High concordance in grading reticulin fibrosis and cellularity in patients with myeloproliferative neoplasms

Olga Pozdnyakova[1], Kaida Wu[2], Abhay Patki[2,5], Scott J Rodig[1], Juergen Thiele[3] and Robert P Hasserjian[4]

[1]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; [2]Sanofi Oncology, Cambridge, MA, USA; [3]University of Cologne, Cologne, Germany and [4]Department of Pathology, WRN244, Massachusetts General Hospital, Boston, MA, USA

The myeloproliferative neoplasms primary myelofibrosis, polycythemia vera, and, rarely, essential thrombocythemia are characterized by variable degrees of bone marrow fibrosis, either at presentation or upon progression. The increasing use of emerging therapies that may alter disease biology and morphology demands accurate and reproducible assessment of fibrosis grade. To assess concordance of hematopoietic cellularity and fibrosis grading, three hematopathologists independently evaluated a total of 728 bone marrow biopsies from 261 patients with myeloproliferative neoplasms on three clinical trials using fedratinib (SAR302503), a *JAK2* inhibitor, including 249 taken at baseline and 479 on therapy. Concordance between the pathologists was evaluated by Pearson correlation coefficient (cellularity) and unweighted kappa statistic (fibrosis grade). There was high correlation of cellularity assessment ($r = 0.92$) and fibrosis grading (kappa $= 0.83$) between the three pathologists. Concordance with World Health Organization (WHO) grade 3 samples was higher compared with grades 0, 1, and 2. Concordance of fibrosis grading in pretreatment samples was superior to that of post-treatment samples (kappa $= 0.83$ and 0.79, respectively, $P = 0.023$). Our analysis suggests that the updated 2008 WHO reticulin fibrosis grading system is highly reproducible, even in patients undergoing *JAK2* inhibitor therapy. This system is practically applicable to establish baseline fibrosis grade as well as changes in fibrosis in subsequent samples on therapy.
Modern Pathology (2014) 27, 1447–1454; doi:10.1038/modpathol.2014.69; published online 25 April 2014

Primary myelofibrosis, essential thrombocythemia, and polycythemia vera are myeloproliferative neoplasms that are associated with mutation in the *JAK2* gene and, less often, with mutations in *MPL* or *CALR* genes.[1–5] These myeloproliferative neoplasms have highly variable propensities to either present with or to develop progressive bone marrow fibrosis: this fibrosing process is intrinsic to primary myelofibrosis and affects a significant subset of polycythemia vera patients and a very small subset of essential thrombocythemia patients.[6–12]

Bone marrow fibrosis is measured by reticulin and trichrome stains of the bone marrow trephine biopsy and is graded on a semiquantitative scale.[13] The classification proposed by the World Health Organization (WHO) and updated in 2008 employs the European Consensus System to grade bone marrow fibrosis, which has four possible grades (0, 1, 2, and 3); distinction between these levels is based on the number, density, and thickness of silver-stained reticulin fibers as assessed by the pathologist (Table 1).[14,15] Previous systems used to grade bone marrow fibrosis included the Manoharan and Bauermeister schemes.[13,14]

Reticulin grade is not included among the factors used in primary myelofibrosis risk stratification systems, such as the Dynamic International Prognostic Scoring System[16,17] and shows variable correlation with clinical parameters of disease progression, such as splenomegaly. However, increased

Correspondence: Dr RP Hasserjian, MD, Department of Pathology, WRN244, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA.
E-mail: rhasserjian@partners.org
[5]Current address: Blueprint Medicines, Cambridge, MA, USA.

reticulin grade is an adverse prognostic factor in primary myelofibrosis, essential thrombocythemia, and polycythemia vera.[18–21] The level of reticulin fibrosis is included among the criteria used to establish a diagnosis of post-polycythemic myelofibrosis and post-thrombocythemic myelofibrosis, which represent clinically progressed stages of polycythemia vera and essential thrombocythemia, respectively.[22] Moreover, bone marrow fibrosis is a dynamic process and has been shown to resolve over time following eradication of the neoplastic clone by allogeneic stem cell transplantation in primary myelofibrosis.[23–25] Regression of bone marrow fibrosis is also associated with successful treatment of other diseases associated with bone marrow fibrosis, such as autoimmune myelofibrosis and chronic myelogenous leukemia.[26–28]

