff for review is around $7.0 \times 10^3/\mu L$.\textsuperscript{13,14} At the author’s laboratory, a lymphocyte number between $5.0$ and $7.0 \times 10^3/\mu L$ in adults as a new finding is reviewed by a senior technician, and reported unless the morphology suggests a condition such as chronic lymphocytic leukemia (CLL; monotonous with smudge cells), hairy cell leukemia, or lymphoma, in which case the slide is forwarded for pathologist review. Values greater than $7.0 \times 10^3/\mu L$ as a new finding always receive pathologist review. It is also necessary (at the risk of stating the obvious) that much higher review levels are appropriate for children, because they have higher normal lymphocyte number values, peaking around 2 to 6 months and declining gradually thereafter.

- **RBC Morphology:** The ISLH general recommendation to limit reporting to at least 2+/moderate atypia is intended to combat the tendency to over-call RBC morphology (for instance, designating three or four types simultaneously as “rare”), which undercuts the utility of the report. Sickle cells in lesser numbers may qualify as a legitimate exception, if they are typical and not artifact. Also, per the International Council for Standardization in Hematology (ICSH), schistocyte levels of greater than 1.0% are an important diagnostic finding in TTP.\textsuperscript{15} Pains should be taken to report schistocytes above that level, and many laboratories do so quantitatively as a percentage of RBCs.

**Validating Slide Review Criteria in Your Laboratory**

Below is a suggested approach for validating slide review criteria, roughly parallel to the original ISLH process.

**Initial Optimization Study**

1. **Sample size:** The ISLH group asked each participating laboratory to collect 1000 samples. A large number of samples is optimal because many of the rules test for uncommon findings.
2. **Sample selection**: During this stage, it is desirable to include as many abnormal samples as possible to challenge the sensitivity of the rules. An ideal mix to challenge rule performance would be approximately equal numbers of true-positive and true-negative samples. It may be most practical to focus on collection of inpatient and/or oncology samples.

3. **Run samples for automated CBC/DIFF** on each instrument make and model to be evaluated. Any optimization studies to set user-definable instrument flagging levels (e.g., Q flags settings on Sysmex instruments [Sysmex Corporation]) should be done before the rule optimization study.

4. **Prepare slides for evaluation**. Per CLSI recommendations, this requires:
   - Three blood films for each sample. One is extra. More may be required for samples with low WBC counts.
   - Two examiners must each perform blinded 200-cell differential reviews (no knowledge of the automated differential count results or each other’s results). If in agreement, these results are averaged.
   - Discrepant results must be refereed by a designated expert (e.g., a pathologist).

5. **Positive results** are defined as abnormal results, which should not auto-release. The ISLH provides recommendations for adults, which include the following:
   - Metamyelocyte: >2%
   - Myelocyte/Promyelocyte: >1%
   - Blasts: >1%
   - Atypical lymphocytes: >5%
   - NRBC: >1%
   - Plasma cells: >1%
   - RBC Morphology: ≥2+/moderate
   - Malaria in any number
   - Platelet clumps: ≥ rare/occasional
   - PLT morphology: ≥2+/moderate
   - WBC morphology - Döhle bodies, toxic gran, toxic vacuolization: ≥2+/moderate

6. **The automated CBC results** should be sorted into four categories (Table 1), based on correlation between the positive/negative status defined by the rules, versus the “truth” defined by the manual differential review.
Table 1. Truth table.

<table>
<thead>
<tr>
<th></th>
<th>Automated CBC/DIFF</th>
<th>Manual Differential / Slide Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>True-Negative (TN)</td>
<td>No rule violations</td>
<td>Not POSITIVE</td>
</tr>
<tr>
<td>True-Positive (TP)</td>
<td>Rule violation(s)</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>False-Negative (FN)</td>
<td>No rule violations</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>False-Positive (FP)</td>
<td>Rule violation</td>
<td>Not POSITIVE</td>
</tr>
</tbody>
</table>

7. **False-negative and false-positive cases** should then be reviewed closely to see if the rules can be adjusted to increase or decrease sensitivity and specificity of each rule. For example: if the PLT Clumping flag misses significant cases of clumping, the rule for review of low platelet counts may need to be adjusted upward to prevent reporting spurious thrombocytopenia. This will result in improved sensitivity but decreased specificity and will increase the rate of slide review. A decision balancing patient safety and efficiency must be made in such cases by the medical director for that laboratory section.

**Validation Study**

1. The entire study is repeated with fresh cases, capturing a representative sampling of patients, for instance by collecting consecutive patient samples over a number of days. Running the study over multiple days allows evaluation of delta checking rules in reducing slide review.

