Reproducibility of the World Health Organization 2008 criteria for myelodysplastic syndromes

by Leonor Senent, Leonor Arenillas, Elisa Luño, Juan Carlos Ruiz, Guillermo Sanz, and Lourdes Florensa

Haematologica 2012 [Epub ahead of print]

doi:10.3324/haematol.2012.071449

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors’ final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.

Haematologica (pISSN: 0390-6078, eISSN: 1592-8721, NLM ID: 0417435, www.haematologica.org) publishes peer-reviewed papers across all areas of experimental and clinical hematology. The journal is owned by the Ferrata Storti Foundation, a non-profit organization, and serves the scientific community with strict adherence to the principles of open access publishing (www.doaj.org). In addition, the journal makes every paper published immediately available in PubMed Central (PMC), the US National Institutes of Health (NIH) free digital archive of biomedical and life sciences journal literature.
Reproducibility of the World Health Organization 2008 criteria for myelodysplastic syndromes

Short title: Reproducibility of WHO 2008 criteria for MDS

Leonor Senent, 1,2a Leonor Arenillas, 1,3a Elisa Luño, 1,4 Juan C. Ruiz, 5
Guillermo Sanz, 1,2 and Lourdes Florensa 1,3

1On behalf of the Spanish MDS cooperative group (GESMD); 2Servicio de Hematología, Hospital Universitario La Fe, Valencia, Spain; 3Laboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar, GRETNHE, IMIM (Hospital del Mar Research Institute), Barcelona, Spain; 4Servicio de Hematología, Hospital Central de Asturias, Oviedo, Spain, and 5Departamento de Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Universidad de Valencia, Valencia, Spain

These authors contributed equally to this work

Correspondence
Leonor Senent, Servicio de Hematología, Hospital Universitario La Fe, Torre F, Planta 7, Bulevar Sur s/n, 46026, Valencia, Spain. Phone: international +34.96.1245876.
Fax: international +34.96.1246201. E-mail: senent_leo@gva.es.
Abstract

Background
The reproducibility of the World Health Organization 2008 classification for myelodysplastic syndromes is uncertain and its assessment was the major aim of this study.

Design and methods
The different peripheral blood and bone marrow variables required for an adequate morphological classification were blindly evaluated by four cytomorphologists in samples from 50 patients with myelodysplastic syndromes. The degree of agreement among observers was calculated using intraclass correlation coefficient and generalized kappa statistic for multiple raters.

Results
The degree of agreement for the percentages of blasts in bone marrow and peripheral blood, ring sideroblasts in bone marrow, and erythroid, granulocytic and megakaryocytic dysplastic cells was strong ($P < .001$ in all instances). After stratifying the percentages according to the categories required for the assignment of World Health Organization subtypes, the degree of agreement was not statistically significant for cases with $5 – 9\%$ blasts in bone marrow ($P = .07$) and $0.1 – 1\%$, in peripheral blood ($P = .47$) as well as for the percentage of erythroid dysplastic cells ($P = .49$). Finally, the interobserver concordance for World Health Organization -defined subtypes showed a moderate overall agreement ($P < 0.001$), being the reproducibility lower for cases with refractory anemia with excess of blasts type 1 ($P = .05$) and refractory anemia with ring sideroblasts ($P = .09$).

Conclusions
The reproducibility of the World Health Organization 2008 classification for myelodysplastic syndromes is acceptable but definition of blast cells and features of erythroid dysplasia needs to be properly refined.
Introduction

For more than 20 years the French-American-British morphologic classification was the base for the diagnosis of myelodysplastic syndromes (MDS), a group of acquired clonal hematopoietic stem cell disorders with very heterogeneous outcomes and characterized by ineffective hematopoiesis, cytopenias, dysplastic morphologic features, and an increased risk of development of acute myeloid leukemia (AML)\(^1,2\). In an attempt to improve its prognostic value, to incorporate other relevant morphological and biological prognostic characteristics, such as grade of myelodysplasia and cytogenetics, and to redefine the border between MDS and AML, an expert panel of the World Health Organization (WHO) proposed in 2001 a new classification system for MDS\(^3\) that was refined in 2008\(^4\). Taking into account the type and number of cytopenias, percentage of cells with dysplastic changes in the different myeloid cell lineages, percentage of blasts in peripheral blood (PB) and bone marrow (BM), percentage of ring sideroblasts in BM, absolute monocyte count and conventional cytogenetics, the revised 2008 WHO classification recognizes 7 subcategories of MDS that are shown in Table 1 \(\text{(modified from Vardiman JW et al.}\)^5\(\)). Although the usefulness of the WHO classification was initially criticized\(^6,7\), it has gained widespread acceptance\(^1,8\). The 2001 WHO classification prognostic value is clearly superior than the FAB classification\(^9,10\) and has been incorporated into the recently defined WHO classification-based Prognostic Scoring System (WPSS)\(^11,12\). However, the reproducibility of the WHO classification is uncertain. The only published study on interobserver agreement according to the 2001 WHO criteria showed an acceptable reproducibility\(^13\). A recent study has evidenced a 20% discrepancy in the WHO 2008 classification-based diagnosis between referring and tertiary care centers\(^14\). The concordance between observers of the 2008 WHO criteria has never been addressed. This issue is not only academically but clinically relevant.

