CLINICAL INVESTIGATION

Ruxolitinib for patients with polycythemia vera who have had an inadequate response or are intolerant to hydroxyurea: a critical appraisal

The goal of this review is to critically analyze data supporting the use of ruxolitinib in polycythemia vera patients resistant or intolerant to hydroxyurea. We analyzed the randomized Phase III study (RESPONSE) and we applied the Grading of Evidence, Assessment, Development and Evaluation approach by evaluating five dimensions of evidence: overall risk of bias, imprecision, inconsistency, indirectness and publication bias. We upgraded the quality of evidence because of large effect size on splenomegaly, hematocrit and symptoms but we downgraded it for performance bias and for indirectness of the comparator. In conclusion, by identifying factors affecting the quality of evidence, we rated the outcomes of ruxolitinib in polycythemia vera patients resistant or intolerant to hydroxyurea as having moderate level of evidence.

Keywords: critical appraisal • GRADE • hydroxyurea • JAK inhibitor • polycythemia vera • ruxolitinib

Giovanni Barosi
Center for the Study of Myelofibrosis, IRCCS Policlinico San Matteo Foundation, 27100 Pavia PV, Italy
Tel.: +39 038 250 363
Fax: +39 038 250 3917
barosig@smatteo.pv.it

What is known about polycythemia vera?

Polycythemia vera (PV) is a myeloproliferative neoplasm (MPN) clinically characterized by variable symptoms (e.g., aquagenic pruritus, erythromelalgia), increased tendency for thrombotic complications and inherent risk of hematological transformation (i.e., myelofibrosis and acute leukemia) [1]. From the biological perspective, PV is a clonal disease in which most blood cells are produced by one or a few abnormal clones; these clones have acquired somatic mutations that confer a survival advantage over normal hematopoietic cells. *JAK2*V617F mutation is present in almost all persons with PV [2–5].

The natural history of PV has been delineated in a large retrospective international study of 1545 patients diagnosed as per 2008 WHO criteria [6]. The median survival was 14.1 years, which was significantly worse than age- and sex-matched US population. Cumulative hazard of leukemic transformation, with death as a competing risk, was 2.3% at 10 years, 5.5% at 15 years and 7.9% at 20 years. In that study, the proportion of patients with arterial thrombosis, venous thrombosis and major hemorrhage at or prior to presentation was 16, 7 and 4%, respectively. After a median postdiagnosis follow-up time of 6.9 years (range: 0–39.3 years), the incidence of arterial thrombosis, venous thrombosis and major hemorrhage was 12, 9 and 4.2%, respectively. These rates of events were lower than those in older reports such as the European Collaboration on Low-Dose Aspirin in Polycythemia Vera study of 1638 patients diagnosed as per Polycythemia Vera Study Group criteria [7]. In that study, the cumulative risk of cardiovascular events (cardiovascular death and nonfatal thrombotic events) after a median follow-up period of 2.8 years reached 5.5 events per 100 person years. In contemporary reports, such as CYTO-PV study [8], the rate of death from cardiovascular events or of major thrombosis was lower (2.7% in patients maintaining low-hematocrit level and 9.8% in those with high hematocrit level) reflecting a more homogeneous patient population and potentially improvements in treatment delivery.

FUTURE SCIENCE part of fsg

On the basis of age ≥65 years and prior thrombosis, patients could be stratified into three groups with risk of cardiovascular events of 2.5, approximately 5.0 and 10.9 per 100 person years [1].

In 2013, the European LeukemiaNet (ELN) and International Working Group for Myelofibrosis Research and Therapy published new response criteria for PV designed for clinical trials [9]. These recommendations revised the previous response criteria [10] and were based on the concept that the definition of response should capture, besides hematological modifications, long-term effects of new drugs, such as absence of vascular events and histologic modifications. Additionally, the myelofibrosis-symptom assessment form (MF-SAF) provided a way to include a patient-reported outcome in the response definition [11].

The primary goal of treatment in PV is prevention of thrombohemorrhagic complications [12]. According to an expert panel recommendations, all PV patients should receive phlebotomy with a target hematocrit of <45%. Antiplatelet therapy chiefly in the form of aspirin has been shown to significantly reduce the risk of cardiovascular events. Higher risk PV patients should receive cytoreductive therapy, most frequently with hydroxyurea (HU) as the first-line agent [12].

An alternative first-line nonleukemogenic cytoreductive treatment option in PV is IFN-α, which current data suggest has a selective suppressive effect on the MPN clone [13–15].

