Five Fully Automated Methods for Performing Immature Reticulocyte Fraction

Comparison in Diagnosis of Bone Marrow Aplasia

Mauro Buttarello, MD,1 Pietro Bulian, MD,1 Giorgio Farina, MD,1 Maria Grazia Petris, MD,2 Valeria Temporin, MD,1 and Lucia Toffolo, MD1

Key Words: Automated analyzers; Immature reticulocyte fraction; Bone marrow aplasia

Abstract

We performed a parallel evaluation of 5 automated reticulocyte counters to produce the immature reticulocyte fraction (IRF). We analyzed 225 samples from healthy control subjects, 115 from patients with various diseases, 38 with advanced aplasia, and 22 in early erythropoietic recovery after chemotherapy or bone marrow transplantation. The reference intervals were different for each instrument (ADVIA 120, 0.04-0.25; CELL DYN 4000, 0.15-0.35; GEN-S, 0.20-0.37; SE 9500 RET, 0.05-0.21; VEGA RETIC: 0.06-0.23). The imprecision, obtained by 1-way analysis of variance on duplicates, was satisfactory for clinical use for all methods (coefficient of variation, 7.6%-20.5% in healthy subjects), although it was higher than the analytic goal based on biologic variability within subjects. The comparison of different methods shows that agreement is good only between SE 9500 RET, CELL DYN 4000, and VEGA RETIC (r² = 0.72-0.78). The study of diagnostic performance in distinguishing aplasia from early bone marrow recovery shows slightly different results (area under the curve from 0.70 for ADVIA 120 to 0.96 for SE 9500 RET). Even with slight differences, the fluorescence-based methods seem to be more robust than other methods for IRF measurement.

Reticulocytes represent immature erythrocytes in the final stage of differentiation. They originate from orthochromatic erythroblasts following the ejection of the nucleus, and they gradually mature, partly in the marrow (3 days on average) and partly in the peripheral blood (1 day).1 Besides morphologic and structural changes, the process of maturation brings with it a gradual reduction in the amount and compactness of the reticulum, which is no longer detectable after the final transition to mature RBCs. The name reticulocyte comes from the microscopic characteristic of the cells after staining with supravital basic dyes that selectively stain the reticulum, giving it the appearance of scattered granules or a fine network or clump.

An attempt was made to classify the maturity of reticulocytes based on the quantity of the reticulum they contain. The first attempt was made by Heilmeyer,2 who divided the cells into 4 categories (the fourth being the most mature). The method was based on microscopic observation after brilliant cresyl blue stain.2 Despite the potential usefulness of a classification of reticulocytes based on their maturity as an index of marrow erythropoietic activity, this practice did not find a clinical application because the results were not reproducible.

In more recent years, it has been demonstrated that the reticulum is predominantly composed of protein and ribosomal RNA.3,4 The introduction of flow-cytometric methods using stains that selectively bind RNA and, therefore, produce signals that are proportionate to the RNA content, has allowed for the classification of individual reticulocytes based on their maturity in a reproducible way.

To express the degree of maturity with cytofluorimetric techniques, 2 strategies have been developed: (1) The first to be adopted was initially based on the use of multipurpose...
cytofluorimeters and used an arbitrary average maturity index that corresponded to the mean fluorescence channel: reticulocyte maturation index.\(^5,6\) (2) The second, introduced more recently, used mostly on dedicated analyzers, divides the reticulocytes into subpopulations (2 or 3) according to the RNA content and, therefore, the different maturity levels.

The term immature reticulocyte fraction (IRF) was proposed to indicate the least mature fraction of reticulocytes.\(^6\) In an attempt to standardize the results obtained from the different methods, it was proposed that the systems that identify 3 populations of differing maturity should define the IRF as the sum of the populations of high and medium immaturity. The IRF, even in its various formulations, has proved to be clinically useful as an early and sensible index of erythropoietic activity. The best clinical use, especially in the classification of anemias based on the evaluation of erythropoietic response, can be obtained by plotting the IRF against the absolute reticulocyte count.\(^7\)

A particularly useful application in cases of reticulocytopenia is the identification of early marrow regeneration following bone marrow transplantation or chemotherapy. This can be noted by the reappearance of reticulocytes with a high RNA content and, therefore, by an increase in the IRF with respect to near zero levels typical of aplasia.

