

Nutzenbewertung



**Gemeinsamer
Bundesausschuss**

**von Arzneimitteln mit neuen Wirkstoffen nach
§ 35a SGB V**

**Bewertung von Arzneimitteln für seltene Leiden
nach § 35a Absatz 1 Satz 10 i.V.m.**

5. Kapitel § 12 Nr. 1 Satz 2 VerfO

Wirkstoff: Ruxolitinib

Datum der Veröffentlichung: 17. Dezember 2012

Inhalt

Tabellenverzeichnis.....	3
Abkürzungsverzeichnis	4
Hintergrund	5
Zusammenfassung der Bewertung	7
1. Einführung.....	10
2. Nutzenbewertung.....	13
2.1. Fragestellung	13
2.2. Zulassungsbegründende Studien	13
2.3. Liste der verwendeten Quellen	14
2.4. Studiencharakteristika.....	14
2.5. Endpunkte und Erhebungsinstrumente	22
2.6. Ergebnisse zum Zusatznutzen	36
2.7. Einschätzung zum Ausmaß des Zusatznutzens durch den pharmazeutischen Unternehmer	50
3. Methodische Anmerkungen und Bewertung der Unterlagen.....	51
3.1. Design und Methodik der Studien	51
3.2. Endpunkte und Effekte.....	52
3.3. Zusammenfassende Darstellung der Studienergebnisse	61
4. Anforderungen an eine qualitätsgesicherte Anwendung	62
Referenzen.....	63
Anhang	66

Tabellenverzeichnis

- Tabelle 1: Charakterisierung der eingeschlossenen Studien
- Tabelle 2: Charakterisierung der Interventionen
- Tabelle 3: Charakterisierung der Studienpopulationen
- Tabelle 4: Matrix der vom pharmazeutischen Unternehmer als patientenrelevant betrachteten Endpunkte aus den Kategorien Mortalität und Morbidität
- Tabelle 5: Matrix der vom pharmazeutischen Unternehmer als patientenrelevant betrachteten Endpunkte aus den Kategorien Lebensqualität und Nebenwirkungen
- Tabelle 6: Matrix weiterer, in den Zulassungsstudien erhobener Endpunkte ohne Eingang in die vorliegende Bewertung
- Tabelle 7: Verzerrungspotenzial auf Studien- und Endpunktebene (Kategorien Mortalität und Morbidität)
- Tabelle 8: Verzerrungspotenzial auf Studien- und Endpunktebene (Kategorien Lebensqualität und Nebenwirkungen)
- Tabelle 9: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie Mortalität
- Tabelle 10: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie Morbidität
- Tabelle 11: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie gesundheitsbezogene Lebensqualität
- Tabelle 12: Ergebnisse zur Mortalität
- Tabelle 13: Ergebnisse zur Morbidität
- Tabelle 14: Ergebnisse zur gesundheitsbezogenen Lebensqualität
- Tabelle 15: Ergebnisse zu Nebenwirkungen
- Tabelle 16: Deskriptive Auflistung häufiger unerwünschter Ereignisse (Grad 3 und 4)
- Tabelle 17: Ergänzende Auswertung der Nebenwirkungen unter Ruxolitinib durch die EMA

Abkürzungsverzeichnis

BAT	Best-Available-Therapy
CRF	Case Report Form
CT	Computertomographie
CTCAE	Common Terminology Criteria for Adverse Events
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
EORTC	European Organisation for Research and Treatment of Cancer
EPAR	European Public Assessment Report (öffentlicher Beurteilungsbericht der EMA)
FACT	Functional Assessment of Cancer Therapy
FDA	Food and Drug Administration
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
IWG-MRT	International Working Group for Myelofibrosis Research and Treatment
JAK	Januskinase
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
MFSAFv2.0	Myelofibrosis Symptoms Assessment Form, Version 2.0
M(C)ID	Minimal (clinical) important difference
MRT	Magnetresonanztomographie
NSAR	Nicht steroidale Antirheumatika
PET-MF	Post-Essentielle-Thrombozythämie-Myelofibrose
PGIC	Patient Global Impression of Change
PMF	Primäre Myelofibrose
PPV-MF	Post-Polycythaemia-vera-Myelofibrose
PT	Preferred Term (der MedDRA)
pU	Pharmazeutischer Unternehmer
QLQ-C30	Quality of Life Questionnaire (Core Questionnaire with 30 Items)
QoL	Quality of Life
SGB V	Fünftes Sozialgesetzbuch
SOC	System Organ Class (Systemorganklasse der MedDRA)
STAT	Signal Transducer and Activators of Transcription (Signaltransduktoren und Aktivatoren der Transkription)
(S)UE	(schwerwiegendes) unerwünschtes Ereignis
TSS	Total-Symptom-Score, additiver Score der Gesamtsymptomschwere
VerfO	Verfahrensordnung des Gemeinsamen Bundesausschusses

Hintergrund

Ruxolitinib ist zugelassen als Arzneimittel zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Gemäß § 35a Absatz 1 Satz 10 SGB V gilt der medizinische Zusatznutzen durch die Zulassung als belegt.

Gemäß § 35a Absatz 2 SGB V entscheidet der Gemeinsame Bundesausschuss (G-BA), ob er die Nutzenbewertung selbst durchführt oder das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) beauftragt.

Ausgehend von der gesetzlichen Vorgabe in § 35a Absatz 1 Satz 10 SGB V, dass der Zusatznutzen eines Orphan Drug durch die Zulassung als belegt gilt, hat der G-BA in seiner Sitzung vom 15. März 2012 das Verfahren der Nutzenbewertung von Orphan Drugs dahingehend modifiziert, dass bei Orphan Drugs zunächst keine eigenständige Festlegung einer zweckmäßigen Vergleichstherapie mehr durch den G-BA als Grundlage der insoweit allein rechtlich zulässigen Bewertung des Ausmaßes eines gesetzlich zu unterstellenden Zusatznutzens erfolgt. Vielmehr wird ausschließlich auf der Grundlage der Zulassungsstudien das Ausmaß des Zusatznutzens durch den G-BA bewertet.

Dementsprechend hat der G-BA in seiner Sitzung am 15. März 2012 den mit Beschluss vom 1. August 2011 erteilten Auftrag an das IQWiG zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a Absatz 2 SGB V in der Weise abgeändert, dass bei Orphan Drugs eine Beauftragung des IQWiG mit der Durchführung einer Nutzenbewertung bei zuvor festgelegter Vergleichstherapie erst dann erfolgt, wenn der Umsatz des betreffenden Arzneimittels die gesetzliche Grenze von 50 Millionen Euro überschritten hat und das Arzneimittel damit einer uneingeschränkten Nutzenbewertung unterliegt (vgl. § 35a Absatz 1 Satz 11 SGB V).

Der G-BA bestimmt gemäß 5. Kapitel § 12 Absatz 1 Nummer 1 Satz 2 der Verfahrensordnung des G-BA (VerfO) das Ausmaß des Zusatznutzens für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Nutzenbewertung). Zur Bestimmung des Ausmaßes des Zusatznutzens von Ruxolitinib zieht der G-BA die Zulassung, die Unterlagen zur Zulassung (insbesondere den EPAR), die pivotalen Studien, auf deren Grundlage die Zulassung des Arzneimittels beruht, sowie das Dossier des pharmazeutischen Unternehmers heran und bewertet die darin enthaltenen Daten nach Maßgabe der im 5. Kapitel § 5 Absatz 7 Nr. 1 bis 4 VerfO festgelegten Kriterien im Hinblick auf ihre therapeutische Relevanz.

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Ruxolitinib in seiner Sitzung am 11. Dezember 2012 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 12. September 2012 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pharmazeutischen Unternehmers in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

Kosten der Therapie für die gesetzliche Krankenversicherung

Anzahl der GKV-Patienten in der Zielpopulation

Die Nutzenbewertung wird am 17. Dezember 2012 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (www.g-ba.de) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

Zusammenfassung der Bewertung

Bei Ruxolitinib handelt es sich um ein Arzneimittel aus der neuen Wirkstoffklasse der Januskinase 1/2-Inhibitoren zur Behandlung von krankheitsbedingter Splenomegalie oder Symptomen bei Erwachsenen mit primärer Myelofibrose (PMF), Post-Polycythaemia-vera-Myelofibrose (PPV-MF) und Post-Essentieller-Thrombozythämie-Myelofibrose (PET-MF).

Die Unterlagen zur Bewertung des Ausmaßes des Zusatznutzens für Ruxolitinib basieren auf den beiden Zulassungsstudien COMFORT I und COMFORT II. Die Bewertung wird gegenüber den dort verwendeten Komparatoren vorgenommen. Beide Studien sind prospektive randomisierte, kontrollierte Vergleichsstudien der Phase III und schließen Patienten der Zielpopulation ein (PMF-, PPV-MF- und PET-MF-Patienten). Die Studien wurden methodisch adäquat durchgeführt. Es wird auf Studienebene von einem niedrigen Verzerrungspotenzial ausgegangen. COMFORT I ist eine doppelblinde, Placebo-kontrollierte Vergleichsstudie. Die Erhebung des primären Endpunkts („Primäranalyse“) erfolgte nach 24 Wochen. Bei der COMFORT II-Studie handelt es sich um eine offene, aktiv kontrollierte Vergleichsstudie gegenüber der nach Ermessen des Prüfarztes besten verfügbaren Therapie, wobei auch eine Kombination mehrerer Therapieoptionen, ein Therapiewechsel sowie eine Nicht-Behandlung als Kontrolle infrage kamen. Die Erhebung des primären Endpunkts erfolgte nach 48 Wochen. In beiden Studien durften alle eingeschlossenen Patienten nach der randomisierten Phase mit Ruxolitinib (weiter)behandelt werden. Diese Extensionsphase dient insbesondere zur Erhebung von Langzeitdaten zur Sicherheit von Ruxolitinib (Ergebnisse liegen noch nicht vor).

Im Folgenden werden die wesentlichen Ergebnisse zusammenfassend dargestellt.

Mortalität

Die beiden Studien waren nicht dafür ausgelegt, Unterschiede hinsichtlich des Gesamtüberlebens aufzudecken. Es konnte zum Zeitpunkt der Erhebung des primären Endpunkts in keiner der Zulassungsstudien ein statistisch signifikanter Gruppenunterschied hinsichtlich der Gesamtmortalität beobachtet werden. Ergebnisse eines Sicherheitsupdates der COMFORT I-Studie 4 Monate nach der Primäranalyse weisen zwar auf einen Trend zur Verlängerung des Gesamtüberlebens unter Ruxolitinib hin, da in der Extensionsphase jedoch alle Patienten mit Ruxolitinib behandelt wurden, ist ein Gruppenvergleich in dieser Studienphase wenig aussagekräftig.

Die Endpunkte Leukämiefreies Überleben und Progressionsfreies Überleben werden als nicht valide Surrogatparameter für den patientenrelevanten Endpunkt Gesamtmortalität eingestuft.

Morbidität

Der primäre Endpunkt war in beiden Studien die Bestimmung des Anteils an Patienten mit einem Rückgang des Milzvolumens um mindestens 35 % nach 24 (COMFORT I) bzw. 48 (COMFORT II) Wochen. Für den primären Endpunkt zeigte sich jeweils eine statistisch signifikante Überlegenheit von Ruxolitinib gegenüber den verwendeten Komparatoren (COMFORT I 41,9 % vs. 0,7 % bzw. COMFORT II 28,5 % vs. 0 %). Für die Veränderung des Milzvolumens wurde in beiden Studien unter Ruxolitinib eine mediane Abnahme um ca. 30 % gegenüber dem Ausgangswert gezeigt, wohingegen das Milzvolumen in den Komparator-Armen zunahm (je 8,5 %). Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass es sich bei der Milzvolumenreduktion um einen Surrogatparameter unklarer Validität für die mit

einer vergrößerten Milz einhergehenden belastenden Symptome handelt. In einer post-hoc-Analyse der COMFORT I-Studie wurde der Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl) untersucht. Es zeigte sich kein eindeutiger Zusammenhang zwischen dem Rückgang des Milzvolumens um $\geq 35\%$ und den untersuchten Oberbauchbeschwerden (Abnahme um $\geq 50\%$), was die Validität des SurrogatEndpunkts bzw. die Eignung der gewählten Responsekriterien infrage stellt.

Die Endpunkte Leukämiefreies Überleben und Progressionsfreies Überleben sind kombinierte Endpunkte, die sich aus Endpunkten aus dem Bereich Morbidität sowie Mortalität zusammensetzen. Für die Kategorie Mortalität wurden diese Endpunkte als nicht validierte Surrogatparameter bewertet. Die Endpunkte wurden in der vorliegenden Bewertung in der Endpunktkategorie Morbidität bewertet, sind jedoch aufgrund ihrer Zusammensetzung sowie fehlender Ergebnisdarstellung zu den Einzelkomponenten in ihrer Patientenrelevanz fraglich und für eine Bewertung zum Ausmaß des Zusatznutzens nicht geeignet. Ergebnisse zu beiden Endpunkten lagen nur für die Studie COMFORT II vor und es konnte jeweils kein statistisch signifikanter Gruppenunterschied gezeigt werden. In der COMFORT I-Studie war das Leukämiefreie Überleben als Langzeitendpunkt mit einer Auswertung nach 144 Wochen geplant. Entsprechende Analysen liegen noch nicht vor.