The decision problem

Approximately 10% of patients with PV treated with HU show resistance, while some experience unacceptable side effects (intolerance) [16–18]. The chief forms of HU-related toxicity include painful cutaneous ulcers (perimalleolar, pretibial, hands, feet), painful oral mucosal lesions, high-grade fever, gastrointestinal intolerance, nonmelanoma skin cancers or pneumonitis [19].

Recognizing intolerance or resistance to HU is important for the decision about when to offer patients a second-line therapy. For this reason, specific criteria have been established by experts in the field, to identify HU intolerance/resistance in PV [16]. For patients on a daily dose of at least 2 g HU for at least 3 months, resistance/intolerance is defined when a patient experiences at least one of the following: a need for phlebotomy to keep hematocrit <45%; platelet count >400 × 10^9 /l and WBC >10 × 10^9 /l; <50% reduction in splenomegaly or no improvement in symptoms related to splenomegaly; absolute neutrophil count <1.0 × 10^9 /l, platelet count <100 × 10^9 /l or hemoglobin <10 g/dl at the lowest dose of HU necessary to achieve complete response (CR) or partial response (PR); or presence of unacceptable HU-related nonhematologic toxicities.

In PV, HU resistance has been associated with an increased risk of death and transformation to myelofibrosis, highlighting the importance of second-line therapeutic options for these patients [17].

Ruxolitinib is a tyrosine kinase inhibitor of JAK1 and JAK2 which prevents activation of JAK–STAT signaling pathway, thought to reduce proliferation of the MPN clone and release of inflammatory molecules [20]. One Phase II and one Phase III randomized trial reported efficacy of ruxolitinib in persons with PV who were resistant or intolerant to HU [21,22]. These data resulted that ruxolitinib, which is marketed in the USA by Incyte Corporation as Jakafi®, received approval in December 2014 from the US FDA for the treatment of patients with PV who have had an inadequate response to or cannot tolerate HU. On March 2015, the EC approved ruxolitinib (as Jakavi®) in Europe for PV patients who are resistant or intolerant to HU.

Appropriate use of ruxolitinib in clinical practice requires analyzing comparative efficacy and safety data. Our goal was to analyze evidence of clinical benefit of ruxolitinib in persons with PV.

Methods

We used the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) methodology to rate confidence in estimates of effect for each outcome [23]. This required assessing overall risk of bias, imprecision, inconsistency, indirectness and publication bias. We explicitly used factors which increased or decreased the quality of evidence which was rated as high, moderate, low or very low [24].

Clinical trials with ruxolitinib in PV

Two publications were selected including one randomized Phase III trial which reported short-term efficacy and safety [22], and one Phase II study which reported long-term efficacy and safety [21]. The Phase III trial was selected for detailed review.

Phase II study

In the Phase II study, patients with PV who were refractory to HU or for whom HU was contraindicated were eligible [21]. The study was initiated before the ELN criteria for HU resistance or intolerance were published [16]. Patients needed to have hematocrit >45% or two phlebotomies within the 24 weeks before enrollment; they were required to have at least one phlebotomy performed within 12 weeks before enrollment. Subjects were randomized to one of three ruxolitinib cohorts: 10 mg twice daily, 25 mg twice daily or 50 mg once daily, respectively. The dose-expansion cohort was determined on the efficacy and safety data of patients who completed at least 56 days of treatment.

CR was defined as hematocrit <45% without phlebotomy, a platelet count equal or lower than 400 × 10^9/l, a WBC count equal or lower than 10 × 10^9/l, a normal spleen by palpation, and no pruritus within the previous week. A PR was defined as hematocrit <45% without phlebotomy after week 4. These criteria differed from the 2009 ELN criteria [10]: actually, splenomegaly was measured using palpation instead of imaging, and symptom evaluation was limited to pruritus.

Changes in spleen length assessment included both the rate of patients with an equal or >50% spleen reduction and the rate of patients who achieved a nonpalpable spleen among those with palpable spleen at baseline. Symptom assessment included both the rate of patients with an equal or >50% reduction and those with resolution of pruritus, night sweats and bone pain.

A total of 34 PV patients from six centers in the USA and Italy were enrolled. Baseline characteristics reflected an advanced disease status with a median hematocrit of 46.7%, and with all patients having received at least one prior therapy (most commonly HU). The majority of patients (70.5%) needed at least two phlebotomies within 24 weeks before the first dose. At data cutoff, the median follow-up duration was 154 weeks (range, 35–179 weeks). Response was achieved in 97% of patients by week 24. Fifty-nine percent of patients achieved a CR and 38% achieved a PR as their best response. The majority of CRs occurred within the first year. Among responding patients, the probability of maintaining a hematocrit <45% without phlebotomy for 48 weeks and 144 weeks was 85 and 61%, respectively. Of the 11 patients who lost their hematocrit response, eight had a subsequent hematocrit <45% without intervening phlebotomies.