In recent years, in addition to the cytofluorimeters, traditional hematology analyzers have appeared on the market that, using different methods, are able to count reticulocytes and quantify the maturity level. The proliferation of differing technologies brings with it the pressing need for standardization,\(^8\) which has been made difficult by the following:

1. The persistence of specific definitions of immature reticulocyte fraction by manufacturers of hematology analyzers
2. The lack of a “universal” stable reference material and of a reference method for quantifying the IRF
3. The rigidity of calibration procedures strictly dependent on the manufacturer and that cannot easily be modified by the user

This study was designed primarily to compare 5 of the most modern analyzers currently on the market with the following aims:

- Calculating and verifying the comparability of reference intervals of the IRF
- Verifying the agreement between the different methods in defining the state of erythropoiesis by using a bivariate matrix (IRF vs absolute count)
- Establishing clinical usefulness of the IRF obtained with each of the 5 methods for differentiating bone marrow aplasia from early recovery on erythropoiesis

### Materials and Methods

In a parallel study at the clinical pathology laboratory of the Ospedale Geriatrico (geriatric hospital), Padua, Italy, during a 6-month period, 5 analyzers able to count reticulocytes and provide IRF indexes in full automation were evaluated. The analyzers considered were the following: (1) ADVIA 120, software version 1.16 (Bayer Diagnostic Division, Tarrytown, NY); (2) CELL DYN 4000, software version 7.5 (Abbott, Santa Clara, CA); (3) GEN-S, software version 1.D (Coulter-Beckman, Hialeah, FL); (4) SE 9500 RET, software version B 00-04 (SYSMEX, Kobe, Japan); and (5) VEGA RETIC (currently known as PENTRA 120), software version 3.16 (ABX, Montpellier, France).

**Table 1** shows the stains, general principles, and methods used to define the levels of maturity. Not all analyzers use the same criteria to define the IRF: ADVIA 120, SE 9500 RET, and VEGA RETIC subdivide the reticulocyte population into 3 fractions, high (H), medium (M), and low (L) immaturity, while the CELL DYN 4000 and the GEN-S classify the cells into 2 fractions; the more immature fraction corresponds to the IRF. To make the results comparable, for the methods that provide 3 populations, the IRF was calculated by summing the fractions of medium and high immaturity.

The instruments were calibrated by company representatives according to the manufacturer’s specifications. During evaluation, they were checked 3 times a day using controls supplied by each manufacturer.

### Table 1
**Analytic Methods and Reference Intervals on Immature Reticulocyte Fraction Determination**

<table>
<thead>
<tr>
<th>Analyzer*</th>
<th>Dye</th>
<th>Technology</th>
<th>Median</th>
<th>2.5°</th>
<th>97.5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVIA 120</td>
<td>Oxazine 750</td>
<td>Absorbance</td>
<td>0.08</td>
<td>0.04</td>
<td>0.25</td>
</tr>
<tr>
<td>CELL DYN 4000</td>
<td>CD4K 530</td>
<td>Fluorescence</td>
<td>0.24</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>GEN-S</td>
<td>New methylene blue</td>
<td>Light scattering</td>
<td>0.28</td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td>SE 9500 RET</td>
<td>Auramine O</td>
<td>Fluorescence</td>
<td>0.13</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>VEGA RETIC</td>
<td>Thiazole orange</td>
<td>Fluorescence</td>
<td>0.14</td>
<td>0.06</td>
<td>0.23</td>
</tr>
</tbody>
</table>

* For proprietary information, see the text.
Samples

To calculate the reference intervals for the analyzers, 225 healthy subjects, according to the National Committee for Clinical Laboratory Standard H44-A7 definition, of both sexes and between 3 and 50 years of age were selected.

In addition, samples from 115 subjects affected by various conditions, such as hemolytic or posthemorrhagic anemia, acute and chronic inflammation, neoplastic diseases, nutritional anemias before and during treatment, and bone marrow aplasia, were analyzed.

