The WHO classification of MDS does make a difference
Robert B. Howe, Anna Porwit-MacDonald, Robert Wanat, Ramin Tehranchi, and Eva Hellström-Lindberg

The purpose of this study was to determine the facility and reliability of the World Health Organization (WHO) classification of myelodysplastic syndromes (MDSs) with several observers reviewing the same diagnostic specimens. We also wanted to determine if the WHO classification provided additional information about predictability of clinical response outcome. To accomplish these goals we reviewed 103 previously diagnosed cases of low-risk MDS. We found 92% interobserver agreement (P < .001). Sixty-four of these patients had been entered into clinical trials using growth factors by the Nordic MDS Study Group. The WHO classification reliably predicted therapeutic response to the combination of granulocyte colony-stimulating factor (G-CSF) and erythropoietin (Epo). The response rate differed significantly between refractor anemia with ringed sideroblasts (RARS) and refractor anemia with multilineage dysplasia and ringed sideroblasts (RCMD/RS) with regard to therapeutic response (75% versus 9%; P = .003). Also, in the group of patients with less than 5% marrow blasts, there was a difference in median survival between patients with unilineage dysplasia (51% surviving at 67 months) and those with multiligneage dysplasia (median survival, 28.5 months; P = .03). (Blood. 2004;103:3265-3270)

© 2004 by The American Society of Hematology

From the Division of Hematology, Oncology and Transplantation, University of Minnesota, Minneapolis; the Department of Pathology, Karolinska Hospital and Karolinska Institute, Stockholm, Sweden; and the Department of Medicine, Huddinge University Hospital and Karolinska Institute, Stockholm, Sweden.


Supported, in part, by the Curtis L. Carlson University of Minnesota/Karolinska Institute Medical Research and Education Program and by a research grant from the Swedish Cancer Foundation. Patients included in the clinical studies were reported by investigators within the Nordic MDS Group. A list of the participating members of the Nordic MDS Group appears in Appendix 1.

Reprints: Robert B. Howe, University of Minnesota, Mayo Mail Code 480, 420 Delaware St SE, Minneapolis, MN 55455; e-mail: howex001@umn.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.
only 4 had blast percentages exceeding 10%. All patients had previously been included in either clinical or laboratory protocols for low-risk MDS and were well characterized regarding clinical and laboratory parameters. In so doing we addressed the difficult questions of defining dysplastic changes in myeloid and megakaryocytic cells, enumerating them, and assessing clear cutoffs for blast percentages in the various categories.\textsuperscript{16,21-23} Sixty-four patients have been included in previously reported studies on the combination of erythropoietin (Epo) and granulocyte colony-stimulating factor (G-CSF)\textsuperscript{24,25} and were reassessed for response to treatment in relation to category in the WHO classification system.

### Patients, materials, and methods

#### Patients

A total of 103 BM samples from MDS patients were evaluated. All patient samples were collected from previous clinical and laboratory studies. The majority of patients were included in the Nordic studies using growth factors (GFs) G-CSF plus Epo for the anemia of MDS,\textsuperscript{10,24} another group was included in the Nordic antithymocyte globulin study,\textsuperscript{25} and a third part were included in laboratory studies on low-risk MDS.\textsuperscript{26} The follow-up of the GF-treated patients was long enough to allow a survival analysis of these patients as of December 1, 2002.

#### Morphologic analysis

The BM samples consisted of either trephine bone marrow biopsy (BMB) or clot preparation from aspirate, and BM and blood smears. Most samples were directly processed at Karolinska or Huddinge Hospital in Stockholm. Others were received from other pathology departments for review. BM biopsies were formalin-fixed and decalcified by routine methods, or more recently, Heidenhein-SuSa fixative (mercuric chloride/formaldehyde/trichloracetic acid/acetic acid) was used. Clot preparations were fixed in routine methods, or more biopsies were formalin-fixed. Sections (4-5 μm) were cut and stained with Giemsa, hematoxylin-eosin, Prussian blue, and Gordon-Sweet reticulin, respectively. Biopsies were evaluated for overall cellularity, representation and maturation of hematopoietic lineages, and presence of abnormal localization of immature precursors (ALIPs), dysmegakaryopoiesis, fibrosis, and iron content. Smears were stained with May-Grünewald-Giemsa and assessed for comparative hematopoietic lineage representation, presence and degree of dysplasia (dyserythropoiesis, dysgranulopoiesis, and dysmegakaryopoiesis), and percentage of blasts (at least 400 nucleated BM cells were counted). Presence and numbers of sideroblasts were evaluated on smears stained with Prussian blue. RSs were defined as erythroid precursors in which one third or more of the nucleus is encircled by 10 or more siderotic granules as demonstrated in an iron stain. Fifteen percent or more of the erythroid precursors in the BM smears must be so identified to qualify as RARS.\textsuperscript{16}