The aim of this study was to assess the reproducibility of the WHO 2008 morphological classification. For this purpose PB and BM samples from 50 patients with MDS were blindly and independently reviewed by four cytomorphologists from three different referral centers with expertise in the diagnosis of MDS. The interobserver concordance in the quantification of the three morphologic characteristics required for the assignment of patients to the appropriate WHO morphological subtype (the number of blast cells in PB and BM, percentage of ring sideroblasts in BM and myelodysplastic features) was evaluated.
Design and Methods

Patients and samples

Samples of BM aspirates and PB from 50 patients with a clearly established diagnosis of MDS according to WHO 2008 criteria and diagnosed at two of the participating centers (Hospital del Mar, Barcelona and Hospital Universitari i Politecnic La Fe, Valencia) were included in this retrospective analysis. In all instances the analyzed samples had been obtained at initial evaluation. The number of cases of the different WHO 2008 morphologic subtypes selected for review was prefixed according to their expected incidence in previous studies. Preference for inclusion in the study was given to cases diagnosed in more recent years and with good quality samples available. Table 2 summarizes the main characteristics of the patients.

Morphologic studies

Four smears from each included patient in the study were available for blind and independent microscopical review by four experienced cytologists from three centers. Two BM and one PB May-Grünwald-Giemsa smears were used for assessing percentages of blasts in PB and BM and percentages of dysplastic cells of the three myeloid cell lines. An additional Prussian blue stained BM smear was used for assessing the percentage of ring sideroblasts. The cytologists had a meeting to discuss the evaluation of dysplasia and diagnosis using training slides. Blasts and ring sideroblasts were defined according to the recent consensus proposals of the International Working Group on Morphology of Myelodysplastic syndrome (IWGM-MDS). The WHO 2008 recommendations for evaluating the morphologic diagnosis of MDS were strictly followed. Thus, BM blast counts were calculated as percentages of all BM nucleated cells. PB and BM differential counts were performed on at least 200 and 500 cells, respectively. Evaluation of dysplasia was based on morphological criteria described in the 2008 WHO publication. Briefly, the following morphologic features of dysplasia were evaluated: 1) dyserythropoiesis: nuclear budding, internuclear bridging, karyorrhexis, multinuclearity, nuclear hyperlobation, megaloblastic changes, stippling basophil, ring sideroblasts and vacuolization, 2) dysgranulopoiesis: nuclear hypolobation (pseudo Pelger-Huët), irregular hypersegmentation, agranularity, pseudo Chediak-Higashi granules, Auer rods and Döhle bodies, and 3) dysmegakaryocytopoiesis: micromegakaryocytes, nuclear hypolobation, and multinucleation. As defined in the WHO 2008 classification, the threshold used for considering a myeloid cell line as dysplastic was the presence of ≥ 10% abnormal cells in the corresponding myeloid lineage. To assess dysplasia at least
200 neutrophils, 200 erythroid precursors and 30 megakaryocytes were evaluated in BM. Information on hemoglobin level, absolute neutrophil and platelet counts was available for observers when performing the morphologic review. In contrast, the observers were blinded to the clinical and cytogenetic data. Consequently, for the purpose of this study, cases of MDS associated with isolated 5q deletion were classified into other MDS morphologic subtypes. All the morphologic characteristics analyzed were recorded in specific forms designed for this purpose by the Spanish Group on Myelodysplastic Syndromes (GESMD) and transferred into a specific database. Once finished the review, no attempt was made to reach a consensus agreement on cases with discrepant results in any of the variables analyzed. The local ethics committees approved the studies which followed the 2000 revision of the Helsinki Declaration.