The diagnostic efficiency of IRF for distinguishing advanced aplasia from an early recovery of the marrow was verified by selecting 60 reticulocytopenic patients in the period following marrow transplantation or following chemotherapy. These latter patients were between 3 and 20 years old and had not received transfusions during the 5-day period immediately preceding the sampling. The corresponding diseases were as follows: acute lymphoblastic leukemia (11 cases), non-Hodgkin lymphoma (9 cases), acute nonlymphoblastic leukemia (21 cases), nervous system neoplasia (8 cases), and other neoplastic disease (11 cases).

The condition of reticulocytopenia was established by using the lower limit of the reference interval previously calculated for each of the 5 methods as follows: ADVIA 120, 27 × 10^3/µL (27 × 10^9/L); CELL DYN 4000, 25 × 10^3/µL (25 × 10^9/L); GEN-S, 20 × 10^3/µL (20 × 10^9/L); SE 9500 RET, 23 × 10^3/µL (23 × 10^9/L); and VEGA RETIC, 30 × 10^3/µL (30 × 10^9/L).10

The criteria used to define the condition of marrow aplasia were as restrictive as possible11,12: polychemotherapy that was begun at least 7 days before and interrupted for not more than 3 days; no more than 5 days after a marrow transplant; in all cases, there was coexistence of leukocytopenia (<2,000/µL [<2.0 × 10^9/L]) and thrombocytopenia (<50 × 10^9/L [<50 × 10^9/L]). By using these criteria, 38 patients were identified as having a condition of aplasia.

To define early marrow recovery following aplasia, we considered the first day of a trend of increasing neutrophils starting from the time of nadir if the increase was greater than 100/µL (0.1 × 10^9/L), or we considered the second consecutive increase if any single increase was less than 100/µL (0.1 × 10^9/L). This was necessary to overcome analytic imprecision of the neutrophil count, which is particularly high at low concentrations. By using these criteria, 22 patients were identified as being in early recovery. All samples were collected by venipuncture into 5-mL tripotassium EDTA evacuated tubes (Vacutainer, Becton Dickinson, Plymouth, England). For the patient with marrow aplasia, samples were taken from leftover blood collected for routine CBC counts. The samples were analyzed by all instruments in parallel and in duplicate within 4 hours following collection. For each patient we used 1 tube that was assigned randomly to each analyzer and rerandomized for duplicates. The sample was allowed to stand for at least 20 minutes between counts. Before evaluation, the stability of the IRF with samples kept at room temperature for the 4 hours following collection was verified. In different experiments, 2 samples with low and 4 with normal or high reticulocyte concentrations were analyzed by each instrument. The counts were performed at 0, 1, 2, 3, and 4 hours following collection. No significant differences were found between IRF values at 0 and 4 hours.

Statistical Analysis

The results obtained represent the average of duplicate counts for most samples. To calculate imprecision for the given method, analysis of variance was used on duplicates.13 This was calculated separately for samples with reticulocytopenia and for those with a normal or high concentration.

The calculation of reference intervals was performed with a nonparametric method (middle 95th percentile of the distribution) since the distribution of reticulocyte counts and of the IRF values are approximately log normal.9 The bivariate matrix of IRF against reticulocytes × 10^9/L was subdivided into 9 areas as defined by intersections obtained by prolonging the threshold of the respective reference intervals. For the aim of this research, the areas characterized by low reticulocyte counts and low, normal, or high IRF values are particularly interesting. The area closest to the origin of the axis is thought to correspond to conditions of aplasia, while the areas with low counts but with normal or high IRF values are typical of early erythropoietic recovery. Since a correlation has been noted between absolute reticulocyte counts and IRF,14,15 this was verified, not only with the entirety of the data, but also in relation to the different erythropoietic conditions characterized by low, normal, or high reticulocyte counts. The calculation of diagnostic efficiency of IRF as provided by the 5 analyzers for distinguishing between the 2 conditions characterized by reticulocytopenia was performed using receiver operating characteristic (ROC) analysis with a nonparametric method to calculate the area under the curve.16,17 Then, the best cutoff value for each method and the corresponding sensitivity and specificity were determined.