Responses on WBC count were achieved in 76 and 73% of patients who had WBC count >10 × 10^9/l or 15 × 10^9/l at baseline, respectively. Responses on platelet count were achieved in 74 and 69% of patients who had platelet count >400 × 10^9/l or >600 × 10^9/l at baseline, respectively.

Among patients with palpable splenomegaly, 70% achieved an equal or >50% reduction in the spleen at week 24. In addition, 44% of patients with palpable splenomegaly reduced their spleen size to nonpalpable at week 24. By week 144, 64% of patients had achieved an equal or >50% reduction in palpable spleen length and 63% had a nonpalpable spleen size. Meaningful improvements in pruritus, night sweats and bone pain were observed within 4 weeks of the beginning of therapy and were maintained through week 144.

Phase III study (RESPONSE)

RESPONSE is an international, open-label, randomized, multicenter Phase III trial [22]. Patients were stratified according to the prior HU therapy (inadequate response or unacceptable side effects) and were randomly assigned, in a 1:1 ratio, to ruxolitinib (starting dose, 10 mg twice daily) or standard therapy judged by the treating physician. Dose increases were allowed to achieve and maintain a hematocrit of <45% in the absence of phlebotomy, reduce splenomegaly (as assessed by palpation), and normalize WBC and platelet counts. Dose reductions or interruptions were mandated for cytopenias of grade 2 or higher.

Patients assigned to standard therapy could cross over to ruxolitinib at week 32 if the primary end point was not met or later in the case of disease progression (phlebotomy eligibility, progression of splenomegaly or both). Data cutoff for the primary analysis occurred when all patients reached week 48 or discontinued therapy.

The primary end point was the proportion of patients who had both hematocrit control and a reduction of 35% or more in spleen volume from baseline at week 32. Spleen volume assessment was centrally reviewed by MRI or CT studies. Hematocrit control was defined as ineligibility for phlebotomy from week 8 to 32 and no more than one instance of phlebotomy eligibility between randomization and week 8. Phlebotomy eligibility was defined as a hematocrit confirmed to be >45% that was at least 3% points higher than the baseline level or a hematocrit of more than 48%, whichever was lower. Secondary end points included the rate of subjects with a primary response (i.e., those in whom both components of the composite primary end point were achieved) at week 32 that was maintained at week 48 and the rate of subjects who had a complete hematologic remission (defined as hematocrit control, platelet count equal or <400×10^9/l and WBC count equal or <10×10^9 /l) at week 32. Other end points included the duration of response, symptom reduction and safety.

The study was sponsored by Incyte Corporation and by Novartis Pharmaceuticals. Representatives of the companies were involved in study design, data collection, interpretation and analyses. Moreover, editorial support was provided by a medical writer funded by the companies. However, the authors have been transparent about conflicts of interest and have largely detailed the role of sponsors in the conduct and reporting of the trial. Furthermore, an independent Data Safety and Monitoring Board oversaw the study conduct. To minimize biases, MRIs used to evaluate changes in spleen volume were read centrally by blinded readers. Results were not provided to the sponsor until data base lock. In order to minimize the potential for differential dose titration of standard therapy and ruxolitinib, equivalent guidance for dose increases and dose reductions were provided in the study protocol.

Risks of bias in RESPONSE study

We first analyzed the internal validity, in other words, risk of bias inherent to the trial design, of RESPONSE. We used the Cochrane Collaboration's risk of bias assessment tool on the domains of allocation concealment, blinding of subjects and personnel, blinding of outcome assessment, incomplete outcomes data and selective reporting [25].

In RESPONSE trial, therapy assignment was not blinded. Subjects knew they were receiving ruxolitinib and were therefore likely be more adherent to therapy than those receiving standard therapy. Because physicians also knew the therapy assignment, they might have been more likely to discontinue therapy in the standard therapy cohort. This could explain why at the data cutoff date, the proportion of subjects who discontinued the treatment in the standard therapy group was higher than of those in the ruxolitinib group (96.4 vs 15.5%; $p < 0.001$). Ninety-eight out of the 108 patients who discontinued the treatment were reported to do it for lack of efficacy. However, in the majority of these patients, lack of efficacy was an expected occurrence, due to the fact that for 58.9% of the patients the standard therapy was HU, in other words, the therapy to which they were resistant or intolerant, and for 15.2% no medication. To blind subjects and physicians in intervention trials when one therapy is active and the other not active is a difficult goal. It has been suggested that when there is the possibility of this type of bias, subjects should be treated according to a strictly enforced prospectively defined protocol to ensure interactions between subjects and physicians in both arms of the study are as similar as possible [26–28]. However, this precaution was not taken in RESPONSE. This performance failure could result in systematic difference in factors other than the intervention of interest.