2. These data then allow calculation of the following data for the laboratories’ actual workload:

   - **Sensitivity** = $\frac{TP}{TP+FN}$
   - **Specificity** = $\frac{TN}{FP+TN}$
   - **Accuracy rate** = $\frac{(TP+TN)}{Total}$
   - **Error rate** = $\frac{(FP+FN)}{Total}$
   - **Review Rate** = $\frac{(TP+FP)}{Total}$
   - **Predictive Value of a Positive Result** = $\frac{TP}{(TP+FP)}$
   - **Predictive Value of a Negative Result** = $\frac{TN}{(FN+TN)}$

The most common errors can also be tabulated (Table 2).

A word of caution: A low review rate, although desirable, says nothing about quality. It would be possible to create rules so liberal that almost nothing is reviewed, because nothing was being caught. Review rates should be developed as part of a larger discussion including quality measurements: sensitivity, specificity, and overall accuracy.
EDUCATIONAL COMMENTARY – CRITERIA FOR AUTOMATED REFLEX TO MANUAL DIFFERENTIAL (cont.)

Table 2. Sample report summary evaluating five different instruments.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True-Negatives</td>
<td>68.3%</td>
<td>61.2%</td>
<td>62.0%</td>
<td>52.1%</td>
<td>50.1%</td>
</tr>
<tr>
<td>True-Positives</td>
<td>11.1%</td>
<td>14.3%</td>
<td>9.8%</td>
<td>19.5%</td>
<td>12.7%</td>
</tr>
<tr>
<td>False-Negatives</td>
<td>12.4%</td>
<td>7.5%</td>
<td>12.1%</td>
<td>3.9%</td>
<td>9.1%</td>
</tr>
<tr>
<td>False-Positives</td>
<td>8.2%</td>
<td>16.9%</td>
<td>16.1%</td>
<td>24.4%</td>
<td>28.0%</td>
</tr>
<tr>
<td>Accuracy rate</td>
<td>(TP + TN)/Total</td>
<td>79.4%</td>
<td>75.6%</td>
<td>71.8%</td>
<td>71.7%</td>
</tr>
<tr>
<td>Error rate</td>
<td>(FP + FN)/Total</td>
<td>20.6%</td>
<td>24.4%</td>
<td>28.2%</td>
<td>28.3%</td>
</tr>
<tr>
<td>Review rate</td>
<td>(TP+FP)/Total</td>
<td>19.3%</td>
<td>31.6%</td>
<td>25.9%</td>
<td>44.0%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>TP/(TP+FN)</td>
<td>47.2%</td>
<td>65.7%</td>
<td>44.8%</td>
<td>83.3%</td>
</tr>
<tr>
<td>Specificity</td>
<td>TN/(FP+TN)</td>
<td>89.3%</td>
<td>78.3%</td>
<td>79.4%</td>
<td>68.1%</td>
</tr>
<tr>
<td>Predictive value,</td>
<td>TP/(TP+FP)</td>
<td>57.5%</td>
<td>45.8%</td>
<td>38.0%</td>
<td>44.4%</td>
</tr>
<tr>
<td>positive</td>
<td>TN/(FN+TN)</td>
<td>84.6%</td>
<td>89.1%</td>
<td>83.6%</td>
<td>93.0%</td>
</tr>
<tr>
<td>Predictive value,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

False-Negatives
- Bands (11-19%) 8.1% 5.9% 7.2% 2.9% 6.2%
- Bands (20-29%) 1.9% 1.0% 1.6% 0.3% 1.0%
- Bands (30-39%) 0.3% 0.3% 0.7% 0.7% 0.7%
- Bands (>40%) 0.7% 0.3% 0.3% 0.3% 0.3%
- NRBC 1.3% 0.3% 2.0% 0.7% 1.0%
- PLT Clumps 1.0%

False-Positives
- Bands/ImmGran 5.2% 12.7% 4.3% 22.1% 4.9%
- PLT Clumps 0.7% 6.2% 0.7% 16.9%
- Blasts 0.3% 0.7% 4.3% 0.3% 0.7%
- NRBC 1.3% 1.6% 1.0% 4.2%
- Atyp/Var Lym 1.0% 0.3% 0.7% 0.3% 1.3%
- Incomplete diffs 0.3% 0.7% 1.6%

Summary

Computerization has allowed the use of complex rules to automatically report normal and near-normal CBC / differential results. This has greatly improved the process for determining if manual slide review is necessary. However, a balance must be found between maximizing the number of patient results allowed to report without review, and minimizing the number of significantly abnormal cases missed by auto-validation. To aid in this effort, the ISLH has published consensus guidelines for differential review.
These guidelines are recommendations that should be customized by each laboratory depending on their patient population, instruments used, and other factors. Validation studies should then be performed in the laboratory to optimize the auto-validation rules chosen, and to verify that the sensitivity, specificity, and overall accuracy are satisfactory.

References


© ASCP 2017