Darüber hinaus wurde zur Erfassung der krankheitsbedingten Symptome in der Studie COMFORT I der Total-Symptom-Score (TSS) verwendet, der sechs typische Symptome der Myelofibrose (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl, Nachtschweiß, Juckreiz, Muskel- und Knochenschmerzen) zusammenfasst und in einem Summenscore zusammenfasst. Der Anteil an Patienten, der nach 24 Wochen eine mindestens 50 %ige Verbesserung der Symptome aufweist war in der Ruxolitinib-Gruppe statistisch signifikant größer als in der Placebo-Gruppe (45,9 % vs. 5,3 %) und auch bei der absoluten Veränderung des Summenscores zeigte sich ein statistisch signifikanter Gruppenunterschied. Der TSS bildet zwar relevante Symptome ab, die Validität des eingesetzten Messinstrumentes und die klinische Relevanz einer 50 %igen Verbesserung der Symptome lässt sich den vorliegenden Unterlagen jedoch nicht eindeutig entnehmen und ist nicht abschließend geklärt.

Des Weiteren wurde der ECOG Performance-Status in beiden Studien erhoben und die Veränderung im Studienverlauf dargestellt. Dieses Messinstrument dient primär dazu, den funktionellen Status des Patienten zu erfassen und wird daher der Endpunktkategorie Morbidität zugeordnet. Es zeigt sich in beiden Studien, dass mehr Patienten ihren ECOG Performance-Status unter Therapie mit Ruxolitinib erhalten oder verbessern konnten und sich weniger Patienten verschlechterten als in den jeweiligen Komparator-Gruppen.

Lebensqualität

Die Lebensqualität wurde in den beiden Studien mithilfe des generischen Fragebogens QLQ-C30 der EORTC bestimmt. In der COMFORT II-Studie wurde zusätzlich der für das Anwendungsgebiet Non-Hodgkin-Lymphom validierte Fragebogen FACT-Lym eingesetzt.

Insgesamt zeigen die Instrumente zur Erfassung der Lebensqualität studien- und endpunktübergreifend eine Verbesserung der Lebensqualität unter Ruxolitinib gegenüber dem jeweiligen Komparator, wobei hier die Ergebnisse aus der verblindeten COMFORT I-Studie die aussagekräftigeren Ergebnisse liefern. Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass bei beiden Studien für die zur Lebensqualität erhobenen Endpunkte

aufgrund der fehlenden bzw. schwer aufrecht zu erhaltenden Verblindung und der Verletzung des ITT-Prinzips in den Auswertungen von einem hohen Verzerrungspotenzial auszugehen ist. Auch die fehlende Krankheitsspezifität des QLQ-C30-Fragebogens ist bei der Interpretation der Ergebnisse zu berücksichtigen.

Nebenwirkungen

Es liegen deskriptive Angaben zu der Anzahl der Patienten, bei denen mindestens ein unerwünschtes Ereignis aufgetreten ist, vor. Ein statistischer Gruppenvergleich wurde nicht durchgeführt. Es kann insgesamt von einem niedrigen Verzerrungspotenzial für die erfassten unerwünschten Ereignisse ausgegangen werden.

In beiden Studien zeigte sich ein hoher Anteil an Patienten mit unerwünschten Ereignissen in jeweils beiden Studienarmen (> 90 %). Auch schwerwiegende Nebenwirkungen und Abbrüche aufgrund von unerwünschten Ereignissen traten studienübergreifend in beiden Behandlungsgruppen jeweils zu gleichen Anteilen auf (je ca. 30 %). Ein unterschiedliches Ergebnis zeigte sich jedoch in den beiden Studien hinsichtlich des Endpunkts „Unerwünschte Ereignisse der CTCAE-Grade 3 und 4“. Während es in der COMFORT I-Studie keinen bedeutsamen Unterschied zwischen den Studienarmen gab (47,1 % bzw. 44,4 %) zeigte sich in der COMFORT II-Studie eine auffällige Differenz zuungunsten von Ruxolitinib (43,2 % vs. 24,7 %). Für den Endpunkt „Unerwünschte Ereignisse, die zu einer Dosisreduktion führten“ zeigten sich in beiden Studien Gruppenunterschiede zuungunsten von Ruxolitinib (COMFORT I: 40 % vs. 9,3 %; COMFORT II: 63 % vs. 15,1 %). Dies spricht für eine insgesamt schlechtere Verträglichkeit von Ruxolitinib gegenüber dem jeweiligen Komparator.

In beiden Studien waren Anämie und Thrombozytopenie die beiden häufigsten hämatologischen unerwünschten Ereignisse und traten bei einem deutlich höheren Anteil an Patienten in der jeweiligen Interventionsgruppe auf. Diese Nebenwirkungen lassen sich auf den Wirkmechanismus von Ruxolitinib als JAK-Inhibitor zurückführen. Thrombozytämie und Anämie haben allerdings nur selten zu einem Studienabbruch geführt (< 1 %). Als unerwünschte Ereignisse unter Ruxolitinib sind weiterhin Blutungen und Infektionen besonders hervorzuheben. Diese wurden seitens der EMA als besondere Risiken identifiziert, die bei der Anwendung von Ruxolitinib genau beobachtet werden müssen.

Insgesamt bleibt bei der Betrachtung der Nebenwirkungen festzuhalten, dass es sich in beiden Studien um einen relativ kurzen Beobachtungszeitraum handelt und die Datenlage noch nicht ausreicht, um die Sicherheit von Ruxolitinib abschließend bewerten zu können. Dies ist insbesondere vor dem Hintergrund wichtig, dass es sich bei Ruxolitinib um einen Vertreter einer neuen Wirkstoffklasse handelt, zu der noch keine vergleichbaren Langzeitsicherheitsdaten zur Verfügung stehen.

Ergebnisse aus Subgruppenanalysen

Die Subgruppenergebnisse für den primären Endpunkt deuten darauf hin, dass Patienten mit JAK2V617F-Mutation studienübergreifend mehr von einer Ruxolitinib-Behandlung profitieren als Patienten ohne diese Mutation. Die Subgruppenanalyse in der COMFORT I-Studie lässt ein besseres Ansprechen von Frauen gegenüber Männern unter Ruxolitinib vermuten; ein Trend, der sich abgeschwächt auch in der COMFORT II-Studie zeigt. Auch eine höhere Thrombozytenzahl (entsprechend einer höheren Ruxolitinib-Anfangsdosierung) scheint mit einem besseren Ansprechen einherzugehen. Als Folge fehlender Interaktionstests im Dossier

und in den Studienberichten können aus den beschriebenen auffälligen Unterschieden jedoch keine konkreten Schlüsse hinsichtlich einer relevanten Interaktion abgeleitet werden.

1. Einführung

Bei Ruxolitinib (Handelsname Jakavi®) handelt es sich um einen Wirkstoff aus der Gruppe der Januskinase 1/2-Inhibitoren zur Behandlung der Myelofibrose. Die Zulassung durch die europäische Kommission wurde am 23. August 2012 erteilt. Seit dem 15. September 2012 ist der Arzneistoff auf dem deutschen Markt verfügbar.

Ruxolitinib ist angezeigt für die Behandlung krankheitsbedingter Splenomegalie oder Symptomen bei Erwachsenen mit

- primärer Myelofibrose (auch bekannt als chronische idiopathische Myelofibrose [PMF]),
- Post-Polycythaemia-vera-Myelofibrose (PPV-MF) oder
- Post-Essentieller-Thrombozythämie-Myelofibrose (PET-MF)

Bislang gibt es in Deutschland keine für diese Indikationen zugelassenen Medikamente. In den USA ist Ruxolitinib seit November 2011 zugelassen.

Myelofibrose ist eine Bluterkrankung, die auf einer Fehlfunktion der hämatopoetischen Stammzellen im Knochenmark beruht. Eine Myelofibrose kann primär (PMF) oder sekundär als Folge einer essentiellen Thrombozytopenie (PET) oder Polyzythämia vera (PV) auftreten. Diese Krankheitsbilder werden auch häufig als myeloproliferative Erkrankungen zusammengefasst (Tefferi und Vainchenker, 2011). Die Myelofibrose zeichnet sich durch eine übermäßige Bildung von Kollagen- und Retikulinfasen im Knochenmark (Fibrose) aus. Als Folge dessen kommt es zu einer gestörten Blutbildung (Mangel an Leukozyten, Thrombozyten und Erythrozyten). Es kommt zu einer Auslagerung der Hämatopoese (insbesondere in die Milz, nachgeordnet auch in die Leber) was zu einer deutlichen Vergrößerung der Organe (Splenomegalie, Hepatomegalie) und damit einhergehenden Symptomen führen kann. Bei progressivem Krankheitsverlauf kann sich die Splenomegalie immer weiter ausprägen und zu einem Milzinfarkt führen (Abdel-Wahab et al., 2009).

Auf molekularer Ebene kommt es unabhängig vom Ursprung der Myelofibrose zu einer klonalen Stammzellproliferation, verbunden mit einem erhöhten Spiegel an verschiedenen inflammatorischen und proangiogenetischen Zytokinen im Serum. Hauptursache ist bei Myelofibrose-Patienten eine Dysregulation bzw. Überaktivierung der JAK/STAT (Januskinase / Signal Transducer and Activators of Transcription) Signalweiterleitung und damit verbunden eine Dysregulation der Hämatopoese und der Funktionen des Immunsystems (Ghoreschi et al., 2009). Diese Überaktivierung des Signalweges lässt sich häufig auf eine Mutation im JAK2-Gen (V617F) zurückführen, die in mehr als 90% der PPV-MF-Patienten und in ca. 60% der Patienten mit PMF bzw. PPV-MF auftritt (Barbui et al., 2011). Myelofibrose-Patienten weisen jedoch auch unabhängig von ihrem JAK2V617F-Mutationsstatus eine Dysregulation der JAK-Signalwege auf (Fachinformation Jakavi®, 2012). Die Überaktivierung von JAK1 geht mit erhöhten Spiegeln proinflammatorischer Zytokine einher. JAK2 führt bei Überaktivierung zu verstärkter Proliferation und Differenzierung hämatopoetischer Vorläuferzellen. Als Januskinase 1/2-Inhibitor greift Ruxolitinib in diesen Signalweg regulierend ein. Der Wirkstoff hemmt die Januskinasen 1 und 2, indem er an die katalytische Tyrosinkinase (TK)-Domäne der Enzyme bindet. Dies stört die Funktionalität der Enzyme, die nicht mehr in der Lage sind,

die Tyrosinreste des Rezeptors oder entsprechende Substrate wie STAT zu phosphorylieren und so zu aktivieren. Durch die Hemmung der überaktivierten JAK1 wird die übermäßige Zytokinproduktion verringert; die Hemmung von JAK2 führt zur Reduktion der extramedullären Hämatopoese und klonalen Myeloproliferation (Gräfe und Siebenand, 2012).

Die Myelofibrose verläuft im initialen Stadium zunächst asymptomatisch. Das mediane Erkrankungsalter liegt bei 65 Jahren (EMA, 2012). Bei der körperlichen Untersuchung der betroffenen Patienten fallen meist die Splenomegalie, die Hepatomegalie und Veränderungen im Blutbild wie Anämie und Thrombozytose auf (Griesshammer et al., 2010). Die typischen klinischen Symptome der Myelofibrose (PMF, PPV-MF und PET-MF) werden sowohl auf die Splenomegalie als auch auf erhöhte Blutspiegel von entzündungsfördernden Zytokinen zurückgeführt. Neben Schwächegefühl und ständiger Müdigkeit (Fatigue) sind dies in erster Linie nächtliche Schweißausbrüche, unangenehmes Wärmegefühl, Fieber, Juckreiz am ganzen Körper, Muskel- und Knochenschmerzen, vaskuläre Ereignisse (Thrombosen, Blutungen), Atemnot, Husten, abdominelle Beschwerden / Oberbauchschmerzen, Völlegefühl (vorzeitige Sättigung) und Gewichtsabnahme, die insgesamt zu beträchtlichen Einschränkungen der Lebensqualität und der Leistungsfähigkeit des Patienten in Beruf und Alltag führen. Hierbei scheinen Fatigue, Pruritus, Nachtschweiß und Knochenschmerzen für die betroffenen Patienten sowohl bei PMF als auch bei PET-MF und PPV-MF besonders belastend zu sein (Mesa et al., 2007).

Die International Working Group for Myelofibrosis Research and Treatment (IWG-MRT) hat fünf unabhängige Risikofaktoren definiert, die mit einer verkürzten Lebensdauer einhergehen (Cervantes et al., 2009):

- Alter > 65 Jahre
- Konstitutionelle Symptome (Gewichtsverlust, Fieber, Nachtschweiß)
- Anämie (Hämoglobin < 10 g/dl oder Notwendigkeit von Erythrozytentransfusionen)
- Leukozytose (Leukozyten > 25000/ μ l)
- Blasten im peripheren Blut $\geq$ 1%

Mit diesen Risikofaktoren können vier unterschiedliche Prognosegruppen (IWG-MRT-Risikostufen) identifiziert werden (Griesshammer et al., 2010):

- Niedrigrisiko (kein o.g. Kriterium liegt vor, mediane Überlebenszeit 135 Monate)
- Intermediärrisiko 1 (ein o.g. Kriterium liegt vor, mediane Überlebenszeit 95 Monate)
- Intermediärrisiko 2 (zwei o.g. Kriterien liegen vor, mediane Überlebenszeit 48 Monate)
- Hochrisiko ($\geq$ 3 der o.g. Kriterien liegen vor, mediane Überlebenszeit 27 Monate)

Therapie der Myelofibrose (PMF, PPV-MF und der PET-MF)

Primäre Therapieziele sind die Behandlung der krankheitsbedingten Splenomegalie bzw. die mit der Erkrankung einhergehenden Symptome. Des Weiteren stehen insbesondere bei der PPV-MF und der PET-MF die Vermeidung thrombotischer oder hämorrhagischer Ereignisse im Fokus der Therapie.