**Statistical analysis**

Agreement between the four observers for continuous quantitative variables (percentages of blasts in PB and BM, percentages of ring sideroblasts in BM and percentages of dysplastic cells of erythroid, granulocytic and megakaryocytic lineages) was evaluated using the intraclass correlation coefficient (ICC)\(^{16}\). The ICC has advantages over Spearman correlation coefficient, because it is adjusted for the effects of the scale of measurements, and allows the assessment of agreement when there are more than two observers. Quantitative variables (percentages of blast cells in PB and BM, percentages of ring sideroblasts in BM and percentage of dysplastic cells in each myeloid lineage) were categorized and also evaluated as categorical variables (Table 3). For this purpose, we used the cutoff levels defined by the WHO classification and an additional cutoff level of 40% of dysplastic cells according to previous data\(^{17}\). To evaluate the concordance between observers in qualitative and categorized quantitative variables, the generalized \textit{kappa} statistic for multiple raters (\(\kappa\)) was calculated. An ICC or generalized \(\kappa\) statistic value of 1 denotes complete agreement between the different observers, while an ICC value of 0 denotes agreement equivalent to chance. Both ICC and generalized \(\kappa\) statistic can be interpreted as follows: 0-0.2 indicates poor agreement; 0.3-0.4 indicates fair agreement; 0.5-0.6 indicates moderate agreement; 0.7-0.8 indicates strong agreement; and >0.8 indicates almost perfect agreement \(^{18}\). The statistical package SPSS, version 19.0 (SPSS Inc., Chicago, IL, USA) was used for ICC calculation and a Microsoft Excel Template to calculate the Kappa statistic.
Results
The degree of concordance between observers for the different morphological characteristics is summarized in Table 3.

Interobserver concordance in blast cells count
There was a strong agreement in the percentage of blast cells in BM considered as a continuous variable. The ICC for this parameter was 0.95 (95% confidence interval [CI], 0.92-0.97; \( P < 0.001 \)). When its degree of concordance was assessed stratifying the variable into 3 categories (<5%, 5 to 9%, and ≥ 10%), according to the thresholds used in WHO classification subtypes, the interobserver concordance was moderate (overall \( \kappa \), 0.57; \( P < 0.001 \)). The degree of agreement was higher and significant when the BM blast percentage was less than 5% (\( \kappa \), 0.72; \( P < 0.001 \)) or equal or greater to 10% (\( \kappa \), 0.65 \( P < 0.001 \)), but it was lower for cases with an intermediate percentage of BM blast cells (5% to 9%) where only a fair agreement was reached (\( \kappa \), 0.29; \( P = 0.07 \)). When this variable was further stratified into four categories, adding a cut-off point of 2% blast cells in BM, the interobserver concordance was fair (overall \( \kappa \) 0.42; \( P < 0.001 \)). The \( \kappa \) values were 0.50 (\( P = 0.002 \)) for cases with less or equal to 2%, 0.28 (\( P = 0.065 \)) for cases with more than 2 and less than 5%, 0.29 (\( P = 0.048 \)) for cases with 5 to less than 10% and 0.60 (\( P < 0.001 \)) for cases with more or equal than 10%.

Interobserver agreement for the percentage of blast cells in PB showed a very good agreement. The ICC for this parameter was 0.82 (95% CI, 0.72-0.89). When the variable was evaluated according to the subcategories used in the WHO classification (absence, less or equal to 1% and more than 1%), the overall kappa score was 0.30 (\( P < 0.002 \)). The agreement between observers was significant in the condition with more than 1% blasts (\( \kappa \), 0.37 \( P = 0.009 \)), but there was no significant agreement between observers in the condition without blasts (\( \kappa \), 0.37; \( P = 0.20 \)) and in the intermediate category of less or equal to 1% blasts (\( \kappa \), 0.09; \( P = 0.47 \)).

