Approval of Jakafi (ruxolitinib) Based on a Home-Grown, Patient-Reported Outcome Instrument: A Case Study

By Ron C. Falcone, PhD and Richard S. Levy, MD

Jakafi® (ruxolitinib) was approved¹ by FDA on 16 November 2011 for the treatment of intermediate or high-risk myelofibrosis (MF) including primary MF, post-polycythemia vera MF (PPV-MF) and post-essential thrombocythemia MF (PET-MF). Prior to this time, there were no approved products, and therefore no precedent for regulatory endpoints in this orphan disease.

MF is a highly symptomatic² myeloproliferative neoplasm (MPN), characterized by elevated levels of inflammatory cytokines and profound enlargement of the spleen associated with extramedullary hematopoiesis.

Ruxolitinib is an orally active new chemical entity (NCE) and the first in a new class of drugs that inhibit Janus kinase (JAK) 1 and 2. The normal role of JAKs in signal transduction is to mediate cytokine and growth factor receptor signaling, which in turn influences cell proliferation, differentiation and cell survival. In 2005, several groups identified the association of the JAK2V617F mutation with MPNs.³⁻⁶

JAK2V617F is a gain-of-function mutation that affects the proliferation of blood cell lineages and the production of high levels of proinflammatory cytokines, leading to the signs and symptoms of MPNs. Although not all cases of MF are related to this mutation, all are characterized by hyperactivation of the JAK-signal transducer and activator of transcription (STAT) pathway.⁷ Ruxolitinib is an ATP mimic which is active regardless of the presence of JAK2V617F⁸⁻⁹ or other mutations subsequently discovered.¹⁰

The regulatory endpoints that provided the basis for a full FDA approval of ruxolitinib were 1) a decrease in splenomegaly and 2) an improvement in MF-related symptoms as measured by a novel patient-reported outcome (PRO) instrument, the modified Myelofibrosis Symptom Assessment Form (MFSAF) version 2.0 diary. By definition, a PRO is any report on the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or other intermediary.
A PRO instrument (i.e., a questionnaire plus the information and documentation that support its use) is a means to capture PRO data used to measure treatment benefit or risk in clinical trials. This case study will describe the 4.5 years from IND to NDA approval of ruxolitinib and the regulatory interactions and patient feedback that drove the overall development plan and inclusion of the PRO instrument into a Phase III trial and product labeling.

**Early Days**
Starting with the first few patients treated with ruxolitinib in the Phase I/II trial, unprecedented symptomatic improvement and reduction in splenomegaly were reported. As there were no effective therapies, physicians were unwilling to enroll patients in survival studies without the possibility of early cross-over, and there were no accepted or appropriate definitions of disease progression. Therefore, Incyte’s goal was to gain agreement on an acceptable basis of approval and ideally to have symptomatic improvement included in the product labeling.

A meeting was held in March 2008 with FDA’s Division of Drug Oncology Products (DDOP) to review efficacy and safety data and discuss possible registration endpoints for MF. FDA indicated an objectively measured spleen size reduction as well as one other clinically relevant benefit might support registration.

Use of symptomatic improvement based on FDA’s draft PRO guidance11 was mentioned as a possible, but difficult and risky approach. In April of 2008, the ongoing Phase I/II protocol was amended to formally collect information on the severity of MF symptoms.

**Development of a Patient-Reported Outcome Instrument**
The development of the PRO instrument and continual interactions with FDA were key factors in the successful incorporation of symptoms into the label.

The original MFSAF, a 19-point questionnaire, was developed by a group of investigators from the Mayo Clinic participating in the Phase I/II study of ruxolitinib in MF. This questionnaire assessed many of the symptoms originally identified by Ruben Mesa, MD (Mayo Clinic, Arizona) in his international Internet-based survey of 456 patients with MF.

In the spring of 2008, assessment of spleen volume by MRI/CT and a number of other potential measures of clinically relevant improvement, including a six-minute walk test (6MWT) were added by amendment to the Phase I/II study. At the same time, Incyte began the process of developing a PRO instrument by following the draft FDA guidance,11 starting with non-structured patient interviews and then cognitive testing.

This work led to the development of a 46-item PRO instrument called the Myelofibrosis Symptom Diary (MFSD). This PRO instrument was developed based on a range of symptoms and signs which were evaluated by severity, frequency, duration, and degree of discomfort. In addition, cognitive testing of patients with MF was successfully performed.

Another meeting was held in September 2008 with FDA’s DDOP, which included a representative of the Study Endpoints and Label Development (SEALD) team to further consider endpoints for a registration study and review of the MFSD. Several potential co-primary or secondary endpoints were discussed, and the 6MWT was recommended as the most viable of the co-primary endpoints.