The *JAK1/JAK2* inhibitor ruxolitinib is a novel agent that has been shown to ameliorate clinical symptoms and splenomegaly and prolong survival in patients with primary myelofibrosis, post-thrombocythemic myelofibrosis and post-polycythemic myelofibrosis.[29–35] Although prior therapies used to treat primary myelofibrosis (hydroxyurea and interferon) have shown little or no effect on bone marrow fibrosis grade,[36,37] treatment with ruxolitinib has been associated with reduction and even complete resolution of bone marrow fibrosis.[38,39] The selective *JAK2* inhibitor fedratinib has also shown efficacy in treating patients with primary myelofibrosis, post-thrombocythemic myelofibrosis, and post-polycythemic myelofibrosis and similarly shows evidence of reduced reticulin fibrosis in treated patients.[40] This accumulating evidence suggests that reticulin grade may represent an additional parameter to measure treatment response, in conjunction with existing clinical and biological parameters such as splenomegaly, symptoms, peripheral counts, and *JAK2* allele burden.[30,31,34,35,41] Indeed, a reticulin fibrosis grade of ≤ 1 (as well as normal age-adjusted normocellularity for primary myelofibrosis and polycythemia vera) is included among required criteria for a complete response to therapy in primary myelofibrosis, essential thrombocythemia, and polycythemia vera patients according to the recently published European LeukemiaNet guidelines.[42,43] However, there has been some controversy as to the reproducibility of reticulin grading and its applicability to routine diagnosis setting outside of a clinical trial setting.[44–48] In order to address this issue, we examined the concordance of reticulin grading in a large number of patients with primary myelofibrosis, post-thrombocythemic myelofibrosis, and post-polycythemic myelofibrosis, in the context of clinical fedratinib trials. Our goal was to determine the reproducibility of the WHO-adopted reticulin grading system in patients before and on treatment with a *JAK2* inhibitor. We also assessed the reproducibility of cellularity estimation in the same samples.

Materials and methods

Patients

The study included 261 patients with myeloproliferative neoplasms on three fedratinib (SAR302503) trials: one Phase 1/2 trial (38 patients), one Phase 2 trial (26 patients), and one Phase 3 trial (197 patients). The diagnoses were primary myelofibrosis in 120 patients (46%), post-polycythemia myelofibrosis in 54 patients (21%), and post-thrombocythemic myelofibrosis in 19 patients (7%). In 68 patients (26%), the myeloproliferative neoplasm subtype before study entry was not available. Before enrollment on the studies, 70% of patients had been treated with hydroxyurea and 5% of patients had received other cytotoxic agents. A total of 728 bone marrow biopsies were evaluated, with 249 biopsies taken at baseline (pre-fedratinib) and 479 biopsies taken on therapy.

Bone Marrow Examination

Bone marrow biopsies were examined for hematopoietic cellularity and fibrosis grade based on the hematoxylin and eosin, silver impregnation reticulin, and trichrome stains. For the Phase 3 and Phase 2 studies, hematoxylin and eosin, reticulin (Gordon and Sweet method), and trichrome (Masson with Wiegert Hematoxylin and Biberich) stains were performed at the Covance Central Laboratory (Indianapolis, IN, USA) according to the manufacturer's protocols. For the Phase 1/2 study, hematoxylin and eosin, reticulin, and trichrome stains were performed locally at different US institutions where the biopsies had originated. Trichrome stains were available for 659 biopsies (90.5% of total biopsies evaluated). Hematopoietic cellularity included only cells of the myeloid, erythroid, megakaryocytic, and lymphoid lineages; stromal cells were explicitly excluded from the cellularity assessment. Hematopoietic cellularity was scored in 10% increments from 0 to 100%. Reticulin fibrosis was graded according to the 2008 updated WHO classification guidelines, from grades 0 to 3 (Table 1). As recommended, reticulin fibers were graded only in areas of hematopoiesis. On the basis of agreement between the three hematopathologists before slide review, biopsies showing heterogeneous morphology with variability in reticulin fibers were graded according to the predominant reticulin grade (>50% of hematopoietic area). For example, if a 60% of the biopsy demonstrated grade 1 and 40% of the biopsy showed grade 2, the final assigned grade was grade 1.

The three hematopathologists (OP, JT, and RH) examined the same slide under a multi-headed microscope and independently scored hematopoietic cellularity and reticulin fibrosis grade. The pathologists were blinded to the patient information (with the exception of patient age and gender) and

Fibrosis/cellularity grading in bone marrow
O Pozdnyakova et al
1449

Table 1 WHO 2008 criteria for grading of reticulin fibrosis[a]

Grade	Description
MF-0	Scattered linear reticulin with no intersections (crossovers), corresponding to normal bone marrow
MF-1	Loose network of reticulin with many intersections, especially in perivascular areas
MF-2	Diffuse and dense increase in reticulin with extensive intersections, occasionally with focal bundles of collagen and/or focal osteosclerosis
MF-3	Diffuse and dense reticulin with extensive intersections and coarse bundles of collagen, often associated with osteosclerosis

[a]Adapted from WHO Classification of Tumors Hematopoietic and Lymphoid Tissue.[15]

treatment status. The final grade was based on agreement among at least two of the pathologists. If all the three pathologists disagreed as to grade or if disagreement with any pathologist was greater than 1 grade, the final grade was established by consensus following discussion among all the three pathologists.