Rating quality of evidence

We analyzed external validity of RESPONSE by evaluating whether the results of the trial could be reasonably applied to persons with PV having the specific therapeutic need of resistance or intolerance to HU. GRADE recommends that the strength of evidence would be assessed according to categorized questions (PICOs) that should include four essential constituents: type of participant (P); intervention (I); comparator (C) and outcome (O). We focused on the specific needs of therapy in patients who were resistant or intolerant to HU therapy, selected by considering the importance of clinical problem, interest of hematologists and indications approved by FDA and EU. As suggested in GRADE, we evaluated four dimensions of trials quality for the question: imprecision; inconsistency; indirectness and publication bias because these address most issues reflecting on the quality of evidence.

The population of interest, best comparator and critical outcome we assumed to be relevant to persons with PV resistant or intolerant to HU therapy are shown in Table 1.

Precision of the estimate of ruxolitinib effect

In RESPONSE, the composite primary end point of both hematocrit control and spleen response at week 32 occurred in a higher proportion of subjects in the ruxolitinib group than in the standard-therapy group (20.9 vs 0.9%; $p < 0.001$). Hematocrit control occurred in a higher proportion of subjects in the ruxolitinib group than in the standard therapy group (60.0 vs 19.6%). Likely, a reduction of 35% or more in spleen volume from baseline occurred in 38.2% of subjects in the ruxolitinib group and 0.9% in the standard therapy group. More patients in the ruxolitinib group than in the standard therapy group had complete hematologic response (23.6 vs 8.9%; $p = 0.003$). At week 32, a total of 49% of patients in the ruxolitinib group, and 5% in the standard therapy group had at least a 50% reduction in the MPN-SAF total symptom score.

We used the GRADE approach to rate the confidence in the estimate of effect by analyzing these results for precision. The GRADE approach is based on the analysis of confidence interval of the effect [29]. We recalculated the results of efficacy of ruxolitinib versus standard therapy (Table 2). The response ratios for the measured outcomes ranged from 3.0 for hematocrit response to 42.7 for spleen response, and the

Table 1. PICO for the role of ruxolitinib in polycythemia vera patients who are refractory or resistant to hydroxyurea therapy.

Constituents of PICO	Selections
Patients	Polycythemia vera patients who have had an inadequate response or are intolerant to hydroxyurea
Intervention	Ruxolitinib.
Comparator	Second-line therapy for polycythemia vera, i.e., interferon or pipobroman, busulphan or ^{32}P
Outcome	Time to vascular event or response

Table 2. Recalculated response ratios of the RESPONSE trial of ruxolitinib in polycythemia vera.

Outcome	Ruxolitinib therapy	Standard therapy	Response ratio (95% CI)
Combined response (spleen response and hematocrit response)	23/100	1/112	23.41 (3.21–170.42)
Hematocrit response	66/110	22/112	3.05 (2.03–4.57)
Spleen response	42/110	1/112	42.76 (5.98–305.3)
Patient reported outcome (MPN-SAF)	36/74	4/81	25.17 (3.48–181.82)

MPN-SAF: Myeloproliferative neoplasm-symptom assessment form.
Data taken from [21].

lower confidence interval boundary closest to no effect (response ratio = 1) ranged approximately from two- to six-times the controls. Thus, the evidence on the precision of the outcomes effect size after ruxolitinib was high and we upgraded the quality of evidence.

Directness of the comparator

We assessed how closely the comparator in RESPONSE trial resembled that of interest, namely, what proportion of the trial subjects received an appropriate second-line therapy for HU resistance or intolerance.

RESPONSE trial design compared ruxolitinib with standard therapy, in other words, the best available therapy. Choosing between several second-line therapies was left to the subject's physician who selected a therapy after randomization using unspecified criteria. Standard therapy included HU in 58.9% of the patients, interferon (IFN) in 12.6%, anagrelide in 7.1%, immunomodulators in 4.5% and pipobroman in 1.8%. No medication was administered in 15.2% of the patients. Six patients received more than one standard therapy. At variance, in our reference outcome definition we selected IFN as the best therapy for patients resistant or intolerant to HU.