However, the DDOP continued to take the position an endpoint based on symptoms was risky. FDA also indicated the concept of fatigue, although important in MF, could not be adequately understood as it was a multidimensional domain.

Consistent with these discussions, the SEALD team provided the following advice on the MFSD: 1) focus on symptoms at their worst severity, 2) use a 24-hour recall period, and 3) use a 0-10 numeric rating scale. FDA also recommended additional patient interviews to establish content validity, which were subsequently conducted.

Because of the excellent results with the 6MWT in the Phase I/II trial, Incyte requested a Special Protocol Assessment (SPA) which included co-primary endpoints based on reduction in spleen volume and improvement in 6MWT. Symptoms of MF were proposed as a secondary endpoint without alpha control, and there was no expectation that these results would be included in labeling,
No agreement was reached on the SPA, because the only demonstration of impaired walk distance in MF was supported by Incyte’s own data. FDA then suggested Incyte submit a new SPA based on a novel definition of disease progression.

In March 2009, Incyte submitted a second SPA based on FDA’s proposed definition of disease progression. In this SPA, symptomatic improvement remained a secondary, non-alpha–controlled endpoint.

Because it was expected that differences between the treatment arms would reach statistical significance well in advance of the median time to progression, FDA did not agree with the second SPA. The agency, however, offered another option which included reduction in spleen volume as the primary endpoint and a secondary endpoint of symptom improvement measured daily using a version of the MFSAF published by Mesa earlier that year.12

The publication described an updated version of the original MFSAF had been used in the Phase I/II trial. Although fatigue was identified as one of the most common and bothersome symptoms of MF, Incyte was advised by FDA to remove fatigue based on previously communicated concerns.

Prior to submitting a new SPA, the qualitative patient interviews and cognitive testing were re-evaluated and shown to support the new version of the MFSAF, the modified MFSAF version 2.0 diary. This instrument contained 7 questions, 6 of which were scored in a composite referred to as the Total Symptom Score (TSS).

“Inactivity,” although included in the daily diary, is considered an outcome and was therefore not included in the scoring of the TSS. The six questions scored in the TSS are listed in Table 1.

In the SPA, although we had performed cognitive testing and demonstrated content validity, we had not completed all the steps to demonstrate “fit for purpose” as outlined in the PRO guidance. Incyte proposed to demonstrate “fit for purpose” by analyzing blinded data at 1 month of the 6-month randomized, placebo-controlled Phase III trial and also the unblinded data obtained at the end of the study.

The newly formed Division of Hematology Drug Products agreed to this new SPA, but no comment was given on the acceptability of demonstrating “fit for purpose” within the registration study. The modified MFSAF v2.0 diary was demonstrated as “fit for purpose” by performing the required activities and analyses as outlined in FDA’s PRO guidance which had become finalized in March of 2009, and a complete Evidence Dossier was included with the NDA.

In the Phase III trial, electronic handheld diaries were used to collect the patient-reported symptom data. Compliance with data entry was 96% of all expected data entered; 98% completed a minimum requirement of 4 out of 7 baseline days; 95% completed a minimum requirement of 20 out of 28 days during Month 6 of the trial; and 94% completed the daily assessment in 1 minute or less.

The test-retest reliability correlation coefficient from Week 7 to Week 8 was 0.97 with placebo and 0.98 with ruxolitinib. Correlation of pain items in the MFSAF with pain scores in the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30 (EORTC QLQ-C30) and Brief Pain Inventory (BPI) was approximately 0.6.

Table 1: The Modified MFSAF Version 2.0 Diary

<table>
<thead>
<tr>
<th>Questions for the 6 Symptoms Comprising the Total Symptom Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. During the past 24 hours, how severe were your worst night sweats (or feeling hot or flushed) due to MF?</td>
</tr>
<tr>
<td>2. During the past 24 hours, how severe was your worst itchiness due to MF?</td>
</tr>
<tr>
<td>3. During the past 24 hours, how severe was your worst abdominal discomfort (feel uncomfortable, pressure or bloating) due to MF?</td>
</tr>
<tr>
<td>4. During the past 24 hours, how severe was your worst pain under the ribs on the left side due to MF?</td>
</tr>
<tr>
<td>5. During the past 24 hours, what was the worst feeling of fullness (early satiety) you had after beginning to eat due to MF?</td>
</tr>
<tr>
<td>6. During the past 24 hours, how severe was your worst bone or muscle pain due to MF (diffuse not joint or arthritis pain)?</td>
</tr>
</tbody>
</table>

Reprinted with permission from Incyte Corporation.
When evaluating the various methods available to guide interpretation of the modified MFSAF v2.0, it was Incyte’s goal to ensure, regardless of the type or magnitude of observed changes, those changes were a) meaningful for patients and b) simple for clinicians and researchers to understand.