Die Therapie der Myelofibrose erfolgt risiko- und altersadaptiert. Patienten mit einem Niedrig-risiko- und einen Intermediärrisiko 1-Profil ohne klinische Probleme sollten aufgrund der relativ guten Prognose einer Beobachtungs-Strategie (Watch and Wait) zugeführt oder in ein entsprechendes Studienkonzept aufgenommen werden (Griesshammer et al., 2010). Die einzige potenziell kurative Therapie der PMF und der PPV-MF ist die allogene Stammzelltransplantation, die allerdings mit einer transplantationsassoziierten Mortalität von bis zu 20 % einhergeht (Griesshammer et al., 2010) und daher nur für ausgewählte Patienten infrage kommt. Patienten, bei denen keine Stammzelltransplantation durchgeführt werden kann, können aktuell nur palliativ mit dem Ziel einer Besserung einzelner Symptome behandelt werden. Dabei stehen die Therapie der Anämie, die Besserung der konstitutionellen Symptome sowie der symptomatischen Splenomegalie im Vordergrund.

Zur Thromboseprophylaxe gilt bei Patienten, die keine Kontraindikationen haben, die tägliche Gabe niedrigdosierter Acetylsalicylsäure (100 mg/Tag) vielfach als indiziert (Griesshammer et al., 2010). Ein häufig gewählter Ansatz zur Behandlung einer Splenomegalie und Hyperproliferation ist die Gabe des zellreduzierenden Medikaments Hydroxycarbamid (Hydroxyurea), wobei die Ansprechraten (50 % Milzverkleinerung) auf diese Therapie entsprechend den Angaben aus der Literatur stark schwanken (Martinez-Trillos et al., 2010; Sirhan et al., 2008). Aufgrund der nicht ausreichenden Wirksamkeit und des schlechten Nebenwirkungsprofils von Hydroxyurea werden aktuell verschiedene alternative Substanzen zur Behandlung der Myelofibrose eingesetzt. Hierzu gehören neben anderen myelosuppressiven Medikamenten wie z.B. Anagrelid, Imatinib oder 2-Chlorodeoxyadenosin, Alkylantien wie Busulfan und Melphalan, immunmodulatorische Substanzen wie Thalidomid, Lenalidomid oder Interferon-alpha sowie zur Behandlung der Anämie Erythropoese-stimulierende Substanzen oder Androgene wie Danazol. Der pharmazeutische Unternehmer führte in seinem Dossier eine systematische Recherche nach RCTs und prospektiv vergleichenden Kohortenstudien mit diesen Substanzen durch und konnte für keine Therapie Studien identifizieren, die Wirksamkeit und Sicherheit dieser Therapieoptionen nachweisen. Die meisten Substanzen besitzen für das Anwendungsgebiet keine Zulassung (in Deutschland). Für Hydroxyurea-resistente Patienten kommt über eine medikamentöse Therapie hinaus zudem eine Splenektomie bzw. eine Milzbestrahlung infrage (Griesshammer et al., 2010).

2. Nutzenbewertung

2.1. Fragestellung

Ruxolitinib ist ein Arzneimittel mit dem Status „Orphan Drug“ zur Behandlung von krankheitsbedingter Splenomegalie bzw. Symptomen bei Erwachsenen mit primärer Myelofibrose (PMF, auch bekannt als chronische idiopathische Myelofibrose), Post-Polycythaemia-vera-Myelofibrose (PPV-MF) oder Post-Essentieller-Thrombozythämie-Myelofibrose (PET-MF). Gemäß § 35a Abs. 1 Satz 10 SGB V gilt für Orphan Drugs der medizinische Zusatznutzen durch die Zulassung als belegt. Der G-BA bestimmt bei Orphan Drugs, die den Umsatz von 50 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Unterlagen. Für diese Fälle wird keine zweckmäßige Vergleichstherapie zur Bewertung des Zusatznutzens herangezogen. In die Bewertung werden gemäß Zulassungsstatus erwachsene Männer und Frauen mit primärer Myelofibrose, Post-Polycythaemia-vera-Myelofibrose oder Post-Essentieller-Thrombozythämie-Myelofibrose eingeschlossen. Die Bewertung von Ruxolitinib erfolgte gegenüber den in den Zulassungsstudien verwendeten Komparatoren.

Die Zulassungsstudien werden mit ihren Charakteristika und den gewählten Endpunkten dargestellt. Für die eingeschlossenen Studien und die als relevant betrachteten Endpunkte wird jeweils das Verzerrungspotenzial bestimmt und die relevanten Ergebnisse dargestellt und einer Bewertung unterzogen.

2.2. Zulassungsbegründende Studien

Pivotal (Phase-III-)Studien:

- COMFORT I (INCB 18242-351): multizentrische, parallele, doppelblinde, randomisierte kontrollierte Studie von Ruxolitinib gegenüber Placebo.
- COMFORT II (CINC424A2352): multizentrische, parallele, offene, randomisierte, kontrollierte Studie von Ruxolitinib gegenüber einer vom Prüfarzt gewählten „besten verfügbaren Therapie“ (BAT, Best-Available-Therapy).

Supportive Dosisescalationsstudie:

- INCB18424-251: multizentrische, offene, nicht-randomisierte Dosisescalationsstudie (Phase I /II, „proof of concept“). Diese Studie wird im EPAR nur ergänzend betrachtet. Sie ist aufgrund ihres Designs nicht zur Beantwortung der Fragestellung geeignet und geht nicht in die vorliegende Bewertung ein.

2.3. Liste der verwendeten Quellen

Studie COMFORT I

- Studienbericht inkl. Studienprotokoll (A Randomized, Double-Blind, Placebo-Controlled Study of the JAK Inhibitor INCB018424 Tablets Administered Orally to Subjects With Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF), Or Post-Essential Thrombocythemia-Myelofibrosis (PET-MF))
- Publiziert als „Verstovsek S, Mesa RA, Gotlib J, Levy RS, Gupta V, DiPersio JF, et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. N Engl J Med 2012, 366(9):799-807.“
- European Public Assessment Report: Ruxolitinib (EMA, 2012)
- Herstellerdossier zu Ruxolitinib

Studie COMFORT II

- Studienbericht inkl. Studienprotokoll “A Randomized Study of the JAK Inhibitor INCB018424 Tablets Compared to Best-Available-Therapy in Subjects with Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF) or Post-Essential Thrombocythemia Myelofibrosis (PET-MF)”
- Publiziert als „Harrison C, Kiladjian JJ, Al-Ali HK, Gisslinger H, Waltzman R, Stalbovskaya V, et al. JAK inhibition with ruxolitinib versus Best Available Therapy for myelofibrosis. N Engl J Med 2012, 366(9):787-798.“
- European Public Assessment Report: Ruxolitinib (EMA, 2012)
- Herstellerdossier zu Ruxolitinib

2.4. Studiencharakteristika

Die Unterlagen zum Nachweis des Ausmaßes des Zusatznutzens für Ruxolitinib basieren auf den beiden Phase-III-Zulassungsstudien COMFORT I und COMFORT II. Diese werden in Tabelle 1 charakterisiert.

Tabelle 1: Charakterisierung der eingeschlossenen Studien

Studie	COMFORT I (INCB 18242-351)	COMFORT II (CINC424A2352)
Design	Parallel, doppelblind, multizentrisch, Placebo-kontrolliert, randomisiert.	Parallel, offen, multizentrisch, kontrolliert, randomisiert, stratifiziert nach Risikogruppe 2 und 3.
Population	Erwachsene mit primärer oder sekundärer Myelofibrose (PMF, PPV-MF, PET-MF) der Risikostufe 2 (Intermediärrisiko 2) oder 3 (Hochrisiko), die unter anderen Therapien resistent oder refraktär waren, diese Therapien nicht vertragen haben oder aus anderen Gründen nicht infrage kommen.	Erwachsene mit primärer oder sekundärer Myelofibrose (PMF, PPV-MF, PET-MF) der Risikostufe 2 (Intermediärrisiko 2) oder 3 (Hochrisiko), die nicht für eine allogene Stammzelltransplantation infrage kommen.
Interventionen (Zahl der randomisierten Patienten)	Ruxolitinib Tabletten, 2 x täglich; Beginn: 2 x 15 mg oder 2 x 20 mg, (entsprechend der Thrombozytenzahl); N=155 <u>Komparator:</u> Placebo 2 x täglich 3 oder 4 Tabletten; N = 154	Ruxolitinib Tabletten, 2 x täglich; Beginn: 2 x 15 mg oder 2 x 20 mg (entsprechend der Thrombozytenzahl); N=146 <u>Komparator:</u> Best-Available-Therapy (BAT) nach Ermessen des behandelnden Prüfarztes (Kombinationen, Therapiewechsel oder Nichtbehandlung erlaubt); N=73
Studiendauer	<u>Screening und Washout:</u> 28 Tage <u>Primäre Datenanalyse:</u> nachdem alle Patienten 24 Wochen und 50% der Patienten 36 Wochen beendet haben. <u>Early Crossover:</u> Alle Patienten mit einer definierten Symptomverschlechterung und einem Milzwachstum $\geq 25\%$ konnten bereits vor der primären Datenanalyse die Studie abbrechen oder zu offener Extensionsphase wechseln (unter Weiterführung oder Neuaufnahme einer Ruxolitinib-Therapie). <u>Extensionsphase (bis Woche 144):</u> Nach der primären Datenanalyse erfolgte die Entblindung und ein Wechsel der Patienten in der Placebo-Gruppe in den offenen Ruxolitinib-Arm war bei Eignung der Patienten möglich. Patienten in der Ruxolitinib-Gruppe konnten die Behandlung fortsetzen. <u>Nachbeobachtung:</u> 28 Tage nach der letzten Dosis. Die Behandlungszeit zur Sammlung von Langzeitdaten hält an; die Studie wird als „active, not recruiting“ geführt.	<u>Screening und Washout:</u> 28 Tage <u>Primäre Datenanalyse:</u> nachdem alle Patienten 48 Wochen beendet haben. <u>Early Crossover:</u> Alle Patienten mit einer definierten Progression (Milzwachstum $\geq 25\%$ oder Splenektomie) konnten die Studie abbrechen oder in die Extensionsphase wechseln (unter Weiterführung oder Neuaufnahme einer Ruxolitinib-Therapie). <u>Extensionsphase:</u> Weiterbehandlung mit Ruxolitinib für alle Patienten bis zu Woche 144 möglich (Nachbeobachtung auf Wunsch länger) <u>Nachbeobachtung:</u> 28 bis 37 Tage nach der letzten Dosis Die Behandlungszeit zur Sammlung von Langzeitdaten hält an; die Studie wird als „active, not recruiting“ geführt.
Ort und Zeitraum der Durchführung	Ort: 68 Zentren in den USA, 6 Zentren in Kanada und 15 Zentren in Australien Zeitraum: 24.08.2009 - 02.11.2010 (Primäranalyse); offene Nach-	Ort: 56 Zentren in 9 europäischen Ländern (Belgien, Deutschland, Frankreich, Großbritannien, Italien, Niederlande, Österreich, Schweden, Spanien)

Studie	COMFORT I (INCB 18242-351)	COMFORT II (CINC424A2352)
	beobachtungsphase läuft noch	Zeitraum: 01.07.2009 - 04.01.2011 (Primäranalyse); offene Nachbeobachtungsphase läuft noch
Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte gemäß Studienprotokoll	<u>Primärer Endpunkt:</u> • Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ (MRT, CT) nach 24 Wochen <u>Sekundäre Endpunkte:</u> • Dauer des Ansprechens (Milzvolumenreduktion) unter Ruxolitinib • Anteil der Patienten mit einer Verbesserung $\geq 50\%$ im TTS • (Prozentuale / absolute) Veränderung des TSS sowie einzelner Symptome des MFSAF v2.0 Fragebogens • Gesamtüberleben <u>Explorative Endpunkte:</u> • Prozentuale Milzvolumenveränderung (MRT / CT) bzw. Veränderung der tastbaren Milzlänge gegenüber Studienbeginn • Leukämiefreies Überleben (geplante Auswertung nach 144 Wochen) • Meta-Analyse des Gesamtüberlebens der Studien COMFORT I und COMFORT II • Veränderung des Körpergewichts gegenüber Studienbeginn • Veränderung der JAK2V617F-Allellast gegenüber Studienbeginn • Veränderungen der Konzentration von phosphoryliertem STAT3 • Veränderungen der (Knochenmarks)-Histopathologie • Veränderung des TSS gegenüber Studienbeginn (LOCF-Auswertung) • Veränderung im ECOG Performance-Status gegenüber Studienbeginn • Veränderungen des QLQ-C30 der EORTC gegenüber Studienbeginn • Veränderungen der Zytokinkonzentration bzw. der Konzentration weiterer Plasmaproteine gegenüber Studienbeginn • Veränderungen der Notwendigkeiten von Transfusionen innerhalb der Studie • Gesamtüberleben nach 144 Wochen der in die Ruxolitinib-Gruppe randomisierten Patienten (kategorisiert nach Risikostufe 2 / 3) • Pharmakokinetik von Ruxolitinib	<u>Primärer Endpunkt:</u> • Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ (MRT, CT) nach 48 Wochen <u>Sekundärer Endpunkte:</u> • Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ (MRT, CT) nach 24 Wochen • Dauer des Ansprechens (Milzvolumenreduktion) • Zeit bis zum ersten Ansprechen (Milzvolumenreduktion) • Progressionsfreies Überleben • Leukämiefreies Überleben • Gesamtüberleben • Veränderungen der Knochenmarkshistomorphologie <u>Explorative Endpunkte</u> • Prozentuale Milzvolumenveränderung (MRT / CT) bzw. Veränderung der tastbaren Milzlänge gegenüber Studienbeginn • Veränderung des Körpergewichts gegenüber Studienbeginn • Veränderung im ECOG Performance-Status gegenüber Studienbeginn • Veränderungen des QLQ-C30 der EORTC gegenüber Studienbeginn • Veränderungen im FACT-Lym Score gegenüber Studienbeginn • Veränderungen der Zytokinkonzentration bzw. der Konzentration weiterer Plasmaproteine gegenüber Studienbeginn • Veränderung der JAK2V617F-Allellast gegenüber Studienbeginn • Veränderungen in den Notwendigkeiten von Transfusionen innerhalb der Studie • Pharmakokinetik von Ruxolitinib <u>Sicherheit und Verträglichkeit:</u> • Erfassung der Frequenz, Dauer und Schweregrad der unerwünschten Ereignisse (inkl. Erfassung der Vitalzeichen, Elektrokardiogramm, Hämatologie, Serumchemie, Urinanalyse)