The choice of second-line myelosuppressive drugs for PV is critical because some drugs administered after HU may enhance the risk of acute leukemia [30]. The dominant reason why we selected IFN in our reference comparator derived from the recently issued recommendations from an ELN panel of experts who stated that IFN-α should be considered as the first choice in the second-line therapy of PV because this drug is reported to be nonleukemogenic [12]. Pipobroman, busulfan and ^{32}P were indicated as second-line therapies reserved for patients with short life expectancy [12]. These recommendations were also supported by the opinion of individual experts in the field [31–33].

With conventional (nonpegylated) IFN preparations, roughly 60% of patients receiving the drug as *de novo* or second-line therapy achieved complete freedom from phlebotomies (80% objective responses) and a significant proportion of them also had improvement in pruritus and splenomegaly [34]). Recent experience with pegylated IFN-α suggests that these preparations may be better tolerated than nonpegylated preparations, while retaining high clinical response rates [35,36]. Hematologic CRs are seen in 80–90% of patients, with the majority showing molecular responses (complete in 15–20%).

Conceivably, a substantial proportion of subjects in the control arm of RESPONSE trial received a therapy different from that their doctors would use and that has been recommended by experts. This represents a potential indirectness of the comparator.

Directness of the end point

Next we assessed how appropriate was the RESPONSE end point (directness of outcome). The primary end point of RESPONSE trial was the combined response on splenomegaly and hematocrit. Besides, a number of response dimensions were included in the trial's secondary end points (WBC, platelet count, symptoms) that reflected the response criteria existing at the time the study was initiated [10].

The value of response as a clinically meaningful end point in PV is an open question. Its appraisal should consider how strongly patients with response are likely to show greater clinical benefit than those without response, and how response relates to the pathophysiological mechanism(s) by which the treatment affects the response.

The evidences on the value of response as a clinically relevant end point in PV are contradictory. In a retrospective study with PV patients, achieving response on hematocrit or hematological parameters did not result in better survival or less thrombosis and bleeding [17]. By contrast, evidence of a correlation between hematological response, in particular reaching a hematocrit lower than 45% or lowering the WBC count, and clinical benefit in term of vascular events has been provided by the CYTO-PV trial [8]. In this trial, patients with *JAK2*V617F-positive PV were randomly assigned to receive either more intensive treatment (target hematocrit; <45%) or less intensive treatment (target hematocrit, 45–50%). The primary composite end point was the time until death from

cardiovascular causes or major thrombotic events. After a median follow-up of 31 months, the primary end point was recorded in five of 182 patients in the low-hematocrit group (2.7%) and 18 of 183 patients in the high-hematocrit group (9.8%) (HR in the high-hematocrit group; 3.91). This provided indirect evidence that control of blood counts is a reliable surrogate for a clinically relevant end point such as reduction of vascular events.

Hematological response as a clinically relevant end point in PV seems to be confirmed by the results of the RESPONSE trial itself, in which the number of cardiovascular events was lower in the arm of ruxolitinib with higher response than in the arm with standard therapy with lower response. Thromboembolic events occurred in one patient in the ruxolitinib group versus six patients in the standard therapy group.

Further evidence on the relevant clinical benefit of decreasing the hematocrit value in PV is the biological mechanism underlying the response. Red cell aggregation increases at high hematocrit levels, creating the potential for vascular stasis. As a result, enhanced interplay between platelet, leukocytes and vessel wall increases the risk of thrombosis [37–39].

In conclusion, we did not find reasons for downgrading the quality of evidence on the benefit of ruxolitinib in patients with resistance or intolerance to HU in PV for indirectness of the outcome.

Adverse events

Both the Phase II [20] and the randomized Phase III study (RESPONSE) [21] of ruxolitinib therapy in PV documented adverse events. The most common grade 3–4 hematological adverse events included anemia, thrombocytopenia and neutropenia. Pooling data from these two studies, grade 3–4 thrombocytopenia occurred in 6.2% of patients, anemia in 3.4% of cases and neutropenia in 2.08%. Analyzing the RESPONSE study, the incidence of reported grade 3–4 hematological adverse events in the ruxolitinib group was 9.9% compared with 5.4% in standard therapy control group (risk ratio: 1.5; 95% CI: 0.57–4.05), suggesting a nonstatistically significant increase in the number of reported severe hematological adverse events in the ruxolitinib group. The RESPONSE study reported infections and nonmelanoma skin cancers in ruxolitinib group, but the risk difference with the standard therapy group was not significant.