#### Reclassification

The original FAB classification was done by the central pathology unit of the Nordic MDS Group studies, Karolinska Hospital, between 1990 and 2001. It could not be excluded that there had been a certain development in the morphologic assessment of BM and technical methods over this long period. Therefore, before reclassification to WHO was done, a reclassification using criteria of the FAB classification\textsuperscript{4} and a consensus allocation to an FAB group were done for each BM sample. This reclassification of FAB is reported in “Results,” and the consensus FAB is shown in Tables 1 and 4. Then, the WHO diagnostic criteria\textsuperscript{16} were applied rigorously. The clinical and laboratory studies followed the guidelines of the national and local ethical committees, respectively, and all patients had given informed consent.

#### Interobserver variation study

The 3 observers (R.B.H., A.P.M., R.W.) reviewed all 103 cases independently. Eight cases with discrepant blast percentages or borderline dysplasia were reviewed by all 3 pathologists together and a diagnostic consensus was reached by performing 500 cell differential counts on BM and reviewing dysplastic changes and RSs together. The 3 observers were blinded to the clinical and laboratory data including cytogenetics, as well as treatment results, until the morphologic review was complete.

#### Cytogenetic analysis

Ordinarily 15 to 30 metaphases (in the majority of cases at least 20) were examined according to standard operating procedures, using either Q- or G-banding, depending on the routines of the particular cytogenetic laboratory at each university hospital in the Nordic MDS Group. Cytogenetic aberrations were grouped according the original IPSS publication.\textsuperscript{6} Patients with 5q− aberrations were classified as described by Giagounidis et al.\textsuperscript{27} According to WHO, 5q− syndrome was defined as a case with less than 5% marrow blasts and 5q− aberration as a single abnormality. Other patients with 5q− abnormality, with or without additional chromosomal aberrations, were included in the other WHO subgroups.

#### Clinical studies

A special analysis with regard to the WHO classification and outcome of treatment was performed on 64 patients included in the Nordic G-CSF–Epo studies. Sixty of these were evaluable for IPSS score, and 62 were evaluable for a response to treatment.

#### Statistical methods

The interobserver concordance was evaluated using the Cohen κ test and the Spearman correlation test. We used χ² analysis to compare groups with regard to response to treatment. Median survival was estimated by Kaplan-Meyer analysis and we used the log-rank test to compare survival between groups.

### Table 1. Cases of MDS reclassified

<table>
<thead>
<tr>
<th>FAB classification</th>
<th>RA, n = 42</th>
<th>RARS, n = 38</th>
<th>RAEB, n = 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RARS</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>RAEB-1</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>RAEB-2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>RCMD</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RCMD/RS</td>
<td>0</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>5q-syndrome</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>CMML</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Unclassified</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Patients were reclassified according to FAB and then according to WHO. The table shows the reclassified FAB diagnoses. Seven cases were difficult to reclassify according to FAB guidelines and were not possible to classify according to WHO guidelines (ie, no megakaryopoiesis available for evaluation). These patients are listed under their original FAB diagnosis.
www.bloodjournal.orgFrom www.bloodjournal.org by guest on April 21, 2017. For personal use only.

Figure 1. Composite of neutrophils and megakaryocytes. Shown are normal (A), hypogranular neutrophils (B), normal (C) and abnormal (D,E) megakaryocytes. Original magnification × 100.