Studie	COMFORT I (INCB 18242-351)	COMFORT II (CINC424A2352)
	<u>Sicherheit und Verträglichkeit:</u> Erfassung der Frequenz, Dauer und Schweregrad der unerwünschten Ereignisse (inkl. Erfassung der Vitalzeichen, Elektrokardiogramm, Hämatologie, Serumchemie, Urinalyse) Darüber hinaus war eine Validierung des MFSAF v2.0 (bzw. des Response-kriteriums „Verbesserung um $\geq 50\%$“) anhand einer Korrelation der Veränderungen im MFSAF v2.0 und folgenden Skalen geplant: QLQ-C30 der EORTC, Patient Global Impression of Change, PROMIS Health Assessment Questionnaire, PROMIS Fatigue-Scale, Brief-Pain-Inventory.	
A priori geplante Subgruppenanalysen (gemäß statistischem Analyseplan)	<u>Untersuchter Endpunkt</u> Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ nach 24 Wochen. Folgende Merkmale sollten untersucht werden: Geschlecht, Alter, Myelofibrose-Typ, Hydroxyurea-Gabe (in den zurückliegenden 3 Monaten), Vorliegen der V617F-Mutation, Milzvolumen zu Studienbeginn, Tastbare Milzlänge, Land, Rasse, Ethnizität.	<u>Untersuchter Endpunkt</u> Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ nach 48 Wochen. Folgende Merkmale sollten untersucht werden: Geschlecht, Alter, Myelofibrose-Typ, Hydroxyurea-Gabe (in den zurückliegenden 3 Monaten), Vorliegen der V617F-Mutation, Milzvolumen zu Studienbeginn, Tastbare Milzlänge, Land, Rasse, Ethnizität.

BAT: Best-Available-Therapy, CT: Computertomographie, ECOG: Eastern Cooperative Oncology Group, LOCF: Last Observation Carried Forward, MFSAF: Myelofibrosis Symptoms Assessment Form, MRT: Magnetresonanztomographie, PET-MF: Post-Essentielle-Thrombozythämie-Myelofibrose, PMF: Primäre Myelofibrose, PPV-MF: Post-Polyzythämie vera Myelofibrose, TSS: Total-Symptom-Score

Beide Studien sind prospektive randomisierte, kontrollierte Vergleichsstudien der Phase III. COMFORT I ist eine doppelblinde, Placebo-kontrollierte Vergleichsstudie. Die Erhebung des primären Endpunkts wurde nach 24 Wochen geplant und durchgeführt. Bei der COMFORT II-Studie handelt es sich um eine aktiv kontrollierte Vergleichsstudie gegenüber der besten verfügbaren Therapie (Best-Available-Therapy, BAT) mit Erhebung des primären Endpunkts nach 48 Wochen. Für die Auswahl der besten verfügbaren Therapie war der jeweilige Prüfarzt zuständig, wobei auch eine Kombination mehrerer Therapieoptionen, ein Therapiewechsel sowie eine Nicht-Behandlung als Kontrolle infrage kamen. Eine Verblindung wurde aus diesem Grund nicht durchgeführt. Alle Patienten konnten nach der primären Analyse auf Wunsch und bei Eignung mit Ruxolitinib behandelt werden. Die Nachbeobachtungszeit (Extensionsphase) hält bei beiden Studien noch an. Sie dient insbesondere zur Generierung von Langzeitdaten zur Sicherheit und Verträglichkeit. In beiden Studien wurden erwachsene Patienten mit diagnostizierter PMF, PPV-MF oder PET-MF unabhängig von ihrem JAK2-Mutations-Status mit Intermediärrisiko 2 (Risikostufe 2) bzw. Hochrisiko (Risikostufe 3) gemäß IWG-MRT-Risikostufen eingeschlossen. Die Patienten mussten eine tastbare Milzlänge von

> 5 cm und eine Lebenserwartung von mindestens 6 Monaten aufweisen. In der COMFORT II-Studie wurden die randomisierten Patienten nach ihrer Risikoklasse (Intermediärrisiko 2 oder Hochrisiko) stratifiziert, wohingegen in der COMFORT I-Studie keine Stratifizierung vorgenommen wurde. Die Populationen der beiden Studien unterschieden sich dahingehend, dass die Patienten in der COMFORT I-Studie gegenüber anderen Therapien resistent oder refraktär sein mussten, diese nicht vertragen haben und – nach der Einschätzung des behandelnden Arztes – nicht für eine andere verfügbare Therapie infrage kommen durften. Diese Einschränkung galt nicht für die Studie COMFORT II. Hier durften die Patienten lediglich nicht für eine Stammzelltransplantation infrage kommen. Der primäre Endpunkt in beiden Studien war jeweils der Anteil an Patienten mit einer Reduktion des Milzvolumens um $\geq 35\%$ gegenüber Studienbeginn nach 24 (COMFORT I) bzw. 48 (COMFORT II) Wochen. Für den primären Endpunkt waren jeweils auch a priori Subgruppenanalysen geplant. Weitere Endpunkte in beiden Studien dienten der Erfassung der krankheitsbedingten Symptomatik, der Lebensqualität und der Untersuchung der Sicherheit.

In Tabelle 2 wird die in den Zulassungsstudien verwendete Intervention und die Medikation in den jeweiligen Vergleichsarmen beschrieben. Zudem werden erlaubte und nicht erlaubte Begleitmedikation innerhalb der Studie beschrieben. In beiden Studien war eine der Zulassung entsprechende zweimal tägliche orale Gabe von Ruxolitinib vorgeschrieben. In Abhängigkeit von der zu Studienbeginn gemessenen Thrombozytenzahl begannen die Patienten mit einer Tagesdosis von 30 mg oder 40 mg. Eine Dosisanpassung war jeweils in Abhängigkeit von der Wirksamkeit und der Verträglichkeit in 5 mg-Schritten erlaubt. Die jeweilige Höchsteinzeldosis lag bei 25 mg (Tagesdosis 50 mg). In der COMFORT I-Studie fand im Vergleichsarm eine entsprechende Placebo-Gabe statt (zweimal täglich je 3 bzw. 4 Tabletten). Im Vergleichsarm der COMFORT II-Studie konnte der Prüfarzt jeweils die nach seinem Ermessen beste verfügbare Therapie (BAT) auswählen, wobei der Einsatz experimenteller Substanzen nicht erlaubt war. Im BAT-Arm waren auch ein Therapiewechsel, eine Kombination mehrerer Therapeutika und eine Nicht-Behandlung erlaubt.

Tabelle 2: Charakterisierung der Interventionen

Studie	Intervention	Kontrolle
COMFORT I (INCB 18242-351)	Ruxolitinib 2 x täglich 15 mg (30 mg/Tag) oder 2 x täglich 20 mg (40 mg/Tag) Thrombozytenzahl > 200.000/ μ l Startdosis 2 x täglich 20 mg (40 mg/Tag) Thrombozytenzahl > 100.000 und < 200.000/ μ l: Startdosis 2 x täglich 15 mg (30 mg/Tag) Dosisanpassungen aufgrund fehlender Wirksamkeit oder Unverträglichkeit waren gemäß Studienprotokoll erlaubt Höchst-dosis: 25 mg (50 mg/Tag)	Placebo 2 x täglich je 3 oder 4 Tabletten Auch im Placeboarm wurden „Dosisanpassungen“ durchgeführt, um die Verblindung aufrecht zu erhalten.
	<u>Begleitmedikation:</u> - Jegliche Begleitmedikation musste im CRF dokumentiert werden <u>Eingeschränkte Begleitmedikation:</u> - Kortikoid-dosen durften 10 mg Prednisonäquivalente nicht überschreiten	

	- Niedrig dosierte ASS (< 150 mg/Tag) und NSAR waren erlaubt - Hämatopoetische Wachstumsfaktoren (z.B. Erythropoietin) waren bei schweren Thrombozytopenien / Neutropenien erlaubt - Dosisanpassung beim Einsatz potenter CYP3A4-Inhibitoren <u>Unerlaubte Medikation:</u> - Acetylsalicylsäure (> 150 mg/Tag) - Weitere experimentelle Substanzen (außer Ruxolitinib) - Hydroxyurea, Interferon, Thalidomid oder Anagrelid - Androgene (Ausnahme: Einsatz von Androgenen in einer zugelassenen Indikation, z.B. Endometriose) - Warfarin oder Heparin	
COMFORT II (CINC424A2352)	Ruxolitinib 2 x täglich 15 mg (30 mg/Tag) oder 2 x täglich 20 mg (40 mg/Tag) Thrombozytenzahl > 200.000/µl: Startdosis 2 x täglich 20 mg (40 mg/Tag) Thrombozytenzahl > 100.000 und < 200.000/µl: Startdosis 2 x täglich 15 mg (30 mg/Tag) Dosisanpassungen aufgrund fehlender Wirksamkeit oder Unverträglichkeit waren gemäß Studienprotokoll erlaubt. Höchst-dosis: 25 mg (50 mg/Tag)	Best-Available-Therapy (BAT) nach Ermessen des Prüfarztes (Kombinationen, Therapiewechsel und Nichtbehandlung waren erlaubt) N = 72, davon u.a.: Keine Medikation (N = 24) Medikation (N = 49) Hydroxycarbamid (N = 34) Glukokortikoide (N = 12) Epoetin alfa (N = 5) Thalidomid (N = 3) Lenalidomid (N = 2) Danazol (N = 3) Interferon alpha (N = 3) Mercaptopurin (N = 3)
	<u>Begleitmedikation:</u> - Jegliche Begleitmedikation musste im CRF dokumentiert werden <u>Einschränkungen / Anpassungen der Begleitmedikation:</u> - Überwachung und ggf. Dosisanpassung bei Patienten, die mit Warfarin oder Heparin behandelt werden (hinsichtlich Thrombozytopenien, Koagulationsparametern) - Hämatopoetische Wachstumsfaktoren (z.B. Erythropoietin) werden nicht empfohlen - Vorsichtiger Einsatz von potenten CYP3A4-Inhibitoren (ggf. Dosisanpassung) - Vorsichtiger Einsatz von milden / moderaten CYP3A4-Induktoren bzw. -Inhibitoren - Niedrige-ASS-Dosen (< 150 mg/Tag) und NSARs waren erlaubt <u>Unerlaubte Medikation:</u> - Hohe Kortikoid-dosen (> 10 mg Prednisonäquivalente) - Acetylsalicylsäure (> 150 mg/Tag) - Potente CYP3A4-Induktoren (Rifampicin, Johanniskraut) - Einsatz weiterer experimenteller Substanzen (außer Ruxolitinib) - Einsatz von Hydroxyurea, Interferon, Thalidomid, Busulfan, Lenalidomid oder Anagrelid war ausschließlich für Patienten der BAT-Gruppe erlaubt.	