Discussion

We presented a systematic critical appraisal of the use of ruxolitinib in PV patients who had an inadequate response or were intolerant to HU. We rated our results for the outcomes of the RESPONSE randomized trial [21] on which FDA and EU approval of ruxolitnib in PV was based, as having an overall 'moderate' level of evidence indicating that 'further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate' [23]. We downgraded the overall quality of evidence for the outcomes of RESPONSE trial because we found risk of performance bias, and we downgraded the quality of evidence in areas such as indirectness in the comparator.

Trial indirectness results from the study conception itself. The authors started from the consideration that few therapeutic resources are available for HU refractory or intolerant PV patients. The comparator chosen in the protocol was a single agent selected among the continued HU, IFN, pipobroman, anagrelide and approved immunomodulators. This choice was driven in large part by the potential increased risk of transformation to acute leukemia with the use of other available cytoreductive agents. Likely, the use of single agent was because many of the available therapies when combined with HU possess an increased potential for leukemic transformation. ^{32}P, busulfan and chlorambucil were excluded form use in this study because of their increased leukemogenic potential and profound myelosuppression.

However, these concepts are not agreed by the scientific community of myeloproliferative neoplasms. Guidelines for treatment of PV published in 2013 [12] and experts' opinion [31–33] indicate IFN as the first choice for patients refractory or intolerant to HU. This choice derives from the nonleukemogenic mechanism of action of IFN and from its potential to suppress the malignant clone of PV with high rate of CR. It is conceivable that the choice of the second-line therapy in PV could depend from centers and countries. In Italy, for example, IFN is not authorized in PV and the Italian centers do not use currently IFN as a second-line therapy. However, this is an argument which should have favored designing a trial in which IFN was the designed comparator of ruxolitinib new experimental therapy.

Better insight of the role of IFN in the treatment of HU-resistant or intolerant PV patients is an aspect which will have close scrutiny as data from the ongoing randomized trial of pegylated IFN-α 2a versus HU in PV and ET patients will mature [40]. A further criticism to RESPONSE trial was the choice of the primary end point [33]. The trial had its focus on short-term control of hematocrit and splenomegaly, responses not unanimously considered as representing benefits to patients. The major reason for using

response as primary end point in the RESPONSE trial was that the importance of response rate as a surrogate end point in trials of cancer treatments is acknowledged by the regulatory agencies and is used for accelerated approval of drugs [41]. In fact, ruxolitinib gained FDA and EU approval for use in patients with PV based on the overall response rate from the RESPONSE study. However, these approvals do not elude the robust debate on whether response is a meaningful outcome in its own right, or purely a surrogate end point for OS and disease control [39]. This is particularly compulsory in PV where the response criteria are heterogeneous, reflecting improvement in different dimension of the disease. Recently published recommendations from the ELN/International Working Group for Myelofibrosis Research and Therapy on the best end point for clinical trials in MPNs recommended direct measures of benefit, and indicated response as the best end point for Phase II efficacy trials, but discouraged its use in Phase III comparative trials [41]. More specifically, the guidelines recommended event-free survival like thrombosis-free survival for Phase III trials in PV. In spite of this, in our analysis we highlighted the existing clinical and biological evidence that reducing the hematocrit in patients with PV is associated with a benefit on cardiovascular events [4]; thus, we did not downgrade the quality of evidence for the indirectness of outcome.

The strength of our analysis relies on the rigorous framework of critical appraisal such as GRADE. GRADE allows for the explicit use of factors that can increase or decrease the quality of the evidence. The GRADE approach has been adopted by more than 70 national and international organizations. Similarly, the Cochrane Collaboration now requires authors to use GRADE for all important outcomes in their systematic reviews.

There are, of course, limitations to our analysis. The dominant one is the paucity of eligible studies, resulting in a small sample size. Our appraisal of precision of the effect size for the use of ruxolitinib in PV could be incorrect if the drug was given to more persons. Improvement in the quality of evidence requires studies with more subjects or more studies with the same end point and design.

Another limitation is that GRADE analysis is based on a high amount of subjectivity. This is particularly true for factors other than study design that should affect our confidence in the estimates of effect.

Conclusion

The choice of second-line therapy in PV is an unmet clinical need. Ameliorating the outcomes of these patients can be of substantial benefit to affected persons. Trials in patients who have had an inadequate response or are intolerant to HU report sustainable efficacy and safety of ruxolitinib in these patients. However, the large apparent reduction in spleen volume, hematocrit and symptoms from ruxolitinib compared with standard therapy likely overestimates the effect size because of the risk of biases and moderate quality of evidence. Using the GRADE approach we uncovered factors affecting the quality of evidence which were otherwise unstated.