Interobserver concordance in ring sideroblasts count
The percentage of ring sideroblasts in BM showed a nearly perfect agreement between observers both analyzed as a continuous variable (ICC, 0.96; 95% CI, 0.93-0.98; \( P < 0.001 \)) and with the 15% cut-off point used in WHO criteria (\( \kappa \), 0.82; \( P < 0.001 \)).
**Interobserver concordance in assessment of dysplasia**

When the degree of dysplasia of the three different hematopoietic cell lines was studied as a continuous variable, the degree of concordance between observers was strong and almost perfect for megakaryocytic lineage with an ICC of 0.91 (95% CI, 0.85-0.95; \( P < 0.001 \)) and for granulocytic lineage with an ICC of 0.89 (95% CI, 0.83-0.94; \( P < 0.001 \)). A substantial agreement was observed for erythroid lineage with an ICC of 0.75 (95% CI, 0.60-0.85; \( P < 0.001 \)). When those variables where stratified according to the 10% cut-off point required by WHO criteria to define a hematopoietic cell lineage as dysplastic, the interobserver agreement was statistically significant for the granulocytic (\( \kappa, 0.40; P=0.04 \)) and megakaryocytic (\( \kappa, 0.49; P < 0.001 \)) lineages. In the erythroid lineage there was poor agreement (\( \kappa, 0.19; P=0.49 \)). When a cut-off point of 40% dysplastic cells was used, the concordance between raters improved for the megakaryocytic and granulocytic lineages but not improved for the erythroid lineage.

**Reproducibility of WHO-defined MDS subtypes**

The overall interobserver concordance for WHO-defined MDS subtypes showed a moderate overall agreement (\( \kappa, 0.43; P < 0.001 \)) (Table 3 and Figure 1). A greater reproducibility was found for patients with RAEB-2 (\( \kappa, 0.60, P < 0.001 \)), RCUD (\( \kappa, 0.5; P < 0.001 \)) and RCMD (\( \kappa, 0.46; P < 0.01 \)). Concordance was lower for RARS (\( \kappa, 0.26, P = 0.09 \)) and RAEB-1 (\( \kappa, 0.29; P = 0.05 \)).

**Discussion**

Despite major advances in the diagnosis of hematologic diseases, cytomorphologic diagnosis remains the cornerstone in the diagnosis of myelodysplastic syndromes (MDS). The current study was designed to evaluate interobserver variability to assign a diagnosis of MDS according to the WHO 2008 classification criteria and to define potential morphological difficulties. In our study, we observed a moderate reproducibility of the WHO 2008 classification.

Mufti et al. (2008)\(^{15} \) pointed out the difficulty of morphological diagnosis of blast cells, however the percentage of blasts in BM is one of the main known prognostic factor\(^{7,19-20} \) and which has been included in the most commonly prognostic scoring systems for MDS, such as the International prognostic scoring system (IPSS) and WPSS. In our work we have an almost perfect agreement in BM blast cells count in cases with < 5% or with \( \geq 10\% \). In those cases with \( \geq 5\% \) and < 10% blast cell count, the rate of concordance showed a moderate agreement. This implies that one patient could be classified as RCMD, RAEB-1 or RAEB-2 depending on the observer. We decided to
evaluate the interobserver concordance for an additional cut-off point of 2% blasts in bone marrow for cases without excess of blasts because this threshold seems to portray prognostic relevance in the revised version of the IPSS. The degree of agreement was adequate for cases with ≤ 2% blast cells but, again, it was not as good for cases between 2% and 5%. Discrepancies in the blast count in BM of patients with myeloid malignancies, including MDS, is partly due to the difficulty in distinguishing between granular blast cells and promyelocytes and the irregular distribution of blast cells in BM. In spite of a good correlation between the percentage of blasts determined by morphologic examination and percentage of CD34+ cells determined by flow cytometry is usually observed, blast enumeration by morphology is the gold-standard method.

The correct assignment of the percentage of blasts in PB is also crucial for a correct diagnosis and classification of patients. In our study, we found a fair agreement in PB blast cell count. This result may be due to the low level of blast cells present in PB. Interobserver discrepancies may best be resolved by increasing the number of cells in the differential counts.

The recognition of dysplastic signs has a crucial value not only for diagnosis and classification of MDS patients, but also a prognostic role in low risk MDS patients. In this regard, Pseudo-Pelger-Huët anomaly in neutrophils and micromegakaryocytes had been correlated with overall survival. Besides, several investigators support that cases with multiligneage dysplasia have a less favorable prognosis than those with only dyserythropoietic dysplasia. Thus, the WHO classification separated cases with refractory anemia in the previous FAB classification into two categories depending on the presence or absence of multiligneage dysplastic features. This distinction has been criticized by some groups, because in clinical practice the assessment of the features of dysplasia is not always easy due to the lack of definition of objective parameters. Poor technical quality of the specimen could also be an obstacle to an accurate diagnosis of dysplasia. In our work we found a moderate but significant interobserver agreement for megakaryocytic and granulocytic dysplasia and a poor agreement for erythroid dysplasia. This is probably because features of dysgranulopoiesis (Pseudo-Pelger-Huët, hypogranularity) and dysmegakaryopoiesis (micromegakaryocytes, non-lobulated nuclei and multiple widely separated nuclei) are less subjective and more reproducible than features of erythroid dysplasia.