ASS: Acetylsalicylsäure, BAT: Best-Available-Therapy, CRF: Case Report Form, NSAR: Nicht steroidale Antirheumatika

Tabelle 3: Charakterisierung der Studienpopulationen (Intention-to-Treat [ITT]-Population)

Studie	COMFORT I		COMFORT II	
Merkmal	Ruxolitinib N = 155	Placebo N = 154	Ruxolitinib N = 146	BAT N = 73
Studienabbrecher in der randomisierten Phase, N (%)	21 (13,5%) <u>Hauptgründe:</u> Tod: 9 (5,8%) UE: 8 (5,2%) Progression: 3 (1,9%) Widerruf: 1 (0,6%) Sonstiges: 0	37 (24,5%) <u>Hauptgründe:</u> Tod: 9 (6,0%) UE: 8 (5,3%) Progression: 12 (7,9%) Widerruf: 5 (3,3%) Sonstiges: 3 (2,0%)	26 (17,8%) ¹⁾ <u>Hauptgründe:</u> UE: 12 (8,2%) Progression: 1 (0,7%) Widerruf: 2 (1,4%) Protokollverletzung: 4 (2,8%) Sonstiges: 7 (4,8%)	24 (33,2%) ¹⁾ <u>Hauptgründe:</u> UE: 4 (5,5%) Progression: 3 (4,1%) Widerruf: 9 (12,3%) Sonstiges: 8 (11,2%)
Alter Median (min-max)	66 Jahre (43 - 91)	70 Jahre (40 - 86)	67 Jahre (35 - 83)	66 Jahre (35 - 85)
≤ 65 Jahre / > 65 Jahre	45,2% / 54,8%	33,8% / 66,2%	47,2% / 52,8%	49,3% / 50,7%
Geschlecht w/m	49,0% / 51,0%	42,2% / 57,1% ³⁾	43,2% / 56,8%	42,5% / 57,5%
Weißer	89,0%	90,3%	80,8% ⁴⁾	91,8% ⁴⁾
MF-Subtyp PMF, N (%) PPV-MF, N (%) PET-MF, N (%) fehlt	70 (45,2%) 50 (32,3%) 35 (22,6%) ---	84 (54,5%) ³⁾ 47 (30,5%) ³⁾ 22 (14,3%) ³⁾ 1 (0,6%)	77 (52,7%) 48 (32,9%) 21 (14,4%) ---	39 (53,4%) 20 (27,4%) 14 (19,2%) ---
Gewicht, Median (min; max)	71,4 kg (45,1; 124,4)	72,2 kg (46,0; 103,8)	68,7 kg (48,0; 109,0)	63,0 kg (45,0; 96,5)
Hydroxyurea- Vorthherapie N (%)	104 (67,1%)	87 (56,5%)	110 (75,3%)	50 (68,5%)
Milzvolumen Median (min-max)	2597,7 cm ³ (478 - 7462)	2566,3 cm ³ (521 - 8881)	2407,6 cm ³ (451 - 7765)	2317,9 cm ³ (728 - 7701)
Tastbare Milzlänge, Median (min-max)	16 cm (0 - 33)	16 cm (5 - 34)	14 cm (5 - 30)	15 cm (5 - 37)
IWG-MRT-Risiko- stufe 3 (hoch), N (%)	90 (58,%)	99 (64,3%)	88 (60,3%)	43 (58,9%)
IWG-MRT-Risiko- stufe 2 (inter- mediär), N (%)	64 (41,3%) ³⁾	54 (35,1%)	58 (39,7%) ³⁾	29 (39,7%) ³⁾
ECOG Performance- Status ²⁾ , N (%)				
0	47 (31,1%)	38 (25,5%)	58 (39,7%)	26 (35,6%)
1	87 (57,3%)	82 (55,0%)	77 (52,7%)	37 (50,7%)
2	14 (9,3%)	25 (16,8%)	10 (6,8%)	9 (12,3%)
3	3 (2,0%)	4 (2,7%)	1 (0,7%)	1 (1,4%)
fehlt	4 (2,6%)	5 (3,2%)	---	---
V617F-Mutation ⁴⁾	72,9 %	79,9 %	---	---

- 1) Todesfälle wurden hier nicht als Abbruchgründe genannt.
- 2) 0: Volle Aktivität, normales Leben möglich, 1: Eingeschränkte Aktivität, leichte Arbeit möglich, 2: Selbstversorgung möglich, nicht arbeitsfähig, nicht bettlägerig, muss > 50 % der Tageszeit ruhen, 3: Selbstversorgung möglich, nicht arbeitsfähig, nicht bettlägerig, muss < 50 % der Tageszeit ruhen, 4: Bettlägerig und völlig pflegebedürftig, 5: Tod
- 3) Daten von 1 Patient fehlen/unbekannt, da alle Unterlagen beim Umzug des Prüfzentrums verloren gingen.
- 4) Daten von mehreren Patienten fehlen aus Datenschutzgründen

BAT: Best-Available-Therapy, BMI: Body Mass Index, ECOG: Eastern Cooperative Oncology Group, m: Männlich, w: weiblich, IWG-MRT: International Working Group for Myelofibrosis Research and Treatment, PET-MF: Post-essentielle Thrombozytämie Myelofibrose, PMF: primäre Myelofibrose, PPV-MF: Post-Polyzythämia vera Myelofibrose, UE: unerwünschtes Ereignis,

Die Populationen zwischen den beiden Studienarmen waren in beiden Studien COMFORT I und COMFORT II im Wesentlichen vergleichbar. Es wird im Folgenden lediglich auf auffällige Unterschiede hingewiesen, deren Bedeutung nicht abschließend beurteilt werden kann.

Ein auffälliger Unterschied zeigte sich in der COMFORT I-Studie hinsichtlich der Altersverteilung. Die Patienten in der Interventionsgruppe waren im Median jünger als die Patienten in der Placebo-Gruppe. In der Primärveröffentlichung zur COMFORT I-Studie von Verstovsek wurde der Unterschied als statistisch signifikant ($p < 0,05$) beschrieben (Verstovsek et al., 2012). Die Patienten der Placebo-Gruppe hatten durchschnittlich einen schlechteren ECOG Performance-Status. Der Anteil an Patienten der Risikostufe 3 in der Placebo-Gruppe war höher (58,1 % vs. 64,3 %). Zusammenfassend ist festzuhalten, dass die Patienten in der Placebo-Gruppe der COMFORT I-Studie älter waren und tendenziell einen schlechteren Gesundheitszustand aufwiesen. Zudem war in der Studie COMFORT I ein größerer Anteil an männlichen Patienten in der Placebo-Gruppe verglichen mit der Interventionsgruppe. In beiden Studien waren die Patienten der Ruxolitinib-Gruppen häufiger mit Hydroxyurea vorbehandelt. Es ist unklar, wie sich dieses Ungleichgewicht auf die gemessenen Therapieeffekte auswirkt.

In beiden Studien brachen mehr Patienten in den Komparator-Armen (Placebo bzw. BAT) die Studie ab. In der COMFORT I-Studie ist eine vergleichbare Anzahl an Patienten wegen Tod (5,8 % vs. 6,0 %) und unerwünschten Ereignissen (5,2 % vs. 5,3 %) ausgeschieden. Ein höherer Anteil an Patienten in der Placebo-Gruppe hat die Studie jedoch aufgrund einer Krankheitsprogression verlassen (1,9 % vs. 7,9 %) bzw. die Einwilligung widerrufen (0,6 % vs. 3,3 %). In der COMFORT II-Studie war in der BAT-Gruppe der Hauptgrund für das Ausscheiden aus der Studie das Zurückziehen der Einwilligung der Patienten (1,4 % vs. 12,3 %).

2.5. Endpunkte und Erhebungsinstrumente

2.5.1. In der Nutzenbewertung berücksichtigte und nicht berücksichtigte Endpunkte (und deren Datenverfügbarkeit)

In Tabelle 4 und Tabelle 5 werden alle Endpunkte aufgeführt, die in die Nutzenbewertung eingeflossen sind. Dabei wurde die Darstellung der Endpunkte im Dossier des pharmazeutischen Unternehmers ergänzt um Endpunkte, die zusätzlich als relevant erachtet werden. In Tabelle 6 werden weitere Endpunkte aufgelistet, die nicht unmittelbar in die Nutzenbewertung eingeflossen sind. Diese Endpunkte sind entweder nicht direkt als patientenrelevant einzustufen (z.B. Fibrosegrad, Laborparameter, Vitalzeichen, Körpergewicht, Plasma-PD-Marker, Phosphorylierung von STAT3), sind zu Validierungszwecken erhoben worden (PROMIS Fatigue-Scale, Patient Global Impression of Change) oder führen aufgrund ihrer Operationalisierung nicht zu einem zusätzlichen Erkenntnisgewinn bei der Beantwortung der vorliegenden Fragestellung (z.B. Anzahl der Patienten mit einer tastbaren Milzlängenreduktion um $\geq 50\%$). Die in Tabelle 6 aufgelisteten Endpunkte können jedoch für die Interpretation der Ergebnisse von Interesse sein und wurden zum Teil im Extraktionsbogen ergänzend dargestellt.

Tabelle 3: Matrix der vom pharmazeutischen Unternehmer als patientenrelevant bezeichneten Endpunkte aus den Kategorien Mortalität und Morbidität, Datenverfügbarkeit

	Mortalität			Morbidität								
Studie	Gesamtüberleben	Leukämiefreies Überleben ¹⁾	Progressionsfreies Überleben ¹⁾	Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$	Prozentuale Milzgrößenveränderung	Beginn des Ansprechens (Rückgang des Milzvolumens um $\geq 35\%$)	Dauer des Ansprechens (Rückgang des Milzvolumens um $\geq 35\%$)	Anteil der Patienten mit einer Verbesserung um $\geq 50\%$ im Total-Symptom-Score (TSS) ²⁾	Absolute / prozentuale Veränderung (TSS)	Veränderung der Einzelsymptome des MFSAFv2.0 gegenüber Studienbeginn	Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden	ECOG Performance-Status ⁹⁾
COMFORT I	+	- ²⁾	-	+ ³⁾	+	- ⁵⁾	- ⁶⁾	+	+	+	+ ⁸⁾	+
COMFORT II	+	+	+	+ ⁴⁾	+	+	+	-	-	-	-	+

+: Daten im Dossier oder in den Studienberichten vorhanden

-: keine Daten im Dossier oder in den Studienberichten vorhanden

1) Diese beiden Endpunkte werden im Dossier der Kategorie Mortalität zugeordnet. Da es sich jedoch jeweils um kombinierte Endpunkte handelt, die sich wesentlich aus Endpunkten der Kategorie Morbidität zusammensetzen, werden sie in der vorliegenden Bewertung unter der Kategorie „Morbidität“ aufgeführt.

2) Dieser Endpunkt sollte in der COMFORT I-Studie als Langzeitendpunkt erst nach 144 Wochen erhoben werden (Daten liegen noch nicht vor) und geht daher nicht in die Bewertung ein.

- 3) Primärer Endpunkt der COMFORT I-Studie, Rückgang des Milzvolumens um $\geq 35\%$ nach 24 Wochen (MRT oder CT); ergänzend dazu wurden im Datenextraktionsbogen die Ansprechraten nach 12, 24 und 48 Wochen dargestellt.
- 4) primärer Endpunkt der COMFORT II-Studie, Rückgang des Milzvolumens um $\geq 35\%$ nach 48 Wochen (MRT oder CT), sekundär wurde dieser Endpunkt auch nach 24 Wochen erhoben und in die Ergebnisdarstellung aufgenommen.
- 5) Der Median wurde nicht separat bestimmt
- 6) In der COMFORT I-Studie nicht berechnet, da das Ansprechen auf die Therapie (bei offener Nachbehandlung mit Ruxolitinib) zum Zeitpunkt des Data-Lock bei der Mehrheit der Patienten noch anhielt und für die Berechnung eines medianen Wertes mittel Kaplan-Meier-Analyse zu wenig Ereignisse aufgetreten waren.
- 7) Gesamtscore aus 6 abgefragten Symptomen, verwendet wurde der MFSAF v2.0.
- 8) Post-hoc-Analyse
- 9) Dieser Endpunkt wurde entgegen der Einschätzung des pU der Endpunktkategorie Morbidität zugeordnet.

CT: Computertomographie, ECOG: Eastern Cooperative Oncology Group, MFSAFv2.0: Myelofibrosis Symptom Assessment Form v2.0, MRT: Magnetresonanztomographie, pU: pharmazeutischer Unternehmer, TSS: Total-Symptom-Score,

Tabelle 4: Matrix der in die Nutzenbewertung eingegangenen Endpunkte aus den Kategorien Lebensqualität und Nebenwirkungen, Datenverfügbarkeit

Studie	Gesundheits- bezogene Lebensqualität		Nebenwirkungen									
	QLQ-C30 der EORTC	FACT-Lym	UES ¹⁾	SUES	UE Grad 3-4	Abbruch aufgrund von UEs	Dosisreduktion aufgrund von UEs	Häufigste UEs, CTCAE Grad 3 und 4 ($\geq 3\%$)	Thrombozytämie (PT)	Anämie (PT)	Blutungen ^{1, 2)}	Infektionen ^{1, 3)}
COMFORT I	+	-	+	+	+	+	+	+	+	+	+	+
COMFORT II	+	+	+	+	+	+	+	+	+	+	+ ⁴⁾	+

+: Daten im Dossier oder in den Studienberichten vorhanden

-: Keine Daten im Dossier oder in den Studienberichten vorhanden

- 1) Dieser Endpunkt wurde zusätzlich zu den vom pharmazeutischen Unternehmer im Dossier präsentierten Endpunkten eingeschlossen.
- 2) Operationalisiert als Einzel-PT "Ekchymose" und als Gruppierung von "Preferred Terms containing the words bleeding, hemorrhage, or epistaxis" gemäß Studienbericht (COMFORT I).
- 3) Operationalisiert als SOC "Infections and infestations"
- 4) Hier konnten nur Daten zum Einzel-PT „Ekchymose“ aus dem Studienbericht extrahiert werden. Eine Auswertung der Blutungsereignisse lag nicht vor.

PT: Preferred Term, SOC: Systemorganklasse, SUE: schwerwiegendes unerwünschtes Ereignis, UE: unerwünschtes Ereignis

Fortsetzung von Tabelle 4.