Statistical Analysis

Concordance of cellularity between the three pathologists was evaluated by Pearson correlation coefficient (AnalystSoft, StatPlus:mac LE—free statistical analysis program for Mac OS. Version 2009; www.analystsoft.com). Concordance of fibrosis grade was evaluated by unweighted kappa statistics and Fisher's exact test (Richard Lowry 2001–2013; www.vassarstats.net). The study protocol was approved by the corresponding institutional research ethics committees and/or institutional review boards before enrollment of patients, and the study was conducted in accordance with the principles set forth by the Declaration of Helsinki.

Results

Concordance of Hematopoietic Cellularity

Hematopoietic cellularity could be assessed in 692/728 biopsies (95.1%). Cellularity could not be evaluated in 36 biopsies because of inadequate/insufficient material for accurate assessment—that is, limited biopsy sample or severe crush artifact. The correlations of cellularity assessment among the three pathologists are shown in Table 2. The average correlation coefficient r value was 0.917. Complete agreement in cellularity assessment between the three pathologists (same or $\pm 10\%$ cellularity) was observed in 565/692 biopsies (81.7%), whereas in 127 biopsies cellularity discordance between at least two of the pathologists was $>20\%$.

Concordance of Fibrosis Grading

Fibrosis could be graded in 665/728 biopsies (91.3%). Fibrosis could not be evaluated in 63 biopsies because of a failed reticulin stain, absence of hematopoietic areas, or inadequate tissue sample. Distribution of the final fibrosis grades assigned by

Table 2 Concordance of hematopoietic cellularity assessment ($n = 692$) between the three pathologists

	r^a
Pathologist 1 *versus* pathologist 2	0.92
Pathologist 1 *versus* pathologist 3	0.90
Pathologist 2 *versus* pathologist 3	0.93

[a]r, Correlation coefficient.

the pathologists was as follows: grade 0: 21 biopsies (3.2%), grade 1: 90 biopsies (13.5%), grade 2: 234 biopsies (35.2%), and grade 3: 320 biopsies (48.1%). Trichrome staining (available in 90.5% of samples) aided the fibrosis grading in which small, often patchy areas corresponding to bundles of thick reticulin fibers were found in grade 2 fibrosis, whereas confluent and extensively stained depositions were characteristically observed in grade 3 fibrosis (Table 1). In 23 cases where reticulin stain was suboptimal, trichrome stain revealed extensive collagen deposition and helped to confirm grade 3 fibrosis. Conversely, trichrome stain results did not change the final fibrosis grade in any cases with adequate reticulin stain. Figure 1 shows representative hematoxylin and eosin and reticulin stains for different fibrosis grades in which all the three pathologists agreed. For fibrosis grading, all three agreed on 552 biopsies (83.0%), two of the three agreed on 111 biopsies (16.7%), and the grading required consensus discussion in two biopsies (0.3%). The concordance of fibrosis grading between the pathologists is shown in Table 3. The unweighted kappa coefficient was equal to or greater than 0.8 that, based on the standards for strength of agreement of the kappa coefficient, qualifies as substantial (0.61–0.80) or almost perfect (0.81–1) agreement according to Landis and Koch.[49] Figure 2 shows frequencies of agreement on individual grades among each pair of pathologists. The highest rate of agreement was observed in grade 3 cases (93.1, 94.0, and 94.4% for all pathologist pairs). For grades 0, 1, and 2, the rate of agreement was 85.3–88.5%. Considering all pathologist pairs together, the rate of agreement for grade 3 was significantly higher than the agreement for grade 2 ($P<0.0001$), grade 1 ($P=0.0001$), and grade 0 ($P=0.03$), whereas there was no significant difference

in the rate of agreement between grade 0 and 1, grade 0 and 2, and grade 1 and 2 ($P > 0.05$ for all comparisons). Regarding the two biopsies requiring consensus discussion, in one case all the three pathologists assigned a different grade (grade 1, grade 2, and grade 3) and in the other case there was a two-grade difference with one of the pathologists (grade 1, grade 1, and grade 3). Both of these biopsies had been taken from patients undergoing treatment with fedratinib (for 24 and 96 weeks) who demonstrated significant heterogeneity in reticulin fibers.