For these reasons, anchor-based methods using the Patient Global Impression of Change (PGIC) responses were utilized. The PGIC is often used in clinical trials to evaluate a patient’s overall sense of whether a treatment has been beneficial or not and has been used across therapeutic categories.13

The specific wording of the PGIC item was: “Since the start of the treatment you’ve received in this study, your MF symptoms are: 1) very much improved, 2) much improved, 3) minimally improved, 4) no change, 5) minimally worse, 6) much worse, 7) very much worse.”

For the present evaluation, patients were grouped as either “responders” or “non-responders” to indicate whether they met the trial’s responder definition of a 50% or greater reduction in their TSS from baseline to Week 24. Next, these groups were categorized or “anchored” to the PGIC response provided at Week 24.

In this way, groups (i.e., responders or non-responders) could be evaluated relative to whether or not they felt their condition had improved. It was expected the proportion of patients labeled as “responders” would only be minimally represented in the “no change” or “worsening” categories and overrepresented in the “improvement” categories compared to non-responders.

The PGIC results showed that over 90% of responders reported a PGIC score of “much improved” or “very much improved,” thus supporting the results of the modified MSFAF v2.0 PRO instrument and interpretation that the changes in symptom scores were clinically meaningful.

When evaluating the various methods available to guide interpretation of the modified MFSAF v2.0, it was Incyte’s goal to ensure, regardless of the type or magnitude of observed changes, those changes were a) meaningful for patients and b) simple for clinicians and researchers to understand.

For these reasons, anchor-based methods using the Patient Global Impression of Change (PGIC) responses were utilized. The PGIC is often used in clinical trials to evaluate a patient’s overall sense of whether a treatment has been beneficial or not and has been used across therapeutic categories.13

The specific wording of the PGIC item was: “Since the start of the treatment you’ve received in this study, your MF symptoms are: 1) very much improved, 2) much improved, 3) minimally improved, 4) no change, 5) minimally worse, 6) much worse, 7) very much worse.”

For the present evaluation, patients were grouped as either “responders” or “non-responders” to indicate whether they met the trial’s responder definition of a 50% or greater reduction in their TSS from baseline to Week 24. Next, these groups were categorized or “anchored” to the PGIC response provided at Week 24.

In this way, groups (i.e., responders or non-responders) could be evaluated relative to whether or not they felt their condition had improved. It was expected the proportion of patients labeled as “responders” would only be minimally represented in the “no change” or “worsening” categories and overrepresented in the “improvement” categories compared to non-responders.

The PGIC results showed that over 90% of responders reported a PGIC score of “much improved” or “very much improved,” thus supporting the results of the modified MSFAF v2.0 PRO instrument and interpretation that the changes in symptom scores were clinically meaningful.

When evaluating the various methods available to guide interpretation of the modified MFSAF v2.0, it was Incyte’s goal to ensure, regardless of the type or magnitude of observed changes, those changes were a) meaningful for patients and b) simple for clinicians and researchers to understand.

For these reasons, anchor-based methods using the Patient Global Impression of Change (PGIC) responses were utilized. The PGIC is often used in clinical trials to evaluate a patient’s overall sense of whether a treatment has been beneficial or not and has been used across therapeutic categories.13

The specific wording of the PGIC item was: “Since the start of the treatment you’ve received in this study, your MF symptoms are: 1) very much improved, 2) much improved, 3) minimally improved, 4) no change, 5) minimally worse, 6) much worse, 7) very much worse.”

For the present evaluation, patients were grouped as either “responders” or “non-responders” to indicate whether they met the trial’s responder definition of a 50% or greater reduction in their TSS from baseline to Week 24. Next, these groups were categorized or “anchored” to the PGIC response provided at Week 24.

In this way, groups (i.e., responders or non-responders) could be evaluated relative to whether or not they felt their condition had improved. It was expected the proportion of patients labeled as “responders” would only be minimally represented in the “no change” or “worsening” categories and overrepresented in the “improvement” categories compared to non-responders.

The PGIC results showed that over 90% of responders reported a PGIC score of “much improved” or “very much improved,” thus supporting the results of the modified MSFAF v2.0 PRO instrument and interpretation that the changes in symptom scores were clinically meaningful.
Jakafi Product Label

The findings described above provided the basis for inclusion of symptoms in the Jakafi product label. In the US pivotal Phase III study, eight symptoms were a secondary endpoint and were measured using the modified MFSAF v2.0 diary. Symptom scores ranged from 0 to 10 with 0 representing “absent” and 10 representing “worst imaginable” symptoms. These scores were added to create the daily total score, which has a maximum of 60; the TSS represents the change from baseline to Week 24.