Results

Reference Intervals

The reference intervals for the 5 methods are reported in Table 1; Figure 1 shows the corresponding distribution histograms.

It can be seen that while the extreme values are relatively superimposable for the ADVIA 120, SE 9500 RET,
Reference intervals for immature reticulocyte fraction counts (n = 225). **A**, ADVIA 120. **B**, CELL DYN 4000. **C**, GEN-S. **D**, SE 9500 RET. **E**, VEGA RETIC. For proprietary information, see the text.
and VEGA RETIC, in reality, a true overlap exists only in the case of the last two, since the ADVIA 120 presents a flattened distribution toward the left. The CELL DYN 4000 and the GEN-S show different intervals with respect to the previous methods, with a threshold shifted toward higher values. Even in this case, there is modest overlap. It is interesting to note that the results obtained with the SE 9500 RET are the same as those obtained on samples from a North American adult population analyzed with the SYSMEX R-1000, which uses the same technology.8

Comparison Between Methods

Table 2 compares the correlation and regression statistics for the 5 methods for all samples analyzed (n = 400). These results are very encouraging in light of the variability between methods using fluorescent dyes, which was noted in an interlaboratory correlation study. They are in line with other published data that compare the SYSMEX R-3000, CELL DYN 4000, and cytofluorimetric methods using thiazole orange. Good agreement clearly can be seen between the SE 9500 RET, CELL DYN 4000, and the VEGA RETIC ($r^2$ between 0.72 and 0.78). Between these 3 instruments and the ADVIA 120 or GEN-S, the low value of $r^2$ (from 0.29 to 0.43) indicates an excessive spread of the data.

Imprecision

The coefficients of variation (CVs) obtained are shown in Table 3.

The results obtained in the healthy population are not different from data previously published by other authors using the Bayer H3, SYSMEX R-1000, CELL DYN 4000, and PENTRA 120.19-21 In reticulocytopenia for the same number of events, the absolute count of cells identified as reticulocytes is lower than normal, and, as a result, so is the number of immature reticulocytes (IRF). As a consequence, there was an increased imprecision with a CV that was 3 times as high as normal.

It can be noted that the CV percentage is low for analyzers like the CELL DYN 4000 and the GEN-S, which usually provide higher IRF values, highlighting an inverse relationship between IRF values and CV percentage.

Two-Parameter Matrix

Figure 2 shows the 2-dimensional matrix for each of the 5 analyzers with all IRF values and their relative absolute reticulocyte count for healthy subjects and for all patients enrolled in the study.

There is a vast amount of literature11,12,22-25 on the clinical usefulness of IRF values (H + M populations) obtained

---

**Table 2**

<table>
<thead>
<tr>
<th>Analyzer</th>
<th>ADVIA 120</th>
<th>CELL DYN 4000</th>
<th>GEN-S</th>
<th>SE 9500 RET</th>
<th>VEGA RETIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>—</td>
<td>18.16</td>
<td>24.61</td>
<td>7.80</td>
<td>9.10</td>
</tr>
<tr>
<td>Slope</td>
<td>—</td>
<td>0.76</td>
<td>0.55</td>
<td>0.64</td>
<td>0.55</td>
</tr>
<tr>
<td>$r^2$</td>
<td>—</td>
<td>0.34</td>
<td>0.31</td>
<td>0.38</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**Table 3**

<table>
<thead>
<tr>
<th>Analyzer</th>
<th>Normal (Mean, CV (%))</th>
<th>Reticulocytopenic (Mean, CV (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVIA 120</td>
<td>0.11, 19.9</td>
<td>0.08, 65.0</td>
</tr>
<tr>
<td>CELL DYN 4000</td>
<td>0.26, 7.6</td>
<td>0.20, 23.4</td>
</tr>
<tr>
<td>GEN-S</td>
<td>0.29, 10.8</td>
<td>0.30, 16.1</td>
</tr>
<tr>
<td>SE 9500 RET</td>
<td>0.15, 15.4</td>
<td>0.08, 43.9</td>
</tr>
<tr>
<td>VEGA RETIC</td>
<td>0.16, 20.5</td>
<td>0.08, 54.8</td>
</tr>
</tbody>
</table>

ANOVA, analysis of variance; CV, coefficient of variation.