Tabelle 5: Matrix weiterer, in den Zulassungsstudien erhobener Endpunkte ohne Eingang in die vorliegende Bewertung, Datenverfügbarkeit

Studie	Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ im Zeitverlauf ¹⁾	Notwendigkeit von Thrombozytenkonzentraten ¹⁾	Notwendigkeit von Erythrozytenkonzentraten ¹⁾	JAK2 Allel-Last ¹⁾	Leukämische Transformation ¹⁾	Anzahl der Patienten mit einer tastbaren Milzlängenreduktion um $> 50\%$ ¹⁾	PROMIS Fatigue-Scale ^{1, 3)}	Patient Global Impression of Change (PGI) ^{1, 3)}	Fibroseggrad ²⁾	Plasma-PD-Marker ⁴⁾	Phosphorylierung von STAT3 ⁵⁾	Weitere Operationalisierungen des TSS ⁶⁾	Körpergewicht	Weitere Auswertungen der Kategorie Nebenwirkungen ⁷⁾	Laborparameter ⁸⁾	Vitalzeichen (z.B. Elektrokardiogramm)
COMFORT I	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMFORT II	-	+	+	+	-	-	-	-	+	+	-	-	+	+	+	+

+: Daten im Dossier oder in den Studienberichten vorhanden

-: keine Daten im Dossier oder in den Studienberichten vorhanden

1) Ergebnisse zu diesen Endpunkten wurden ergänzend im Datenextraktionsbogen dargestellt.

2) Wird in den Studienbericht auch als Knochenmarkhistomorphologie beschrieben.

3) Diese beiden Skalen wurden in der COMFORT I-Studie zur Validierung des MFSAFv2.0 erhoben und waren im Studienprotokoll nicht explizit als Endpunkte erwähnt.

4) z.B. C-reaktives Protein, Interleukin-6, Tumor-Nekrose-Faktor-alpha (TNF- α)

5) Bei STAT3 handelt es sich um einen Transkriptionsfaktor, der durch Januskinasen phosphoryliert und somit aktiviert wird. Durch Messung des Phosphorylierungsgrades kann indirekt die Hemmung der Januskinasen analysiert werden.

6) z.B: Zeitpunkt bis zum Erreichen des Responsekriteriums (Symptomreduktion $> 50\%$)

7) z.B. Nebenwirkungen mit einem vermuteten Kausalzusammenhang zur Therapie, Nebenwirkungen nach Systemorganklasse, klinisch bedeutsame Nebenwirkungen.

8) z.B. Blutbild (Thrombozyten, Erythrozyten, Neutrophile,...), Leberenzyme, Kreatinin.

JAK: Januskinase, TSS: Total-Symptom-Score, STAT: Signal Transducer and Activators of Transcription

2.5.2. Verzerrungspotential auf Studien- und Endpunktebene

Das Verzerrungspotential der Studien wurde mit dem 'Risk of Bias Tool', sowohl auf Studienebene als auch zu einzelnen Endpunkten, getrennt eingeschätzt (vgl. IQWiG-Methodenpapier 4.0, S. 117 ff oder Cochrane Handbook for Systematic Reviews of Interventions 5.1.0, Chapter 13). Eine vollständige Übersicht in Tabellenform hinsichtlich der Darstellung des Verzerrungspotentials einzelner Endpunkte, findet sich im Extraktionsbogen der jeweiligen Studie.

Da es sich hier um adäquat durchgeführte randomisierte, kontrollierte Studien handelt, wird das Verzerrungspotenzial auf Studienebene als niedrig eingestuft. Auf Endpunktebene wurde bei der Bewertung des Verzerrungspotenzials insbesondere die fehlende bzw. schwer aufrecht zu erhaltende Verblindung und die nicht bei allen Endpunkten durchgeführte ITT-Auswertung

berücksichtigt (siehe jeweilige Datenextraktionsbögen zur Begründung der Einschätzung des Verzerrungspotenzials).

Tabelle 6: Verzerrungspotenzial auf Studien- und Endpunktebene (Kategorien Mortalität und Morbidität)

		Mor- talität	Morbidität										
Studie	Studienebene	Verlängerung der Lebensdauer	Leukämiefreies Überleben	Progressionsfreies Überleben	Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$	Prozentuale Milzgrößenveränderung	Beginn des Ansprechens (Rückgang des Milzvolumens um 35%)	Dauer des Ansprechens (Rückgang des Milzvolumens um 35%)	Anteil der Patienten mit einer Verbesserung um $\geq 50\%$ im TSS	Absolute / prozentuale Veränderung (TSS)	Veränderung der Einzelsymptome des MFSAF v2.0 gegenüber Studienbeginn	Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden	ECOG Performance-Status
COMFORT I	N	N	-	-	N	H	-	-	N	H	H	H	H
COMFORT II	N	N	N	N	N	H	H	H	-	-	-	-	H

ECOG: Eastern Cooperative Oncology Group, H: hoch, MFSAF: Myelofibrosis Symptoms Assessment Form, N: Niedrig, TSS: Total-Symptom-Score

Die Tabelle wird in Tabelle 7 fortgesetzt.

Tabelle 7: Verzerrungspotenzial auf Studien- und Endpunktebene (Kategorien Lebensqualität und Nebenwirkungen)

	Gesundheits- bezogene Lebensqualität		Nebenwirkungen									
Studie	QLQ-C30 der EORTC	FACT-Lym	UE	SUE	UE Grad 3-4	Abbruch aufgrund von UE	Dosisreduktion aufgrund von UE	Häufigste UEs, CTCAE Grade 3 und 4 (≥ 3%)	Thrombozythämie	Anämie	Blutungen	Infektionen
COMFORT I	H	-	N	N	N	N	N	N	N	N	N	N
COMFORT II	H	H	N	N	N	N	N	N	N	N	N	N

H: hoch, EORTC: European Organisation for Research and Treatment of Cancer, N: niedrig, SUE: schwerwiegendes unerwünschtes Ereignis, UE: unerwünschtes Ereignis

Fortsetzung von Tabelle 6

2.5.3. Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz

Zur Beurteilung der Patientenrelevanz der vorgelegten Endpunkte wurden das Dossier und die zugehörigen Studienberichte und die in diesen Quellen zitierte Literatur herangezogen. Des Weiteren wurden eigene Recherchen zur Validität und Patientenrelevanz der Endpunkte durchgeführt. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in den Tabellen Tabelle 9, Tabelle 10 und Tabelle 11.

Tabelle 8: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie Mortalität (basierend auf Angaben aus dem Dossier und dem Studienbericht)

Mortalität	
Endpunkt / Messinstrument	Gesamtüberleben
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Anzahl der Todesfälle bis zur Woche 24 (COMFORT I) bzw. 48 (COMFORT II), ITT-Auswertung <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU gibt an, dass die Patientenrelevanz für diesen Endpunkt prinzipiell unbestritten ist.
Einschätzung zur Qualität u. Patientenrelevanz	Der Einschätzung des pU zur Patientenrelevanz kann gefolgt werden.
Angabe und Einschätzung eines Schwellenwertes / MID	-
Endpunkt / Messinstrument	Leukämiefreie Überlebenszeit Progressionsfreies Überleben
Anmerkung	Bei den Endpunkten Leukämiefreies Überleben und Progressionsfreies Überleben handelt es sich gemäß Herstellerdossier des pU um Surrogatendpunkte für den Endpunkt Mortalität. Es wurde eine fokussierte Recherche nach möglichen Validierungsstudien durchgeführt, die eine Korrelation zwischen dem Endpunkt Leukämiefreies Überleben bzw. dem Endpunkt Progressionsfreies Überleben und dem eigentlichen patientenrelevanten Endpunkt Gesamtüberleben in der Indikation (Myelofibrose) belegen. Im Rahmen dieser Recherche konnte keine Studie im Indikationsgebiet gefunden werden, sodass die Validität des Endpunkts weiterhin unklar bleibt. Diese Endpunkte wurden in der vorliegenden Bewertung als nicht valide Surrogatparameter für den patientenrelevanten Endpunkt Gesamtmortalität eingestuft. Zur Wertung der Endpunkte Leukämiefreies Überleben und Progressionsfreies Überleben siehe auch Tabelle 10.

Tabelle 9: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie Morbidität (basierend auf Angaben aus dem Dossier und Studienbericht)

Morbidität	
Endpunkt / Messinstrument	Leukämiefreie Überlebenszeit
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> <i>"Leukemia free survival was defined as the interval between randomization and the date of the bone marrow blast count of 20% or greater OR the date of the first peripheral blast count of 20% or greater that was subsequently confirmed to have been sustained for at least 8 weeks OR the date of death from any cause, whichever occurs first."</i> <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU gibt an, dass die Patientenrelevanz für diesen Endpunkt prinzipiell unbestritten ist.
Einschätzung zur Qualität u. Patientenrelevanz	Es handelt sich um einen kombinierten Endpunkt, der sich aus Endpunkten verschiedener Endpunktkategorien (Mortalität und Morbidität) mit unterschiedlicher Relevanz und Schwere zusammensetzt. Für die Kategorie Mortalität wird dieser Endpunkt als nicht validiertes Surrogat bewertet. Der Endpunkt wurde in der vorliegenden Bewertung in der Endpunktkategorie Morbidität bewertet, ist jedoch aufgrund seiner Zusammensetzung sowie fehlender Ergebnisdarstellung zu den Einzelkomponenten in seiner Patientenrelevanz fraglich und für eine Bewertung zum Ausmaß des Zusatznutzens nicht geeignet.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	-
Endpunkt / Messinstrument	Progressionsfreies Überleben
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> <i>"Progression-free survival (PFS) was defined as the interval between randomization and the occurrence of any one of these events:</i> <i>• For a spleen volume increase, the progression date was the date of the first MRI showing a 25% or greater increase in spleen volume from the on-study nadir (including baseline)</i> <i>• For leukemic transformation, the progression date was the date of the bone marrow blast count of 20% or greater obtained on bone marrow aspirate OR the date of the first peripheral blast count of 20% or greater that was subsequently confirmed to have been sustained for at least 8 weeks.</i> <i>• For splenic irradiation, splenectomy, or death, the date of progression was the actual date of the event."</i> <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU gibt an, dass die Patientenrelevanz für diesen Endpunkt prinzipiell unbestritten ist.
Einschätzung zur Qualität u. Patientenrelevanz	Es handelt sich um einen kombinierten Endpunkt, der sich aus Endpunkten verschiedener Endpunktkategorien (Mortalität und Morbidität) mit unterschiedlicher Relevanz und Schwere zusammensetzt. Für die Kategorie Mortalität wird dieser Endpunkt als nicht validiertes Surrogat bewertet. Der Endpunkt wurde in der vorliegenden Bewertung in der Endpunktkategorie Morbidität bewertet, ist jedoch aufgrund seiner Zusammensetzung sowie

	fehlender Ergebnisdarstellung zu den Einzelkomponenten in seiner Patientenrelevanz fraglich und für eine Bewertung zum Ausmaß des Zusatznutzens nicht geeignet.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	-
Endpunkt / Messinstrument	Anteil der Patienten mit einem Rückgang des Milzvolumens um ≥ 35 %
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Im Dossier operationalisiert über die Anzahl der Patienten mit einem Rückgang $\geq 35\%$ nach 24 Wochen (COMFORT I) bzw. 48 Wochen (COMFORT II) gemessen mittels MRT oder CT. <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Die Milzgröße korreliert mit der Krankheitslast und den Allgemeinsymptomen (Cervantes, 2011). Patienten werden durch ein zusätzliches Gewicht von bis zu 10 kg belastet. Eine massiv vergrößerte Milz ist durch die Bauchdecke nach außen sichtbar und drückt gleichzeitig auf innere Organe (Magen, Darm, Lunge). Neben Schwierigkeiten bei Alltagstätigkeiten wie z.B. Sitzen, Gehen und Beugen führt eine vergrößerte Milz bei Mahlzeiten durch Druck auf den Magen zu einem vorzeitigen Sättigungsgefühl. Weiterhin kann arterieller Bluthochdruck entstehen oder verstärkt werden. Betroffene erleben dies als Beeinträchtigung des Gesundheitszustandes. Folglich ist eine Änderung der Milzgröße patientenrelevant (Dossier, Modul 4, S. 26). Der pU führt darüber hinaus zur Begründung der Patientenrelevanz die Kriterien der IWG-MRT an, nach denen eine Reduktion der tastbaren Milzlänge um mindestens 50% als klinisches Ansprechen gewertet wird (Tefferi et al., 2006). Es wird ausgeführt, dass dies (empirisch ermittelt) einem Volumenrückgang um ca. 35% entspricht (Verstovsek et al., 2010) und dieser Schwellenwert wird folglich als Wert für eine Dichotomisierung der Patienten in Responder und Non-Responder gewählt. Der pU verweist zudem auf das European Myelofibrosis Network (EUMNET), das ähnliche Kriterien verwendet: eine 50%ige Reduktion der Milzlänge bei einer Länge von > 10 cm und eine 30% Reduktion der Milzlänge bei einer Länge von < 10 cm werden hier als Schwellenwert für einen klinisch relevanten Therapieerfolg erachtet (Barosi et al., 2005).
Einschätzung zur Qualität u. Patientenrelevanz	Angegebene Quellen im Dossier und in den Studienberichten: • Cervantes, 2011 • Barosi et al., 2005 • Tefferi et al., 2006 • Verstovsek et al., 2010 Die Publikation von Cervantes (2011) stellt eine narrative Arbeit dar, in der die Therapieoptionen zur Behandlung einer Splenomegalie dargelegt werden. Es wird – ohne konkrete Daten heranzuziehen – auf einen möglichen Zusammenhang zwischen der Milzgröße und krankheitsbedingten Symptomen (z.B. Oberbauchbeschwerden, Appetitlosigkeit, Fatigue) hingewiesen. Angaben zur Validierung dieses Zusammenhangs finden sich in der vorgelegten Publikation nicht. Die Publikation von Barosi et al. (2005) stellt ein Konsensuspapier des European Myelofibrosis Network (EUMNET) dar. Ziel der Arbeit war es, geeignete Response-Kriterien abzuleiten, um klinisch relevante Therapie-