Concordance of Fibrosis Grading in Pre- and Post-Fedratinib Groups

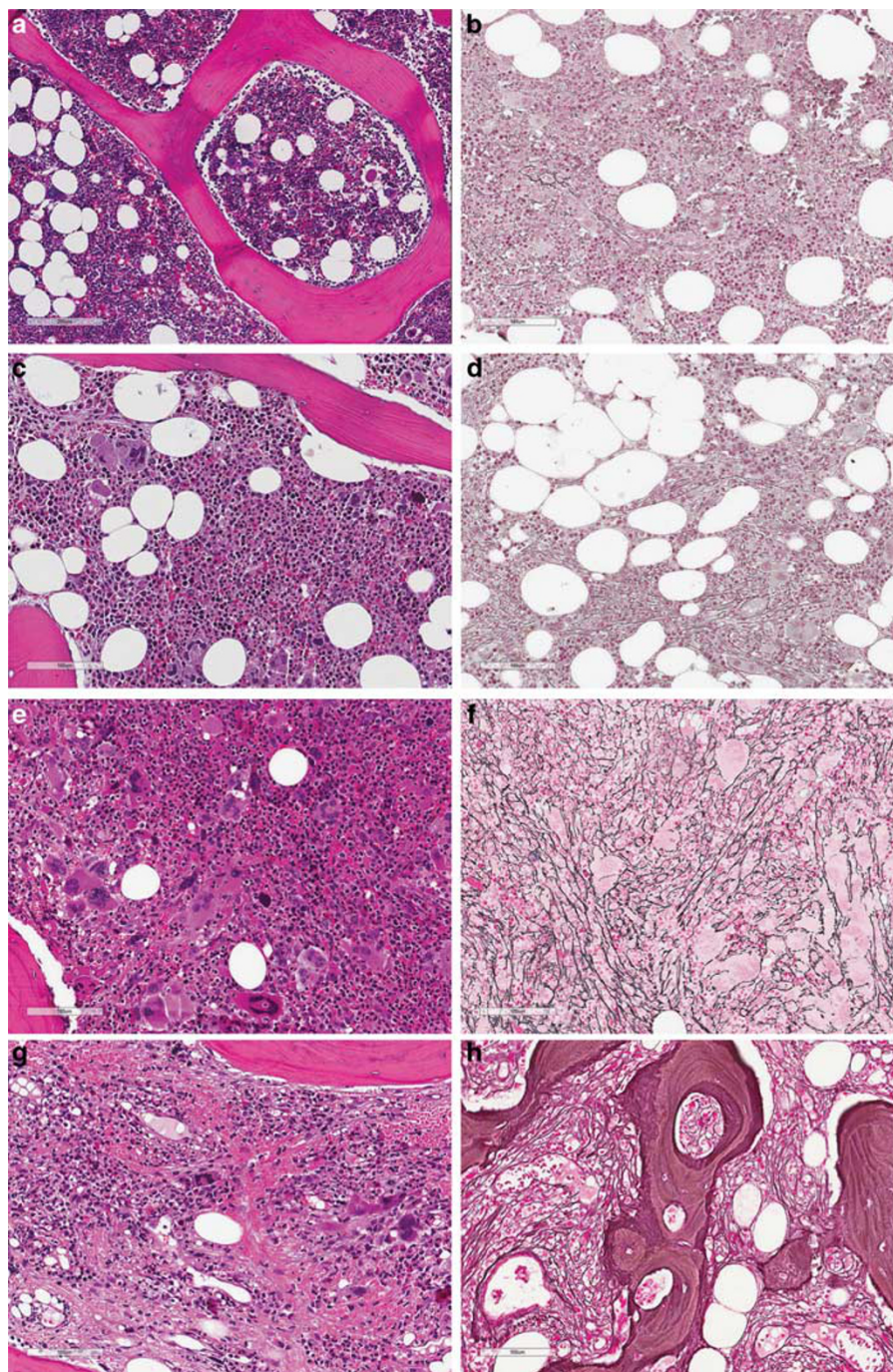To evaluate reproducibility of fibrosis grading in patients undergoing treatment with fedratinib, we

Figure 1 Examples of bone marrow samples where all the three pathologists agreed on fibrosis grade. (**a** and **b**) Grade 0 ((**a**) hematoxylin and eosin, (**b**) reticulin). (**c** and **d**) Grade 1 ((**c**) hematoxylin and eosin, (**d**) reticulin). (**e** and **f**) Grade 2 ((**e**) hematoxylin and eosin, (**f**) reticulin). (**g** and **h**) Grade 3 ((**g**) hematoxylin and eosin, (**h**) reticulin). Magnification scale indicated in lower left of each image.

performed separate analyses for the biopsies taken at baseline (pre-fedratinib treatment; $n = 233$) and during or after fedratinib therapy ($n = 168$); the latter group excluded all patients within the treatment group on the Phase 3 study, some of whom received placebo. The agreement between each pair of pathologists for the treated group in comparison to the baseline group is shown in Table 4. Considering all pathologist pairs together, the rate of agreement for baseline samples was significantly higher than that for post-treatment samples ($P = 0.023$). However, all kappa values for the treated group were at least in the 'substantial correlation' category (Table 4).

Discussion

Bone marrow fibrosis of significantly varying degree may be associated with the three major myeloproliferative neoplasm entities either at presentation, such as in primary myelofibrosis, or upon disease progression in polycythemia vera and essential thrombocythemia. The updated 2008 WHO classification has adopted the European Consensus on grading of bone marrow fibrosis,[14] a four-tiered system with grades 0–3.[15] Development of bone marrow fibrosis represents a stepwise evolution from absent or minimal reticulin fibrosis to marked reticulin and/or collagen fibrosis often associated with osteosclerosis,[7–9] and precise fibrosis grade assignment is essential both for initially classifying myeloproliferative neoplasms and identifying disease progression.[15,22,28] Although currently reticulin fibrosis grade is not widely accepted for risk assessment in patients with primary myelofibrosis, recent studies have shown that higher grades of bone marrow fibrosis independently predict poor outcome. Thus, concomitant use of clinical and morphological scoring systems may allow a better prediction of survival and patient risk stratification.[18,21,28] Further study is needed to correlate changes in bone marrow fibrosis with clinical and biological markers of disease evolution in myeloproliferative neoplasms.