* For proprietary information, see the text.
Figure 2. Two parameter matrix: relationship of immature reticulocyte fraction (IRF) and absolute reticulocyte count. The black dots show the samples that the SE 9500 RET placed in the “aplastic” category. A, SE 9500 RET. B, ADVIA 120. C, CELL DYN 4000. D, GEN-S. E, VEGA RETIC. Values are given in Système International units; conventional units are $\times 10^3/\mu L$ and the conversion factor is 1.0. For proprietary information, see the text.
with the R series of SYSMEX analyzers, which have the same technology as the SE 9500 RET. To verify the agreement between methods, the SE 9500 RET therefore was used for comparison to see where the other methods placed the same patients previously defined as “aplastic” by the SE 9500 RET. This is shown by the black dots in Figure 2. From the position of these dots, it can be seen that for the VEGA RETIC and the CELL DYN 4000, most patients were placed in the area of aplasia, a few in the area of early marrow recovery, and only 1 was defined as “normal.” The ADVIA 120 had a higher number of patients in the area of early marrow recovery, while for the GEN-S, the majority of the patients were placed in this area.

Correlation and regression statistics between IRF and reticulocyte counts for the 5 analyzers on the total number of samples, separately subdivided according to reticulocyte count (low, normal, or high), are reported in Table 4. The results indicate the existence of a correlation that is modest between these variables or lacking in the field of reticulocytopenia.

Clinical Sensitivity

Figure 3 shows the ROC curves relative to the diagnostic efficiency of the IRF produced by the different analyzers in distinguishing patients with advanced aplasia from an early recovery of erythropoiesis following aplasia.

The diagnostic efficiency (measured as the area under the curve) is higher for the SE 9500 RET (0.96), followed by the VEGA RETIC (0.90), the CELL DYN 4000 (0.87), the GEN-S (0.71), and the ADVIA 120 (0.70). The statistical tests do not indicate any statistical difference among the first 3 analyzers, while they show a significant difference between each of these (SE 9500 RET, VEGA RETIC, CELL DYN 4000) and the other 2 analyzers (GEN-S and ADVIA 120).

The cutoff values were extrapolated by a nonparametric method from the individual ROC curves. These values Table 5 correspond to the best association between sensitivity and specificity, given the hypothesis that both have the same undesirability of errors. Comparing the cutoff values obtained in this way with the lower limit of the reference interval, a substantial overlap can be noted for the VEGA RETIC, SE 9500 RET, and the ADVIA 120, while values are slightly higher for the CELL DYN 4000 (0.18 vs 0.15) and even higher for the GEN-S (0.29 vs 0.20).

Discussion

When monitoring certain clinical conditions, such as early recovery following bone marrow or stem cell transplantation or recovery after chemotherapy, the IRF has been shown to be a more sensitive parameter than reticulocyte or neutrophil counts. The usefulness for monitoring erythropoietin, iron, B₁₂, and folic acid therapies or, more generally, for diagnosing and classifying anemias seems more sporadic. A large part of these studies are based on the use of multipurpose flow cytometers and thiazole orange as a stain for RNA, or dedicated cell
counters (manufactured by SYSMEX) using auramine O. Both of these methods, either because of the need for experienced personnel (in the first case) or because of the elevated initial cost (in the latter), can be used in middle-sized to large laboratories. The possibility of having traditional hematologic analyzers that can count reticulocytes and provide the relative level of maturity has extended the possibility of using the IRF as a new parameter. The increase in the number of methods, however, has brought with it the need for comparison and standardization.

This research highlights the difference in the reference intervals that are strictly method-dependent. Even when there is overlap between methods that provide the IRF as the sum of the H + M populations (ADVIA 120, SE 9500 RET, and VEGA RETIC) or between the counters that produce a single value (CELL DYN 4000 and GEN-S), there is an acceptable interchangeability only between the SE 9500 RET and the VEGA RETIC (slope, 0.83; \( r^2 = 0.72 \)).