WHO classification includes a uniform threshold of 10% for dysplasia in each myeloid lineage; however, as discussed in the paper of Parmentier et al., this level of dysplasia is highly questionable, it is particularly low in the megakaryocytic lineage where the number of cells analyzed is smaller than in the other series. We analyzed
the interobserver concordance with a cut-off point of 40% dysplastic cells and we found that the agreement improved in the megakaryocytic and granulocytic lineage but not in the erythroid lineage. These results agree with those by Matsuda et al.\textsuperscript{17} and Germing et al.\textsuperscript{28} who proposed to raise the threshold of dysmegakaryopoiesis from 10 to 40%.

The prognostic value of the WHO classification is already known.\textsuperscript{7,9,10,12} Howe et al.\textsuperscript{13} analyzed the reproducibility of the 2001 WHO classification showing a 92% of agreement among three reviewers. Their discrepancies were related to the identification and enumeration of dyspoiesis in neutrophils and megakaryocytes and in those cases with borderline blast percentages.

Recently, Naqvi et al.\textsuperscript{14} analyses the discrepancies in morphologic diagnosis of MDS between referral and tertiary centers setting up morphological differences in 12% of patients. However, they did not analyze the causes for that discrepancy.

The current work, although reviewing a rather limited number of samples, is the first to analyze the correlation between observers of the 2008 WHO morphological criteria. We found a nearly moderate and significant concordance to define the 2008 WHO MDS subtypes ($\kappa$, 0.43; $P < 0.001$). Most differences concerned the distinction of unilineage and multilineage dysplasia; consequently some patients were classified as RARS or RCMD depending on the recognition of dysplasia in one or more myeloid lineages. As previously described by Howe et al.\textsuperscript{13}, we also had difficulties in assigning MDS subtype in those cases with borderline blast cell percentages. In fact, a substantial agreement was obtained only in cases with less than 2% or 5% or more than 10% of blast cells in BM.

To sum up, the WHO 2008 classification can be applied with a moderate interobserver concordance. Discrepancies are frequent and may have a potential negative impact in the assignment of prognosis and therapy planning in the individual patient. The degree of agreement could improve if dyserythropoiesis features are refined. Future studies should evaluate the potential increment in the threshold for considering a cell lineage as dysplastic in order to enhance the recognition of multilineage dysplasia. Finally, MDS diagnosis is complex, requires an accurate application of the WHO criteria, and should be performed by experienced morphologists. Despite all those measures, it must be highlighted that the value of cytomorphology alone for the classification of MDS is limited. In this regard, the development, standardization, and incorporation into our daily practice of other techniques, such as flow cytometry and molecular studies, will likely allow us in the near future to better diagnose, characterize, and classify this heterogeneous group of myeloid neoplasms.
References


Acknowledgments
The authors wish to thank Dr. Ignacio Lorenzo for his helpful comments on statistical issues and Luis Benlloch for data management.

Funding
This work was supported in part by research funding from “RTICC” (Red Temática de Investigación Cooperativa en Cáncer, RD06/0020/0031, RD07/0020/2004), Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo, Spain (FI07/00107; PI07/1009 and PI11/02010); SGR 541 (“Agència de Gestió d'Ajuts Universitaris i de Recerca”, “Departament d'Innovació, Universitats i Empresa”) y COST-European Genetic and Epigentic Study on AML ans MDS (BM080).

Authorship and Disclosures
Conception and design: LS, LA, EL, GS, and LF. Collection of samples: LS, LA, and LF; Morphological review: LS, LA, EL, and LF; Statistical analysis: JCR; Data interpretation: All authors; Manuscript writing: LS, LA, EL, GS, and LF; Final approval of manuscript: All authors. The authors have no financial conflict of interest to declare.

Figure legends
Figure 1.
The presence of granulated blast cells (arrow) makes difficult the distinction between blast cells and promyelocytes (discontinous arrow) so that the number of blast cells may differ and the same patient be classified as MDS with or without excess of blasts.

Figure 2.
Evaluation of dysplastic features in erythropoiesis such as megaloblastoid changes (arrow) and cytoplasmic changes (discontinous arrow) are poorly reproducible and this justifies that the agreement between observers in evaluation of dyserythropoiesis is not good.
Figure 1
Figure 2.