In this study we investigated the reproducibility of the WHO fibrosis grading system in 261 patients with primary myelofibrosis, post-polycythemic myelofibrosis, and post-thrombocythemic myelofibrosis who were enrolled in trials with fedratinib, a selective *JAK2* inhibitor by unweighted kappa statistics, a well-established approach for the evaluation of observer agreement for categorical data.[49] To the best of our knowledge, this is the largest study evaluating interobserver fibrosis grading agreement to date and it shows high concordance of fibrosis grading for the entire cohort, with a kappa coefficient equal to or greater than 0.8 for all pathologist pairs (Table 3 and Figure 1). All the three pathologists agreed on a particular grade in 83% samples, which is a higher rate than reported in the

Table 3 Concordance of fibrosis grading ($n = 665$ biopsies) between the three pathologists

	Kappa
Pathologist 1 *versus* pathologist 2	0.80[a]
Pathologist 1 *versus* pathologist 3	0.84[b]
Pathologist 2 *versus* pathologist 3	0.84[b]

[a]Substantial correlation.
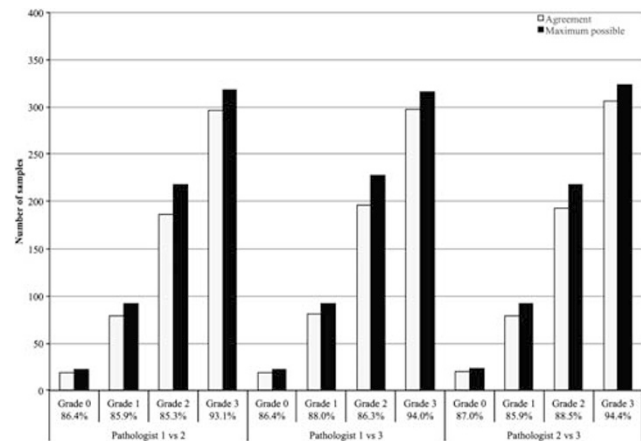[b]Almost perfect correlation.

Figure 2 Frequency of agreement on each individual reticulin grade between each pair of pathologists.

published literature. For example, Wilkins *et al*.[45] reported a fibrosis grading statistical analysis on 370 biopsies from patients with essential thrombocythemia using a log-linear modeling of pairwise interobserver agreement method, showing a strength of association for reticulin fibrosis of 5.1 (95% CI, 4.0–6.4) with the three hematopathologists agreeing within one grade of one another in 69% of cases. However, this study employed a 5-grade system (0–4) that would be expected to produce more frequent variance in grade between observers than a 4-grade system used in our study. Moreover, this study found a stronger agreement between pathologists scoring in the reticulin grade compared with all other criteria reported in the same study (such as megakaryocytic morphology and clustering) and even assignment of WHO diagnosis. This study therefore concluded that reticulin grade was the dominant independent predictor of the WHO diagnosis of essential thrombocythemia.[45] Interestingly, a Danish group has reported similar experience, in which the degree of concordance of morphological assessment among seven hematopathologists improved from 53% (CI, 46.6–58.5%) to 60% (CI, 54.1–65.8%) by adding reticulin fibrosis grade to the assessment criteria.[50] However, concerning the latter study the number of panelists was higher than that in our study (seven *versus* three), which would result in lower chance of complete agreement; in addition, the inclusion of many control specimens may have lowered the overall rate of concordance.

Table 4 Concordance of fibrosis grading at baseline and during/after fedratinib treatment

Pathologist	Baseline			Treated			P-value (baseline versus treated)
	Proportion of cases with agreement	Percent agreement	Kappa	Proportion of cases with agreement	Percent agreement	Kappa	
1 *versus* 2	208/224	92.9%	0.82[a]	140/160	87.5%	0.76[b]	0.10
1 *versus* 3	213/228	93.4%	0.86[a]	144/161	89.4%	0.80[b]	0.08
2 *versus* 3	212/228	93.0%	0.85[a]	147/165	89.1%	0.82[a]	0.32

[a]Almost perfect correlation.
[b]Substantial correlation.