Results

FAB re-evaluation

A re-evaluation of the original FAB group occurred mostly in patients with blast percentages around 5%, poor iron staining, or otherwise suboptimal samples. Patients for whom it was impossible to define a FAB group were all deemed unclassifiable according to WHO. In Tables 1 and 4, these patients are included under their original FAB diagnosis. After the FAB re-evaluation, no patient fell outside the inclusion criteria for the clinical studies.

WHO classification consensus

The interobserver variation study showed 92% agreement among the 3 reviewers of BM morphology. A significant concordance was achieved while using WHO classification (κ 0, 909, P < .001; Spearman r, 0.939, P < .01, 2-tailed).

Of the 103 cases reviewed, there was consensus in 95 when reviewed independently (Table 1). The remaining 8 were resolved by joint review of the pathology specimens, peripheral blood counts, and cytogenetics (Table 2). Most disagreements concerned definition of dysmorphic changes in megakaryocytes: size, aberrations in maturation and number, and granulocytes: pseudo Pelger-Huet changes and abnormalities of granulation as well as their enumeration (Figure 1). The arbitrary cutoff of 10% abnormal cells in a given cell line was sometimes difficult to apply. The quantity and quality of the specimen and the staining characteristics were often central in the final judgment. There were also occasional difficulties in applying the criterion of more than 15% RSs to define RARS, although true RSs were usually easily identified if the iron stain was well prepared. In summary, 4 examples of RA called by one reviewer were called RCMD by a separate reviewer and 3 cases called RCMD/RS by one were categorized as RARS by the others. One case diagnosed RA by one observer was called RAEB-1 by 2 observers who counted more than 5% blasts in the BM (Table 2). This distinction may prove important, as discussed in “WHO classification and outcome of GF treatment.”

The percent blasts at the cutoff between RAEB-1 and RAEB-2, or RAEB-1 and other diagnostic categories was sometimes an issue. Repeat 500 cell differential counts were used to assign the category. Among those cases universally agreed on, all 3 observers called one RAEB, but differed on the blast percentage (5.5%-11%). Foci of blasts led one to classify this case as RAEB-2. In another 3 cases the threshold of 5% blasts resulted in change of the diagnosis from RA, RARS, or RCMD to RAEB-1. Cases defined as 5q–syndrome were more readily classified. One case of RAEB was reclassified as 5q–syndrome following review of the cytogenetics and finding fewer than 5% blasts on review.

Eleven cases demonstrated the 5q–aberration by conventional cytogenetics. Eight of these had megakaryocytic dysplasia and fewer than 5% BM blasts and were classified as 5q–syndrome. A single case resembled 5q–syndrome morphologically but had karyotype 46,XY. Two others had additional dysplasia and complex chromosome abnormalities. One had more than 5% BM blasts and was classified as RAEB-1.

WHO classification and outcome of GF treatment

Most RA and RARS patients had low IPSS scores (10 low, 2 Int-1), whereas RAEB-1 and RAEB-2 patients had higher scores (2 low, 6 Int-1, 1 Int-2, 2 high). RCMD and RCMD/RS patients were in the intermediate IPSS range (11 low, 8 Int-1, 3 Int-2, 0 high; Table 3).

Sixty-four cases entered into treatment studies with G-CSF and Epo were included in the 103 reclassified cases. These are outlined in Table 4. According to the Nordic response criteria,11 a complete response (CR) was defined as an increase in hemoglobin to at least 11.5 g/dL and a 100% reduction in transfusion need for 4 weeks or longer. Partial response (PR) was defined as either a 100% reduction in transfusion need or an increase in hemoglobin more than 1.5 g/dL.10,28-30

Although clinical responses were noted in each category (Table 5), CRs were most common in RA (67%) and RARS (50%). There was a striking difference in response rate between RARS and RAEB-2, whereas RAEB-1 and RAEB-2 patients had higher scores (2 low, 6 Int-1, 1 Int-2, 6 high). RCMD and RCMD/RS patients were in the intermediate IPSS range (11 low, 8 Int-1, 3 Int-2, 0 high; Table 3).