A definitive analysis of whether ruxolitinib alters the natural history of patients with PV who are refractory or intolerant to HU, requires one or more large randomized trials with event-free survival as the primary or co-primary end point comparing what is now considered in most of the centers the second-line best available therapy, namely pegylated IFN.

Financial & competing interests disclosure

G Barosi was supported by a grant from Associazione Italiana per la Ricerca sul Cancro (AIRC, Milano) 'Special Program Molecular Clinical Oncology 5x1000' to AGIMM (AIRC-Gruppo Italiano Malattie Mieloproliferative), and participated in advisory boards from Novartis. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- RESPONSE Phase III study showed that ruxolitinib was effective in controlling the hematocrit, reducing spleen size and improving symptoms in patients with polycythemia vera who had inadequate response to or had unacceptable side effects from hydroxyurea.
- By applying the GRADE approach we upgraded the quality of evidence because of large effect size on splenomegaly, hematocrit and symptoms but we downgraded it for performance bias and indirectness of the comparator.
- The overall moderate level of evidence of RESPONSE trial indicates that further research is likely to have an important impact on our confidence in the estimate of effect.

References

Papers of special note have been highlighted as:
• of interest

1 Tefferi A, Barbui T. Polycythemia vera and essential thrombocythemia: 2015 update on diagnosis, risk-stratification and management. *Am. J. Hematol.* 90(2), 162–173 (2015).

2 Kralovics R, Passamonti F, Buser AS *et al.* A gain-of-function mutation of *JAK2* in myeloproliferative disorders. *N. Engl. J. Med.* 352(17), 1779–1790 (2005).

3 Baxter EJ, Scott LM, Campbell PJ *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 365(9464), 1054–1061 (2005).

4 James C, Ugo V, Le Couédic JP *et al.* A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 434(7037), 1144–1148 (2005).

5 Levine RL, Wadleigh M, Cools J *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7(4), 387–397 (2005).

6 Tefferi A, Rumi E, Finazzi G *et al.* Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia.* 27(9), 1874–1881 (2013).

7 Landolfi R, Marchioli R, Kutti J *et al.* Efficacy and safety of low-dose aspirin in polycythemia vera. *N. Engl. J. Med.* 350(2), 114–124 (2004).

8 Marchioli R, Finazzi G, Specchia G *et al.* Cardiovascular events and intensity of treatment in polycythemia vera. *N. Engl. J. Med.* 368(1), 22–33 (2013).

9 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* 121(23), 4778–4781 (2013).

10 Barosi G, Birgegard G, Finazzi G *et al.* Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. *Blood* 113(20), 4829–4833 (2009).

11 Emanuel RM, Dueck AC, Geyer HL *et al.* Myeloproliferative neoplasm (MPN) symptom assessment form total symptom score: prospective international assessment of an abbreviated symptom burden scoring system among patients with MPNs. *J. Clin. Oncol.* 30(33), 4098–4103 (2012).

12 Barbui T, Barosi G, Birgegard G *et al.* Philadelphia-negative classical myeloproliferative neoplasms: critical concepts and management recommendations from European LeukemiaNet. *J. Clin. Oncol.* 29(6), 761–770 (2011).

13 Huang BT, Zeng QC, Zhao WH *et al.* Interferon α-2b gains high sustained response therapy for advanced essential thrombocythemia and polycythemia vera with *JAK2V617F* positive mutation. *Leuk. Res.* 38(10), 1177–1183 (2014).

14 Stauffer Larsen T, Iversen KF *et al.* Long term molecular responses in a cohort of Danish patients with essential thrombocythemia, polycythemia vera and myelofibrosis treated with recombinant interferon alpha. *Leuk. Res.* 37(9), 1041–1045 (2013).

15 Silver RT, Kiladjian JJ, Hasselbalch HC. Interferon and the treatment of polycythemia vera, essential thrombocythemia and myelofibrosis. *Expert Rev. Hematol.* 6(1), 49–58 (2013).

16 Barosi G, Birgegard G, Finazzi G *et al.* A unified definition of clinical resistance and intolerance to hydroxycarbamide in polycythaemia vera and primary myelofibrosis: results of a European LeukemiaNet (ELN) consensus process. *Br. J. Haematol.* 148(6), 961–963 (2012).

17 Alvarez-Larrán A, Pereira A, Cervantes F *et al.* Assessment and prognostic value of the European LeukemiaNet criteria for clinicohematologic response, resistance, and intolerance to hydroxyurea in polycythemia vera. *Blood* 119(6), 1363–1369 (2012).

18 Kiladjian JJ, Chevret S, Dosquet C *et al.* Treatment of polycythemia vera with hydroxyurea and pipobroman: final results of a randomized trial initiated in 1980. *J. Clin. Oncol.* 29(29), 3907–3913 (2011).