The relatively high interobserver agreement rate of fibrosis grade in the current study could be explained by several factors. First, except for the Phase 1/2 study, hematoxylin and eosin, reticulin, and trichrome stains were performed centrally at the Covance Central Laboratory under optimized conditions that improved stain quality, especially consistency of reticulin fiber intensity. In biopsies with suboptimal reticulin stain, the staining was repeated and was only considered uninterpretable if the staining failed a second time. Second, a large proportion of the biopsies in the study demonstrated WHO grade 3 fibrosis (320 biopsies, 48.1%), which showed the highest rate of agreement between the three pathologists (93.1–94.4%). These results are not surprising, as it is generally easier to diagnose an end-stage pathologic process. The rates of agreement for grades 0, 1, and 2 were <90% and the rate of agreement was significantly lower when compared with grade 3. Finally, the WHO classification grading criteria were extensively discussed before reviewing the slides in a training session between the three pathologists. On the basis of this training session, the pathologists agreed to grade fibrosis only in areas of hematopoiesis following the strict adherence to the WHO fibrosis grading system.[15] Although the WHO Classification does not provide guidance on handling cases with mixed grades, the pathologists decided in advance to assign grade based on the grade comprising the majority of the hematopoietic area.

It is not unusual to encounter biopsies with mixed grades, especially in biopsies from patients on therapy.[36,37] We compared the concordance of fibrosis grading in baseline and samples to those taken from patients on therapy; such an analysis has never been reported before our knowledge. We found that there was significantly lower agreement in samples taken from patients on fedratinib therapy ($P=0.023$). Nevertheless, even in these treated samples, the kappa statistics for all observer pairs were still within the 'substantial correlation' or 'near perfect correlation' ranges. These results validate our approach and suggest high reproducibility of the WHO fibrosis grading system,[15] even in patients undergoing therapy with disease-modifying agents such as fedratinib.

During the slide review and grading, the pathologists were blinded to the patients' treatment status. However, retrospective review of the selected slides from the treated group showed frequent low-grade reticulin fibrosis in spite of significant collagen staining with trichrome, indicating a divergence between reticulin fibrosis and collagen fibers; these cases often showed persistent osteosclerosis as well (data not shown). This observation suggests that *JAK2* inhibitors may alter the disease biology by causing regression of reticulin fibers before regression of collagen fibrosis or osteosclerosis. This notion is indirectly supported by a case report that showed complete resolution of bone marrow fibrosis only after 168 weeks of therapy with ruxolitinib, a *JAK1/JAK2* inhibitor,[38] and particularly after long term (5 years) of ruxolitinib treatment[51] and is similar to findings reported after stem marrow transplantation in primary myelofibrosis.[24] A unified approach is needed to address the issues of fibrosis grade heterogeneity, which may have contributed to the lower fibrosis grading concordance in treated patients, as well as the divergence in some treated cases between lower reticulin grades in the presence of collagen on trichrome stain and significant osteosclerosis. The WHO grading system does not currently provide guidance on these issues but could be slightly modified to take into account these observations in treated patients. For example, the use of the 'majority' grade in heterogeneous cases could be adopted.

In an attempt to standardize fibrosis grading assessment, Teman *et al*[47] have proposed quantification of fibrosis and osteosclerosis by a computer-assisted approach using a color deconvolution algorithm that graded reticulin staining as a percentage of black pixels in three representative areas in a core biopsy that were later averaged. Osteosclerosis was assessed by manually selecting the bony trabeculae on a scanned slide and calculating the ratio of bony trabecular area to total biopsy area. Although the idea of a consistent objective fibrosis scoring system is appealing, this computer-assisted approach has practical limitations that preclude its clinical use currently. Moreover, evaluation of only three randomly selected hematopoietic areas (total area of $1.5\,mm^2$) may

Fibrosis/cellularity grading in bone marrow
O Pozdnyakova et al
1453

generate a significant bias concerning objectivity, particularly in cases with heterogeneity of reticulin grade that require careful examination of the entire slide.[36,37] In addition, Teman et al.[47] validated only the computer-assisted scoring of grades 2 and 3. If such computer-based system is to be employed in clinical practice, it is imperative that the scoring software also recognizes grades 0 and 1 for treatment efficacy assessment. We believe that visual assessment of fibrosis by a hematopathologist remains the current gold standard for the diagnosis and serial monitoring of patients with myeloproliferative neoplasms.

In conclusion, accurate grading of bone marrow fibrosis is important for the initial diagnosis, classification, and risk stratification of the different myeloproliferative neoplasm entities; in the new era of the *JAK1/JAK2* and *JAK2* inhibitor agents, reticulin grade is a compelling candidate, along with existing clinical and biological factors, to measure treatment response.[38] Our study showed that the current WHO fibrosis grading system[15] is practically applicable and highly reproducible for establishing baseline fibrosis and subsequent fibrosis grades (which may increase or decrease) in patients undergoing therapy.

Acknowledgments

Disclosure/conflict of interest

Drs Pozdnyakova, Thiele, Rodig, and Hasserjian received consulting income from Sanofi Oncology for reviewing biopsies within the context of the trials described. Drs Wu and Patki are employees of the Sanofi Oncology.

References

1 Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med 2005;352:1779–1790.

2 Pikman Y, Lee BH, Mercher T, et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med 2006;3:e270.

3 Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. N Engl J Med 2013;369:2379–2390.

4 Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with non-mutated JAK2. N Engl J Med 2013;369:2391–2405.