Imprecision of this parameter is higher for the systems (VEGA RETIC, ADVIA 120, and SE 9500 RET) that obtain the IRF as a sum of H + M and provide lower numeric values than the other 2 analyzers. Imprecision increases considerably (with the exception of GEN-S) when it is evaluated on reticulocytopenic samples and, thus, with a low number of events referred to the IRF. Furthermore, it can be seen that the CV percentages are higher for the methods in which the IRF represents a small fraction of the total, as in the SE 9500 RET, ADVIA 120, and VEGA RETIC (mean IRF = 0.08). For the CELL DYN 4000 and the GEN-S, in fact, the mean IRF was 0.20 to 0.30, and imprecision was the same as that for total reticulocyte counts.\(^{10}\) The imprecision thus seems to be inversely proportionate to the number of events counted as IRF. At particularly low levels, the imprecision also may be influenced by the analytic sensitivity of the method. The most widely accepted approach to evaluate the limits of acceptability of imprecision is based on biologic variability.\(^{26,27}\)

The only data available on biologic variability within individuals (Cvi) are distinct for H and M fractions and have values of 10% and 13%, respectively, for the SYSMEX and 33% and 17% for the Bayer H3 systems.\(^{20}\) In the present study, the H and M populations were added together, and to calculate the analytic goal (0.5 Cvi or less), the Cvi of the M fractions was used since this fraction was greater than the H fraction. If these criteria are applied, the analytic goal was not reached by several of these systems. However, for clinical use, all the methods can be considered acceptable. Moreover, the biologic variability within individuals should be reconsidered in light of the proposal to define the IRF as the sum of H and M populations.

The results obtained from the analysis of correlation between IRF and the absolute reticulocyte count indicate modest correlation, but significantly different from zero when the samples were considered on the whole. These data are very similar to those reported in other publications.\(^{14,15}\) It becomes necessary to separately evaluate the behavior in certain areas of the graph. In fact, while for all systems the correlation is very low in conditions of reticulocytopenia or reticulocytosis, it does exist (\( r \) between 0.24 and 0.43) in the area that corresponds to healthy subjects. This could be explained by considering the fact that in healthy subjects there is a condition of stable erythropoiesis, in which the greater number of circulating reticulocytes corresponds to a greater fraction of young reticulocytes. In reticulocytopenia and reticulocytosis, which are basically transitory states, the 2 parameters are independent of each other and have different responses to erythropoietic factors.

The different allocation of reticulocytopenic samples in the 5 systems studied is interesting and unpredictable a priori. In fact, if certain systems (SE 9500 RET, VEGA RETIC, CELL DYN 4000) present a substantial agreement in classifying samples from subjects with advanced aplasia (reduced reticulocyte concentrations and reduced IRF), the ADVIA 120 and GEN-S placed part of the samples, classified as aplasia with the first 3 analyzers, in the area of normal or even increased IRF (which is the area corresponding to an early erythropoietic response).

When using the IRF in differential diagnosis between marrow aplasia and early erythropoietic response, both conditions characterized by reduced reticulocyte counts, the ROC analysis shows a different efficiency for the different methods. Methods using fluorescence and argon laser showed greater sensitivity and are more robust than methods using a helium-neon laser or diode laser, which measure light scattering or absorbance.

It can be confirmed that the IRF values provided by different systems do not have the same information content, at least in the conditions and with the software versions studied. The best cutoff values calculated with ROC analysis do not substantially differ from the lower limit of the reference interval, with the exception of the GEN-S where it was in an uncertain area, with IRF values that were shared between healthy subjects and those with aplasia, thus forcing the discrimination to rely on the absolute reticulocyte count.

The usefulness of the IRF as a “new” reticulocyte parameter was confirmed by all methods. Two problems remain to be resolved: standardization and different diagnostic efficiency that presently seem to be greater for fluorescence-based instruments. This leads to 2 consequences: (1) the need to use the same method in the sequential monitoring of marrow aplasia, and (2) caution in the use of bivariate graphs produced by some analyzers for differential diagnosis between marrow aplasia and early erythropoietic response.
References

27. Fraser GC, Petersen PH. Desirable standards for laboratory tests if they are to fulfill medical needs. Clin Chem. 1993;39:1447-1455.