In the RARS group, 5 of 6 evaluable patients belonged to the good response category and one to the intermediate category. Of the 9 patients in the RCMD/RS group, 3, 5, and 1
Patients belonged to the good, intermediate, and poor response categories, respectively. The impact of unilineage versus multilineage dysplasia on survival was then analyzed in patients with less than 5% blasts (FAB RA and RARS). Patients with unilineage dysplasia had a better survival (51% surviving at 67 months versus median survival 28.5 months, \( P = .03 \); Figure 2). No difference was seen regarding outcome and survival in patients with and without RSs \( (P = .44) \).

Discussion

Much of the progress in understanding and management of malignant disease can be credited to the development and application of classification and staging systems that allow medical investigators to study comparable diseases in comparable patients. When a new classification system is presented, it is essential to compare it with the old classification both regarding its usability (interobserver variation) and its usefulness in the evaluation of clinical studies. The WHO classification has previously been evaluated with respect to prognosis\(^{17} \) of essentially untreated cases. The present study was designed to investigate interobserver variability, to define potential diagnostic difficulties, in particular regarding dysplastic features, and to evaluate the usefulness of the classification in predicting outcome of proerythroid GF treatment in a large cohort of patients mainly belonging to the low-risk categories.

For the MDSs the FAB and IPSS have served this role well. Although they have been widely applied and generally accepted, some deficiencies have been identified. The application of cytogenetics has identified a unique syndrome, that of 5q\(-\). It has also made possible the IPSS, a powerful prognostic tool.\(^6\) CMML has emerged as a myeloid disorder, which may exhibit characteristics of a myelodysplastic and a myeloproliferative syndrome. Admittedly, some cases of leukemia feature dysplastic changes in multiple cell lines and some cases emerge from previously diagnosed MDS. Others may have very few dysplastic features. These variants may represent different subclasses.

The purpose of our study was 2-fold. We wanted to apply the WHO diagnostic criteria to patients studied previously to assess the ability of several observers to apply the criteria consistently and to reassess outcome measurements in previous reports on patients to determine if the expanded number of diagnostic categories refines our ability to predict treatment response.

Our first goal was reached by having 3 separate observers, 2 pathologists and a clinical hematologist, review the peripheral blood and BM specimens. Peripheral blood count and cytogenetic data were then applied according to the WHO criteria. There was a significant interobserver agreement. Discrepancies among observers nearly always related to the identification and enumeration of dyspoiesis in neutrophils and megakaryocytes. Although we reached consensus by reviewing pathologic material jointly, subtle hypogranularity of neutrophils was sometimes disputed. There was usually consistent agreement on pseudo Pelger-Huet cells. The quantitation of 10% dysplastic cells in 2 or more myeloid cell lines was sometimes difficult. There was occasional disagreement on the dyspoietic changes and enumeration of megakaryocytes. When aspirates were inadequate, we relied on BM sections. Such discrepancies may best be resolved by obtaining fresh pathologic material (BM and peripheral blood) and ensuring adequate staining, especially for iron. Samples with borderline blast percentages should have 500 cell differential counts.

Borderline blast percentages changed the diagnosis in 5 cases (4.9%). In each instance we repeated 500 cell differential counts on BM samples to reach the final diagnosis. Although this seemed arbitrary, it was a practical solution. There was occasional difficulty identifying and enumerating RSs. Although we applied generally accepted criteria, the result depended heavily on the quality of the specimens. It is therefore recommended that at least 2 serial BM samples be taken in any patient with borderline values for marrow blasts, sideroblasts and dysplastic features, or when the BM specimen is of suboptimal quality.

Sixty-four of the cases had been entered into clinical trials using G-CSF and Epo by the Nordic MDS Group. Because these were low-risk patients (RA, RARS, and RAEB by FAB criteria), we cannot comment on the application of WHO criteria to other categories. The referees for the Nordic MDS Group applied the FAB criteria as understood from the original report and according to the standard use of these criteria in clinical studies and publications.\(^7\) Our reclassification applying WHO criteria factored those cases with multilineage dysplasia out of the low-risk group. Because the amount of dysplasia allowed by FAB was not entirely clear, one might argue that some cases that we classified as RCMD might be termed “unclassifiable” by FAB. However, unclassifiable MDS according to FAB has been used mainly for patients in whom it was not possible to estimate the percentage of blasts. For

| Table 4. Cases of MDS reclassified*† |
|-------------------------|---------|---------|
| WHO classification     | Original FAB classification |
| RA                      | RA, n = 24 | RARS, n = 22 | RAEB, n = 18 |
| RARS                    | 0        | 9        | 0        |
| RAEB-1                  | 0        | 0        | 12       |
| RAEB-2                  | 0        | 0        | 4        |
| RCMD                    | 13       | 0        | 0        |
| RCMD/RS                 | 0        | 11       | 0        |
| 5q-syndrome             | 7        | 1        | 0        |
| Unclassified            | 1        | 1        | 2        |