The secondary alpha-controlled endpoint was the proportion of patients with at least a 50% reduction (i.e., improvement) in TSS. At baseline, the mean TSS was 18.0 in the ruxolitinib group and 16.5 in the placebo group.

A higher proportion of patients in the ruxolitinib group (45.9%) had a 50% or greater reduction in TSS than in the placebo group (5.3%, P<0.001), with a median time to response of less than four weeks.

The TSS was included in the final product label (Figure 1) because the endpoint was alpha-controlled, the data collected was complete, the treatment effect size was robust and clinically meaningful and all of the MF symptoms (abdominal discomfort, pain under left ribs, night sweats, itching, bone/muscle pain, and early satiety) were driving the TSS (not just one or two symptoms). In addition, Figure 2 displays the proportion of patients with at least a 50% improvement in each of the individual symptoms that comprise the TSS, indicating that all six of the symptoms contributed to the higher TSS response rate in the group treated with ruxolitinib. Please see the complete Prescribing Information for Jakafi.

Discussion

The decision to incorporate a PRO instrument into the Phase III trial was worthwhile for Incyte, because it led to a faster approval on an endpoint in which there was great confidence based on the Phase I/II study. Another advantage was that symptom results were
included in the product label, enabling promotion of symptomatic improvement in this highly symptomatic disease.

The risk associated with the use of a PRO instrument was that it had not completed demonstration of “fit for purpose” prior to conducting the registration study. This risk was mitigated by having numerous discussions with FDA throughout the SPA process and an SPA agreement which included spleen size reduction as the only primary endpoint.

Therefore, the real risk was being unable to include the symptom data in the package insert, rather than drug approval, because symptomatic benefit would still be able to support a positive benefit-risk profile with an achieved primary endpoint. It is possible without the PRO data, an accelerated approval with a more complex regulatory path could have been the outcome rather than full approval.

The question arises as to whether a PRO instrument should be developed and included in all clinical development plans. Obvious factors in this consideration include the presence of symptoms in the patient population, the ability of the drug to improve some symptoms without making others worse, and the drug’s overall toxicity profile.

An “off-the-shelf” instrument may often not be directly suitable, and therefore a PRO instrument may need to be developed or modified for use in the specified patient population. This could be realized by following the steps in FDA’s final guidance in 2009.

Alternatively, FDA encourages the formation of collaborative groups in the “precompetitive” stage to increase the efficiency of joint efforts and to lessen the resource burden upon any individual person or company. To enable this approach, FDA issued the draft guidance, Qualification Process for Drug Development Tools, which provides the framework for FDA interactions to identify data needed to support a conclusion that, within the stated context of use, can support regulatory decision-making.

Once qualified, these tools can be used in multiple drug development programs. It is less labor intensive for FDA to deal with PROs developed by “precompetitive” industry and academic teams than to evaluate separate tools from each sponsor; however, the individual sponsor approach may be more appropriate and far more efficient for rare diseases where other companies are not interested in the effort.

In this context, it may be important for biopharmaceutical companies developing home-grown PRO measures to not restrict use by competitors, as such an approach might result in FDA being unable to devote the resources to evaluate a different tool for each company.

In conclusion, the successful incorporation of a PRO instrument by Incyte can be attributed to several factors. MF is a serious, life-threatening, orphan disease without any prior approved medical therapies, and therefore there were no regulatory precedents on the endpoints required for approval.

The iterative process with FDA started early in development and led to a flexible approach to define registration and labeling endpoints. The Phase III study was double-blind, and the historical problem of missing data was avoided by collecting data every day with an easy-to-use electronic diary.

Most importantly, MF is a highly symptomatic condition and ruxolitinib treatment was demonstrated to have a robust benefit on the TSS, with changes in each of the individual symptoms that comprise the TSS contributing to the overall benefit.

References

Authors

**Ron C. Falcone, Ph.D.** is the vice president of Regulatory Affairs at Incyte Corporation. He has over 35 years of experience in pharmaceutical development. Prior to joining Incyte, Dr. Falcone worked at Wyeth, AstraZeneca (ICI Americas, Zeneca), and Ortho-Biotech Oncology (Centocor). **Richard S. Levy, M.D.** is the executive vice president and chief drug development and medical officer at Incyte Corporation, where he led the clinical development of ruxolitinib for myelofibrosis. Dr. Levy has over 20 years of experience in clinical research, regulatory affairs, and project leadership in the biopharmaceutical industry, including his previous positions at Celgene, DuPont Pharmaceuticals, and Novartis.

© 2012 by the Regulatory Affairs Professionals Society. All rights reserved.