19 Antonioli E, Guglielmelli P, Pieri L *et al.* Hydroxyurea-related toxicity in 3,411 patients with Ph'-negative MPN. *Am. J. Hematol.* 87(5), 552–554 (2012).

20 Quintás-Cardama A, Vaddi K, Liu P *et al.* Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: therapeutic implications for the treatment of myeloproliferative neoplasms. *Blood* 115(15), 3109–3117 (2010).

21 Verstovsek S, Passamonti F, Rambaldi A *et al.* A Phase 2 study of ruxolitinib, an oral JAK1 and JAK2 inhibitor, in patients with advanced polycythemia vera who are refractory or intolerant to hydroxyurea. *Cancer* 120(4), 513–520 (2014).

22 Vannucchi AM, Kiladjian JJ, Griesshammer M *et al.* Ruxolitinib versus standard therapy for the treatment of polycythemia vera. *N. Engl. J. Med.* 372(5), 426–435 (2015).

• **The Phase III clinical trial on which US FDA and EU based the approval of ruxolitinib in refractory or intolerant polycythemia vera patients.**

23 Guyatt GH, Oxman AD, Vist GE *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650), 924–926 (2008).

24 Guyatt GH, Oxman AD, Kunz R *et al.* Going from evidence to recommendations. *BMJ* 336(7652), 1049–1051 (2008).

25 Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0).* Higgins JPT, Green S (Eds). The Chochrane Collaboration (2011).

26 Boutron I, Guittet L, Estellat C *et al.* Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med.* 4(2), e61 (2007).

27 Hróbjartsson A, Forfang E, Haahr MT *et al.* Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int. J. Epidemiol.* 36(3), 654–663 (2007).

28 Wood L, Egger M, Gluud LL *et al.* Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336(7644), 601–605 (2008).

29 Guyatt GH, Oxman AD, Kunz R *et al.* GRADE guidelines 6. Rating the quality of evidence – imprecision. *J. Clin. Epidemiol.* 64(12), 1283–1293 (2011).

30 Finazzi G, Caruso V, Marchioli R *et al.* Acute leukemia in polycythemia vera: an analysis of 1638 patients enrolled in a prospective observational study. *Blood* 105(7), 2664–2670 (2005).

31 Vannucchi AM. How I treat polycythemia vera. *Blood* 124(22), 3212–3220 (2014).

32 Sever M, Newberry KJ, Verstovsek S. Therapeutic options for patients with polycythemia vera and essential thrombocythemia refractory/resistant to hydroxyurea. *Leuk. Lymphoma* 55(12), 2685–2690 (2014).

33 Pardanani A, Tefferi A. Is there a role for JAK inhibitors in BCR-ABL1-negative myeloproliferative neoplasms other than myelofibrosis? *Leuk. Lymphoma* 55(12), 2706–2711 (2014).

34 Kiladjian JJ, Chomienne C, Fenaux P. Interferon-alpha therapy in bcr-abl-negative myeloproliferative neoplasms. *Leukemia* 22(11), 1990–1998 (2008).

35 Quintás-Cardama A, Kantarjian H, Manshouri T *et al.* Pegylated interferon alfa-2a yields high rates of hematologic and molecular response in patients with advanced essential thrombocythemia and polycythemia vera. *J. Clin. Oncol.* 27(32), 5418–5424 (2009).

36 Kiladjian JJ, Cassinat B, Chevret S *et al.* Pegylated interferon-alfa-2a induces complete hematologic and molecular responses with low toxicity in polycythemia vera. *Blood* 112(8), 3065–3072 (2008).

37 Pearson TC, Wetherley-Mein G. Vascular occlusive episodes and venous haematocrit in primary proliferative polycythaemia. *Lancet* 2(8102), 1219–1222 (1978).

38 Thomas DJ, Marshall J, Russell RW *et al.* Effect of haematocrit on cerebral blood-flow in man. *Lancet* 2(8045), 941–943 (1977)

39 Wilson MK, Karakasis K, Oza AM. Outcomes and end points in trials of cancer treatment: the past, present, and future. *Lancet Oncol.* 16(1), e32–e42 (2015).

40 ClinicalTrialsDatabase: NCT01259856. https://clinicaltrials.gov/ct2/show/NCT01259856

41 Barosi G, Tefferi A, Besses C *et al.* Clinical end points for drug treatment trials in BCR-ABL1-negative classic myeloproliferative neoplasms: consensus statements from European LeukemiaNET (ELN) and Internation Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT). *Leukemia* 29(1), 20–26 (2015).