*Includes in clinical growth factor trials; see Hellström-Lindberg et al\(^{10}\) and Tehranchi et al\(^{26}\)
†Patients were reclassified according to FAB and then according to WHO. The table shows the reclassified FAB diagnoses. Four cases were difficult to reclassify to FAB and could not be reclassified according to WHO guidelines (i.e., there was no megakaryopoiesis available for evaluation). These patients are listed under their original FAB diagnosis.

| Table 5. Response rates by WHO classification* |
|-------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| RA                      | RARS, n = 8 | RCMD, n = 12 | RCMD/RS, n = 11 | RAEB-1, n = 11 | RAEB-2, n = 4 | 5q- n = 8 | Unclassified, n = 4 |
| CR                      | 2        | 2        | 1        | 0        | 2        | 0        |
| PR                      | 0        | 2        | 2        | 0        | 4        | 0        |
| NR                      | 1        | 2        | 6        | 10       | 8        | 2        | 4        | 2        |

RARS versus RCMD/RS (CR + PR), \( P = .003 \).
*For response criteria, see “WHO classification and outcome of GF treatment” and Hellström-Lindberg et al\(^{10}\) and Cheson et al\(^{28,29}\)
purposes of comparison with other clinical studies using the FAB classification we allowed the originally refereed diagnoses to stand. We reassigned our 64 cases to the appropriate WHO categories and classification we allowed the originally refereed diagnoses to stand. Our approach is exemplified by RARS and RAEB-2. A challenge is to learn how to apply the WHO criteria widely. The present study suggests that in patients with less than 5% marrow blasts, a very important negative variable for response to proerythroid GF treatment as well as for survival is nonerythroid multilineage dysplasia. The presence of multilineage dysplasia apparently lessens the chance of response to therapy, in particular in cases with RS. The difference between RARS and RCMD/RS was more pronounced (P = .003) than for patients with unilineage versus multilineage dysplasia in general. A significant difference in survival and leukemic transformation between RARS and RCMD/RS has been shown previously, and our results underline that this also translates into a significant difference in response to GF treatment. Thus, a thorough analysis of dysplastic features should be recommended in MDS with less than 5% marrow blasts before initiation of GF treatment, and this analysis should be weighted together with other predictive variables, such as IPSS and predictive G-CSF–Epo score.

We conclude that the WHO criteria can be applied successfully with good interobserver agreement. The importance of good specimens and staining was emphasized. In doubtful cases, a second BM biopsy and clinical observation may be necessary. Concordance might be further improved if the definitions of dysplastic granulocytes and megakaryocytes could be refined. One drawback of the classification system is the increased number of diagnostic categories. This requires more patients per study to make application of statistics meaningful. Using the WHO criteria for classifying MDS refines the ability to predict treatment response as exemplified by RARS and RCMD/RS. Patients likely to show complete responses are also identified. We recommend the broad application of the WHO MDS criteria, both to assess its utility further and to aid in the interpretation of clinical trials.

Appendix 1
Participating members of the Nordic MDS Group were from the Swedish and Norwegian Centers, as follows: Inger-Marie Dahl, Ingunn Dybedal, and Jon Magnus Tangen (Norway); and Petar Antunovic, Jan Astermark, Lena-Maria Engström, Eva Hellström-Lindberg, Olle Linder, Lars Nilsson, Herman Nilsson-Ehle, and Gunnar Öberg (Sweden).

References


The WHO classification of MDS does make a difference

Robert B. Howe, Anna Porwit-MacDonald, Robert Wanat, Ramin Tehranchi and Eva Hellström-Lindberg