	erfolge messen zu können. In diesem Rahmen wurde festgelegt, dass eine tastbare Reduktion der Milzgröße um $\geq 50\%$ als klinisch relevant zu bewerten ist. Eine formale Validierung der konsentierten Kriterien wurde in dieser Arbeit nicht vorgenommen. Bei der Arbeit von Tefferi et al. (2006) handelt es sich um ein Konsensuspapier der International Working Group for Myelofibrosis Research and Treatment (IWG-MRT). Auch hier lag der Fokus der Arbeit auf der Konsentierung klinisch relevanter Ansprechkriterien zur Einschätzung eines Therapieerfolges und eine Reduktion der tastbaren Milzgröße um $> 50\%$ wird als patientenrelevant eingestuft. Eine Validierung der konsentierten Kriterien findet sich nicht. Über die angegebenen Publikationen hinaus wurde in der COMFORT I-Studie in einer post-hoc-Analyse untersucht, inwieweit die drei abdominellen Symptome des MFSAF (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl) mit der objektiv gemessenen Milzgröße korrelieren. Hier zeigte sich kein klarer Zusammenhang zwischen dem Rückgang des Milzvolumens und den untersuchten Oberbauchbeschwerden. Zwar wurde bei 63,3% der Patienten, die eine $\geq 35\%$ige Reduktion des Milzvolumens erreicht hatten auch ein Rückgang der Oberbauchbeschwerden um $\geq 50\%$ festgestellt; bei 36,7% der Patienten mit $\geq 35\%$iger Reduktion des Milzvolumens konnte eine entsprechende Abnahme der Symptomatik jedoch nicht festgestellt werden. Andererseits war bei ca. der Hälfte der Patienten ein Rückgang der Oberbauchbeschwerden um $\geq 50\%$ nicht mit einer Abnahme des Milzvolumens um $\geq 35\%$ assoziiert. Zusammenfassend wird der Endpunkt „Rückgang des Milzvolumens um $\geq 35\%$“ als Surrogatendpunkt für die mit einer Milzvergrößerung einhergehenden Symptome (z.B. Oberbauchbeschwerden) eingestuft. Ein eindeutiger Zusammenhang, der sich insbesondere für eine Bewertung des Ausmaßes des Zusatznutzens eignet, konnte nicht gezeigt werden. Damit bleiben der Stellenwert und die Validität dieses Endpunkts unklar. Auch die EMA hat im Rahmen einer Beratung (Scientific-Advice) dazu geraten, neben dem Milzvolumen, weitere (patientenrelevante) Endpunkte (z.B. Veränderungen von Symptomen) zu erheben.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	Der pU verwendet als Responsekriterium einen Milzvolumenrückgang um $\geq 35\%$. Er begründet dies damit, dass eine 50%ige Abnahme der tastbaren Milzlänge ein international anerkanntes Responsekriterium für die Definition eines klinischen Ansprechens ist (Barosi et al., 2005; Tefferi et al., 2006) und dies einer ca. 35%igen Abnahme des Milzvolumens entspricht (Verstovsek et al., 2010). Daten aus der COMFORT I-Studie zur Korrelation der Abnahme der tastbaren Milzlänge um $\geq 50\%$ und der messbaren Milzvolumenabnahme um $\geq 35\%$ bestätigen diesen Zusammenhang allerdings nur bedingt. Zwar war der Anteil der Patienten, die das jeweilige Kriterium erreicht hatten vergleichbar (41,9% vs. 39%), es konnte jedoch auf individueller Basis keine gute Korrelation zwischen den beiden Responsekriterien gezeigt werden (post-hoc-Analyse, siehe Datenextraktionsbogen S. 83). Es ist zudem darauf hinzuweisen, dass es sich bei der 50%igen Reduktion der tastbaren Milzlänge um ein konsentiertes Kriterium zur Beurteilung eines Therapieerfolges handelt und für diesen Schwellenwert keine klinischen Daten zum Nachweis der Relevanz vorgelegt wurden. Die Bedeutung des gewählten Schwellenwertes kann anhand der vorliegenden Daten nicht abschließend geklärt werden.
Endpunkt / Messinstrument	MFSAF v2.0 (Myelofibrosis Symptoms Assessment Form); in der Studie wird hauptsächlich der Total-Symptom-Score ausgewertet (Summenscore aus 6 Items des MFSAF v2.0, ohne „Inaktivität“)¹

Operationalisierung und Validierung: Begründung seitens des pharmazeutischen Unternehmers	<u>Operationalisierung:</u> Der hier verwendete modifizierte Fragebogen MFSAF v2.0 (in Form eines elektronischen Patiententagebuchs) besteht aus 7 Items, die die für die Myelofibrose charakteristischen Symptome abfragen: • Oberbauchbeschwerden • Schmerzen unter den Rippen • Völlegefühl • Nachtschweiß • Juckreiz • Muskel- und Knochenschmerzen • Inaktivität Patienten bewerteten jeden Abend diese Symptome und die daraus resultierende Einschränkung der Aktivität für den jeweiligen Tag mit 0 (nicht vorhanden) bis 10 (schlimmstmöglich). Während der Nacht wurden die Daten automatisch übertragen und der Gesamtscore (TSS, bestehend aus 6 Symptomen, ohne das Item Inaktivität) wurde zentral und automatisiert berechnet, ohne dass der Patient oder der Prüfarzt das Ergebnis erhielten. Das Ergebnis der Einzelsymptome wird zum Gesamtscore „Total-Symptom-Score“ summiert (Range: 0-60). <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Symptome sind die vom Patienten wahrgenommenen Krankheitszeichen. Sie bedeuten für ihn sowohl eine Verschlechterung des Gesundheitszustandes als auch eine Beeinträchtigung der Lebensqualität. Deswegen ist Symptomatik ein patientenrelevanter Endpunkt. Zur exakten Erfassung der Symptomatik wurde in der COMFORT I-Studie der speziell auf die Symptome der Myelofibrose zugeschnittene Fragebogen „Myelofibrosis Symptoms Assessment Form“ (MFSAF v2.0) eingesetzt, der die patientenrelevanten und krankheitscharakteristischen Symptome erfasst und quantifiziert (Dossier, Modul 4, S. 26). Es werden vom pU mehrere Publikationen angeführt, die die Patientenrelevanz des MFSAF untermauern sollen (Mesa et al., 2011; Mesa et al., 2009; Mesa et al., 2011). Die vorgelegte Arbeit von Mesa et al. (2009) beschreibt die Entwicklung und erstmalige Validierung des Messinstrumentes zur Erfassung der für die Myelofibrose-typischen Symptome. Basierend auf einer Befragung von Myelofibrose-Patienten zu besonders belastenden Krankheitssymptomen wurde ein 20 Items umfassender Fragebogen entwickelt. Ein Vergleich mit der etablierten Skala „Memorial Symptom Assessment Scale (MSAS)“ zeigte einen klaren Zusammenhang zwischen korrespondierenden Symptomen (z.B. „Fatigue“ und „Lack of Energy“, „Sweats“ and „Night Sweats“). Es handelt sich allerdings dabei nicht um die in der COMFORT I-Studie eingesetzten MFSAF v2.0 (ein aus 7 abgefragten Symptomen bestehendes Messinstrument), sondern um die ursprüngliche Form des MFSAF, der insgesamt 20 Items enthält (inkl. des Brief-Fatigue-Inventory als separates Erhebungsinstrument). Des Weiteren führt der pU eine Publikation zur Validierung des MFSAF an, die auf einer offenen Phase II-Studie für den Wirkstoff Ruxolitinib basiert (Mesa et al., 2011). Hier wurde eine modifizierte Version des MFSAF verwendet, die 15 krankheitsspezifische Items abfragt und damit nicht der in der COMFORT I-Studien verwendeten Version (mit 7 Items) entspricht. In dieser Studie wurde der Zusammenhang zwischen den Items des MFSAF v2.0 und Myelofibrose-spezifischen objektiven Parametern untersucht (Splenomegalie, Anämie, Body-Mass-Index, 6-Minute-Walk), die potenziell mit den belastenden Symptomen in Verbindung stehen. Zwischen
--	---

	einzelnen Parametern konnte zwar eine Korrelation nachgewiesen werden (z.B. Milzgröße und Fatigue), in weiten Teilen konnte jedoch kein klarer und signifikanter Zusammenhang zwischen einzelnen Items des MFSAF v2.0 und objektiven Parametern nachgewiesen werden, sodass die Validität des verwendeten Fragebogens nicht abschließend beurteilt werden kann. Es wäre an einem größeren Patientenkollektiv zu klären, ob die fehlende Signifikanz durch die zu geringe Patientenzahl erklärt werden kann. Der MFSAF v2.0 sollte zudem in der COMFORT I-Studie validiert werden (u.a. durch Analyse der Korrelation mit den Skalen EORTC QLQ-C30, PROMIS Fatigue-Scale, Patient Global Impression of Change [PGIC] und Brief-Pain-Inventory). Die Daten werden deskriptiv im Studienbericht erwähnt und auf dem „53rd Annual Meeting of the American Society of Hematology“ (2011, San Diego) als Poster präsentiert (Mesa et al., 2011). Eine Korrelation mit dem PGIC zeigt, dass 91,2% der Ruxolitinib-Patienten, die eine 50%ige Verbesserung des Total-Symptom-Scores erreichten, ihren Gesundheitszustand durch die Therapie als „viel“ oder „sehr viel“ verbessert bewerteten. Demgegenüber hatte sich der Gesundheitszustand bei 73,6% der Non-Responder in der Placebo-Gruppe nicht verändert oder sogar verschlechtert. Des Weiteren wurde die Korrelation der MFSAF v2.0 mit der EORTC QLQ-C30 untersucht: Responder ($\geq 50\%$ige Abnahme des TSS-Scores) unter Ruxolitinib zeigten statistisch signifikant größere Verbesserungen der Lebensqualität (inkl. der einzelnen Subskalen) als Patienten unter Placebo-Behandlung, bei denen sogar eine Verschlechterung der Lebensqualität festgestellt wurde. Auch ließ sich ein statistisch signifikanter Zusammenhang zwischen einer Symptomlinderung um $\geq 50\%$ im TSS und der PROMIS Fatigue-Scale nachweisen. Dieses Messinstrument misst das für Myelofibrose-Patienten besonders belastende Symptom Fatigue (Müdigkeit).
Einschätzung zur Qualität u. Patientenrelevanz	Vom pharmazeutischen Unternehmer angegebene Quellen: - Mesa et al., 2011 - Mesa et al., 2009 - Mesa et al., 2011 In den beiden vom pU angeführten Quellen wurde die Entwicklung und Validierung des MFSAF beschrieben (Mesa et al., 2011; Mesa et al., 2009). Es wurde allerdings in der COMFORT I-Studie eine modifizierte Version des Messinstrumentes (bzw. ein daraus abgeleiteter Summen-score) herangezogen, sodass die vom pU herangezogenen Quellen nur bedingt zur Validierung des Messinstrumentes herangezogen werden können. Die im Rahmen der COMFORT I-Studie erhobenen Daten zur Validierung des MFSAF v2.0 bzw. des daraus abgeleiteten Total-Symptom-Score zeigen allerdings einen konsistenten Zusammenhang zwischen einer Symptomlinderung um $\geq 50\%$ im Total-Symptom-Score und etablierten Messinstrumenten zur Erfassung der Lebensqualität (EORTC QLQ-C30, Patient Global Impression of Change, PROMIS Fatigue-Scale). Diese Daten können als Hinweis darauf gedeutet werden, dass ein Ansprechen im Total-Symptom-Score ($\geq 50\%$ige Abnahme der Symptome) bei den meisten Patienten mit einer spürbaren Verbesserung des Gesundheitszustandes bzw. der Lebensqualität einhergeht. Zur ausführlichen Validierung des Messinstrumentes wird im Studienbericht auf ein separates Evidenz-dossier verwiesen, das allerdings nicht als Quelle zur Verfügung stand, sodass eine vollständige methodische Bewertung der Daten nicht möglich ist. Auffällig ist zudem, dass zu einer a priori geplanten Analyse der Korrelation der MFSAF mit der Brief-Pain-Inventory keine Daten im Studienbericht existieren.