5 Tefferi A, Lasho TL, Finke CM, et al. CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. Leukemia 2014; PMID: 24402162 (e-pub ahead of print).

6 Thiele J, Kvasnicka HM, Schmitt-Graeff A, et al. Follow-up examinations including sequential bone marrow biopsies in essential thrombocythemia (ET): a retrospective clinicopathological study of 120 patients. Am J Hematol 2002;70:283–291.

7 Thiele J, Kvasnicka HM, Schmitt-Graeff A, et al. Dynamics of fibrosis in chronic idiopathic (primary) myelofibrosis during therapy: a follow-up study on 309 patients. Leuk Lymphoma 2003;44:949–953.

8 Barosi G, Rosti V, Bonetti E, et al. Evidence that prefibrotic myelofibrosis is aligned along a clinical and biological continuum featuring primary myelofibrosis. PLoS One 2012;7:e35631.

9 Buhr T, Busche G, Choritz H, et al. Evolution of myelofibrosis in chronic idiopathic myelofibrosis as evidenced in sequential bone marrow biopsy specimens. Am J Clin Pathol 2003;119:152–158.

10 Barbui T, Thiele J, Passamonti F, et al. Initial bone marrow reticulin fibrosis in polycythemia vera exerts an impact on clinical outcome. Blood 2012;119:2239–2241.

11 Ellis JT, Peterson P, Geller SA, et al. Studies of the bone marrow in polycythemia vera and the evolution of myelofibrosis and second hematologic malignancies. Semin Hematol 1986;23:144–155.

12 Mesa RA, Green A, Barosi G, et al. MPN-associated myelofibrosis (MPN-MF). Leuk Res 2011;35:12–13.

13 Kuter DJ, Bain B, Mufti G, et al. Bone marrow fibrosis: pathophysiology and clinical significance of increased bone marrow stromal fibres. Br J Haematol 2007;139:351–362.

14 Thiele J, Kvasnicka HM, Facchetti F, et al. European consensus on grading bone marrow fibrosis and assessment of cellularity. Haematologica 2005;90:1128–1132.

15 Thiele J, Kvasnicka HM, Tefferi A, et al. Primary myelofibrosis. In: Swerdlow SH, Campo E, Harris NL, et al (eds). WHO Classifications of Tumours of Haematopoietic and Lymphoid Tissues, 4th edn. IARC Press: Lyon, France; 2008, pp 44–47.

16 Passamonti F, Cervantes F, Vannucchi AM, et al. A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). Blood 2010;115:1703–1708.

17 Gangat N, Caramazza D, Vaidya R, et al. DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. J Clin Oncol 2011;29:392–397.

18 Kvasnicka HM, Thiele J. The impact of clinicopathological studies on staging and survival in essential thrombocythemia, chronic idiopathic myelofibrosis, and polycythemia rubra vera. Semin Thromb Hemost 2006;32:362–371.

19 Thiele J, Kvasnicka HM. Grade of bone marrow fibrosis is associated with relevant hematological findings-a clinicopathological study on 865 patients with chronic idiopathic myelofibrosis. Ann Hematol 2006;85:226–232.

20 Vener C, Fracchiolla NS, Gianelli U, et al. Prognostic implications of the European consensus for grading of bone marrow fibrosis in chronic idiopathic myelofibrosis. Blood 2008;111:1862–1865.

21 Gianelli U, Vener C, Bossi A, et al. The European Consensus on grading of bone marrow fibrosis allows

a better prognostication of patients with primary myelofibrosis. Mod Pathol 2012;25:1193–1202.

22 Barosi G, Mesa RA, Thiele J, *et al.* Proposed criteria for the diagnosis of post-polycythemia vera and post-essential thrombocythemia myelofibrosis: a consensus statement from the International Working Group for Myelofibrosis Research and Treatment. Leukemia 2008;22:437–438.

23 Sale GE, Deeg HJ, Porter BA. Regression of myelofibrosis and osteosclerosis following hematopoietic cell transplantation assessed by magnetic resonance imaging and histologic grading. Biol Blood Marrow Transplant 2006;12:1285–1294.

24 Thiele J, Kvasnicka HM, Dietrich H, *et al.* Dynamics of bone marrow changes in patients with chronic idiopathic myelofibrosis following allogeneic stem cell transplantation. Histol Histopathol 2005;20:879–889.

25 Kroger N, Kvasnicka M, Thiele J. Replacement of hematopoietic system by allogeneic stem cell transplantation in myelofibrosis patients induces rapid regression of bone marrow fibrosis. Fibrogenesis Tissue Repair 2012;5(Suppl 1):S25.

26 Hasserjian RP, Boecklin F, Parker S, *et al.* ST1571 (imatinib mesylate) reduces bone marrow cellularity and normalizes morphologic features irrespective of cytogenetic response. Am. J Clin Pathol 2002;117:360–367.

27 Pullarkat V, Bass RD, Gong JZ, *et al.* Primary autoimmune myelofibrosis: definition of a distinct clinicopathologic syndrome. Am J Hematol 2003;72:8–12.

28 Thiele J, Kvasnicka HM, Schmitt-Graeff A, *et al.* Bone marrow changes in chronic myelogenous leukaemia after long-term treatment with the tyrosine kinase inhibitor STI571: an immunohistochemical study on 75 patients. Histopathology 2005;46:540–550.

29 Verstovsek S, Kantarjian HM, Estrov Z, *et al.* Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. Blood 2012;120:1202–1209.

30 Verstovsek S, Mesa RA, Gotlib J, *et al.* A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. N Engl J Med 2012;366:799–807.

31 Passamonti F, Maffioli M, Cervantes F, *et al.* Impact of ruxolitinib on the natural history of primary myelofibrosis: a comparison of the DIPSS and the COMFORT-2 cohorts. Blood 2014;123:1833–1835.

32 Verstovsek S, Mesa RA, Gotlib J, *et al.* Efficacy, safety and survival with ruxolitinib in patients with myelofibrosis: results of a median 2-year follow-up of COMFORT-I. Haematologica 2013;98:1865–1871.

33 Harrison C, Kiladjian JJ, Al-Ali HK, *et al.* JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. N Engl J Med 2012;366:787–798.

34 Cervantes F, Vannucchi AM, Kiladjian JJ, *et al.* Three-year efficacy, safety, and survival findings from COMFORT-II, a phase 3 study comparing ruxolitinib with best available therapy for myelofibrosis. Blood 2013;122:4047–4053.

35 Mesa RA, Gotlib J, Gupta V, *et al.* Effect of ruxolitinib therapy on myelofibrosis-related symptoms and other patient-reported outcomes in COMFORT-I: a randomized, double-blind, placebo-controlled trial. J Clin Oncol 2013;31:1285–1292.

36 Thiele J, Kvasnicka HM, Schmitt-Graeff A, *et al.* Bone marrow histopathology following cytoreductive therapy in chronic idiopathic myelofibrosis. Histopathology 2003;43:470–479.

37 Thiele J, Kvasnicka HM, Schmitt-Graff A, *et al.* Therapy-related changes of the bone marrow in chronic idiopathic myelofibrosis. Histol Histopathol 2004;19:239–250.

38 Wilkins BS, Radia D, Woodley C, *et al.* Resolution of bone marrow fibrosis in a patient receiving JAK1/JAK2 inhibitor treatment with ruxolitinib. Haematologica 2013;98:1872–1876.

39 Kvasnicka H-M, Thiele J, Bueso-Ramos CE, *et al.* Exploratory analysis of the effect of ruxolitinib on bone marrow morphology in patients with myelofibrosis. ASCO Meeting Abstracts 2013;31:7030.

40 Jamieson C, Hasserjian RP, Gotlib J, *et al.* Effect of treatment with the JAK2-selective inhibitor fedratinib (SAR302503) on bone marrow histology in patients with myeloproliferative neoplasms with myelofibrosis. Blood 2013;122:2823.

41 Silver RT, Kiladjian JJ, Hasselbalch HC. Interferon and the treatment of polycythemia vera, essential thrombocythemia and myelofibrosis. Expert Rev Hematol 2013;6:49–58.

42 Barosi G, Mesa R, Finazzi G, *et al.* Revised response criteria for polycythemia vera and essential thrombocythemia: an ELN and IWG-MRT consensus project. Blood 2013;121:4778–4781.

43 Tefferi A, Cervantes F, Mesa R, *et al.* Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. Blood 2013;122:1395–1398.

44 Thiele J, Kvasnicka HM. Myelofibrosis—what's in a name? Consensus on definition and EUMNET grading. Pathobiology 2007;74:89–96.

45 Wilkins BS, Erber WN, Bareford D, *et al.* Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. Blood 2008;111:60–70.

46 Campbell PJ, Bareford D, Erber WN, *et al.* Reticulin accumulation in essential thrombocythemia: prognostic significance and relationship to therapy. J Clin Oncol 2009;27:2991–2999.

47 Teman CJ, Wilson AR, Perkins SL, *et al.* Quantification of fibrosis and osteosclerosis in myeloproliferative neoplasms: a computer-assisted image study. Leuk Res 2010;34:871–876.

48 Barbui T, Thiele J, Vannucchi AM, *et al.* Problems and pitfalls regarding WHO-defined diagnosis of early/prefibrotic primary myelofibrosis versus essential thrombocythemia. Leukemia 2013;27:1953–1958.

49 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.

50 Madelung AB, Bondo H, Stamp I, *et al.* World Health Organization-defined classification of myeloproliferative neoplasms: morphological reproducibility and clinical correlations–the Danish experience. Am J Hematol 2013;88:1012–1016.

51 Kvasnicka HM, Thiele J, Bueso-Ramos CE, *et al.* Effects of five-years of ruxolitinib therapy on bone marrow morphology in patients with myelofibrosis and comparison with best available therapy. Blood 2013;122:4055.