	Auch die EMA (2012) kritisiert im EPAR (S. 52), dass der pU den Total-Symptom-Score zwar als valides Messinstrument darstellt, der Validierungsprozess jedoch nicht klar nachvollziehbar ist. Nichtsdestotrotz betrachtet die EMA jegliche Symptomlinderung als klinisch relevant (S. 82) und befindet die Ergebnisse der COMFORT I-Studie hinsichtlich dieses Endpunkts als bedeutsam. Insgesamt kann festgehalten werden, dass die mit diesem Mess-instrument erfassten Symptome in dieser Indikation von Bedeutung sind, wobei hier vor allem Fatigue, Juckreiz, Nachtschweiß und Knochenschmerzen von betroffenen Patienten mit PMF, PPV-MF und PET-MF als besonders belastend empfunden werden (Mesa et al., 2007) und eine Erfassung der Veränderung der belastenden Symptomatik als sinnvoll erachtet wird.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	Der pU wählt in der COMFORT I-Studie eine Verbesserung um $\geq 50\%$ im Total-Symptom-Score als Grenze, um einen klinisch relevanten Unterschied zu definieren und Patienten in Responder oder Non-Responder einzuteilen. Daten aus der COMFORT I-Studie zeigen, dass eine 50%ige Verbesserung des Total-Symptom-Scores mit einer Verbesserung im Patient Global Impression of Change, der Lebensqualität (gemessen mit Hilfe der EORTC QLQ-C30) und der PROMIS Fatigue-Scale einhergeht. Die Daten können als Hinweis darauf gedeutet werden, dass eine 50%ige Verbesserung des Total-Symptom-Score bei den meisten Patienten mit einer spürbaren Verbesserung des Gesundheitszustandes einhergeht. Die Methodik der Validierung ist allerdings nicht klar nachvollziehbar, sodass die klinische Relevanz des Schwellenwertes nicht abschließend beurteilt werden kann.
Endpunkt / Messinstrument	ECOG Performance-Status
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Veränderung des ECOG Performance-Status gegenüber Studienbeginn. Bedeutung der einzelnen Stufen: 0: Volle Aktivität, normales Leben möglich; 1: Eingeschränkte Aktivität, leichte Arbeit möglich; 2: Selbstversorgung möglich, nicht arbeitsfähig, nicht bettlägerig, muss < 50% der Tageszeit ruhen; 3: Selbstversorgung möglich, nicht arbeitsfähig, nicht bettlägerig, muss > 50% der Tageszeit ruhen; 4: Bettlägerig und völlig pflegebedürftig; 5: Tod <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Die Lebensqualität stellt die Auswirkungen der Erkrankung bzw. deren Linderung auf den Alltag des Patienten und sein Empfinden dar. Somit ist sie unmittelbar patientenrelevant (Modul 4, S. 26). Der pU gibt an, dass Änderungen im ECOG Performance-Status drastische Veränderungen des Gesundheitszustandes und damit auch der Lebensqualität anzeigen und dient als Maß für Einschränkungen in Beruf und Alltag. Der pU führt aus, dass es sich beim ECOG Performance-Status um einen der am häufigsten verwendeten Parameter zur Charakterisierung des Gesundheitszustandes handelt.
Einschätzung zur Qualität u. Patientenrelevanz	Vom pU angegebene Quellen: Der pU legt in seinem Dossier und auch im Studienbericht keine Quellen zur Patientenrelevanz des verwendeten Messinstrumentes vor. Es handelt sich beim ECOG Performance-Status um ein etabliertes, generisches Messinstrument zur Erfassung des funktionellen Status von Krebspatienten (Oken et al., 1982). Er ermöglicht eine grobe Einteilung in 6 Stadien und dient in der Praxis dazu zu untersuchen, wie gut der Patient im Alltag zurechtkommt und er wird häufig als Prognoseinstrument verwendet.

	Da dieses Messinstrument daher nicht geeignet ist, um detailliert die Lebensqualität der Patienten zu erheben und therapiebedingte Änderungen der Lebensqualität zu erfassen, wird dieser Endpunkt in der vorliegenden Bewertung in die Endpunktkategorie Morbidität eingestuft. Eine ergänzende Betrachtung des Performance-Status und dessen Veränderung im Verlauf der Studie wird trotz der beschriebenen Limitationen aber durchaus als sinnvoll erachtet.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	-

#: Lt. dem 5. Kapitel § 12 Absatz 1 der Verfo des G-BA, gilt der medizinische Zusatznutzen nach § 35a Absatz 1 Satz 3 Nummer 2 und 3 SGB V durch die Zulassung belegt. Nachweise gemäß § 5 Absatz 1 bis 6 müssen nicht vorgelegt werden.

1: Die Festlegung erfolgte in Absprache mit der FDA.

ECOG: Eastern Cooperative Oncology Group, EMA: European Medicines Agency, IWG-MRT: International working Group for Myelofibrosis Research and Treatment, MFSAF: Myelofibrosis Symptom Assessment Form

Tabelle 10: Beschreibung der Endpunkte und Erhebungsinstrumente der Kategorie gesundheitsbezogene Lebensqualität (basierend auf Angaben aus dem Dossier und Studienbericht)

Gesundheitsbezogene Lebensqualität	
Endpunkt / Messinstrument	QLQ-C30 der EORTC (generische Skala zur Erfassung der Lebensqualität, Range je Unterskala 0-100)
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Generischer Fragebogen zur Erfassung der Lebensqualität von Krebspatienten, bestehend aus • 5 funktionellen Skalen (körperliche Funktionen, Rollenfunktion, kognitive Funktionen, emotionale Funktionen und soziale Funktionen), • 3 Symptomskalen (Fatigue, Schmerz, Übelkeit / Erbrechen), • 2 Globalbeurteilungen (Lebensqualität und allgemeiner Gesundheitszustand) sowie • einige Einzel-Items für Symptome, die von onkologischen Patienten häufig berichtet werden (Dyspnoe, Schlaflosigkeit, Verstopfung, Diarrhoe, finanzielle Belastung) Bei den funktionellen Skalen und den Globalurteilen bedeutet eine höhere Zahl eine bessere Lebensqualität. Bei den Symptom(skalen) ist eine höhere Zahl mit mehr Symptomen / Einschränkungen gleichzusetzen. <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Die Lebensqualität stellt die Auswirkungen der Erkrankung bzw. deren Linderung auf den Alltag des Patienten und sein Empfinden dar. Somit ist sie unmittelbar patientenrelevant (Modul 4, S. 26). Der pU gibt an, dass der Lebensqualitätsfragebogen QLQ-C30 der EORTC als Standardinstrument gilt und vor allem in onkologischen Therapiestudien häufig eingesetzt wird (Aronson et al., 1993). In der angeführten Publikation wird die Entwicklung und Validierung des Messinstrumentes beschrieben. Eine weitere angegebene Quelle beschreibt die (methodisch adäquate) Ableitung einer klinisch relevanten Differenz für den Fragebogen (Osoba et al., 1998). Hier wurde eine ankerbasierte Methode gewählt und die Verbesserung bzw. Verschlechterung der Skala mit einer 7 Kategorien umfassenden globalen Skala (very much / moderately / a little worse, no change, very much / moderately / a little better) in Beziehung gebracht. 300 Brustkrebs-

	patientinnen wurden in die Analyse eingeschlossen. Eine geringe Veränderung ging mit einer Veränderung der Skala von 5-10 Punkten einher. Eine moderate Veränderung korrelierte mit einer Veränderung um 10-20 Punkte und für eine vom Patienten als sehr groß empfundene Veränderung war eine Veränderung von mehr als 20 Punkten auf der Skala notwendig.
Einschätzung zur Qualität u. Patientenrelevanz	Vom pU angegebene Quellen: • Aaronson et al., 1993 • Osoba et al., 1998 Prinzipiell handelt es sich um ein etabliertes und validiertes Messinstrument, das die Lebensqualität adäquat erfasst und klinisch relevante Änderungen darzustellen vermag (Aaronson et al., 1993; Osoba et al., 1998). Kritisch ist allerdings anzumerken, dass der Fragebogen nicht speziell für Patienten mit Myelofibrose validiert ist und folglich krankheits-spezifische Auswirkungen auf die Lebensqualität nur schwer erfassen kann. Dies wird auch seitens der EMA im EPAR angemerkt (EMA, 2012).
Angabe und Einschätzung eines Schwellenwertes / MID[#]	Der pU gibt an, dass für diesen Fragebogen zur Erfassung der Lebensqualität eine Gruppendifferenz von mindestens 10 Punkten notwendig ist, um eine klinisch relevante Änderung festzustellen (Dossier, Modul 4, S. 72). Dieser Schwellenwert ergibt sich aus einer methodisch adäquaten (s.o.) Arbeit von Osoba et al. (1998). In der vorgelegten Arbeit wurde eine Veränderung der Lebensqualität um > 10 Punkte in den Subskalen des QLQ-C30 von den betroffenen Krebspatienten als spürbare Verbesserung der Lebensqualität eingestuft. Es fand also eine Korrelation auf patienten-individueller Ebene statt. In der COMFORT II-Studie wurde jedoch ein <i>Gruppenunterschied</i> von > 10 Punkten im Studienverlauf als klinisch relevant betrachtet, was nicht mit den Daten von Osoba et al. in Einklang zu bringen ist. Dies ist insbesondere zu kritisieren, da es bereits zu Studienbeginn Unterschiede in den Werten der Subskalen gab. Das vom pU beanspruchte Relevanzkriterium kann folglich in der angewendeten Form nicht als validiert angesehen werden. Es sollte zudem bei der Interpretation der Relevanzschwelle beachtet werden, dass der Zusammenhang nicht explizit an Myelofibrose-Patienten gezeigt worden ist.
Endpunkt / Messinstrument	FACT-Lym
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Patienten werden aufgefordert, die Lebensqualität anhand von 2 Fragebögen (FACT-G [generisch] und LymS [krankheitsspezifisch]) innerhalb der letzten 7 Tage zu beurteilen. Es bestehen mehrere Möglichkeiten der Auswertung dieser Skalen: • FACT-G Total-Score (Range 0-108): ein generischer Fragebogen mit 27 Items (4 Bereiche werden abgefragt: Physical Well-Being (WB), Social/Family WB, Emotional WB, and Functional WB) • FACT-LymS Score (Range 0-60): krankheitsspezifische Unterskala (15 Items) • FACT-Lym Total-Score (Range 0-168): FACT-G Total-Score + LymS • FACT-TOI Score (Range 0-116): Physical WB + Functional WB + LymS Zur Operationalisierung wurde jeweils eine Veränderung der Summenscores gegenüber Studienbeginn nach 8, 16, 24 und 48 Wochen dargestellt. Höhere Werte bedeuten eine höhere Lebensqualität. <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u>

	Die Lebensqualität stellt die Auswirkungen der Erkrankung bzw. deren Linderung auf den Alltag des Patienten und sein Empfinden dar. Somit ist sie unmittelbar patientenrelevant (Modul 4, S. 26). Der pU gibt an, dass er mit diesem Messinstrument spezifischere Einflüsse der Myelofibrose-Symptomatik auf die Lebensqualität erfassen will, da dieser speziell auf die Symptomatik hämatologischer Neoplasien ausgerichtet ist. Literaturangaben zur Validierung des eingesetzten Messinstrumentes werden vom pU im Dossier nicht vorgelegt.
Einschätzung zur Qualität u. Patientenrelevanz	Vom pU angegebene Quellen: Quellen zur Patientenrelevanz des verwendeten Messinstrumentes lassen sich dem Dossier nicht entnehmen. Prinzipiell ist die krankheitsspezifische Erfassung der Lebensqualität zu begrüßen. Angaben zur Entwicklung und Validierung des FACT-Lym-Fragebogens können als Abstract dem Kongressband des jährlichen Treffens der „American Society of Hematology“ (2005) entnommen werden (Cella et al., 2005). Aus dieser Publikation ergibt sich, dass der Fragebogen speziell für Patienten mit Non-Hodgkin-Lymphom entwickelt wurde und es ist somit fraglich, ob er auch geeignet ist die Lebensqualität von Patienten mit Myelofibrose spezifisch zu erfassen. Bei dem FACT-G Total Score handelt es sich um ein etabliertes und validiertes generisches Messinstrument zur Erfassung der Lebensqualität von Krebspatienten (Cella et al., 1993). Insgesamt ist die Validität des verwendeten Summenscores nicht eindeutig nachvollziehbar. Die EMA beschreibt dieses Messinstrument im EPAR (EMA, 2012) als Selbsterhebungsinstrument für Krebspatienten mit funktionellen Subskalen, macht allerdings keine Angaben zur Qualität und Validität des Erhebungsinstrumentes.
Angabe und Einschätzung eines Schwellenwertes / MID[#]	Der pU präsentiert in seinem Studienbericht zu der COMFORT II-Studie klinisch relevante Schwellenwerte (MIDs) zur Beurteilung eines Therapieerfolgs. Er verweist auf Publikationen von Carter et al. (2008) und Webster et al. (2003), aus denen die jeweiligen Schwellenwerte hervorgehen: • FACT-Lym Total Score: 6,2-11,2 • FACT-TOI: 5,5-11 • FACT-G: 3-7 • FACT-LymS: 2,9-5,4 Bei der Publikation von Carter et al., aus der die MIDs zum FACT-LymS, dem FACT-TOI und dem FACT-Lym Total Score stammen, handelt es sich lediglich um einen Kongressbeitrag, in dem der statistisch signifikante Zusammenhang zwischen einer Veränderung im ECOG Performance-Status (als Anker) und der jeweiligen Skala bei Patienten mit einem Non-Hodgkin-Lymphom beschrieben wird. Des Weiteren wird ein verteilungsbasierter Ansatz beschrieben (1/3 bzw. 1/2 Standardabweichung). Eine umfassende Beschreibung der gewählten Methodik findet sich nicht, was eine abschließende Beurteilung der postulierten MIDs schwierig macht. Die MID zum FACT-G, die aus der Publikation von Webster et al. (2003) hervorgeht kann als etabliert akzeptiert werden. Die beschriebenen MIDs wurden allerdings nicht in der relevanten Indikation (Myelofibrose) bestimmt und diese Unsicherheit sollte bei der Interpretation der Daten berücksichtigt werden.

[#]: Lt. dem 5. Kapitel § 12 Absatz 1 der VerfO des G-BA, gilt der medizinische Zusatznutzen nach § 35a Absatz 1 Satz 3 Nummer 2 und 3 SGB V durch die Zulassung belegt. Nachweise gemäß § 5 Absatz 1 bis 6 müssen nicht vorgelegt werden

2.6. Ergebnisse zum Zusatznutzen

In die vorliegende Bewertung gingen die folgenden Endpunkte ein (vgl. Matrix der Endpunkte in Tabelle 4 und Tabelle 5). Die Ergebnisse richten sich (soweit nicht anders angegeben) dabei nach dem Zeitpunkt der primären Endpunkterhebung (a priori definiert nach 24 Wochen [COMFORT I] bzw. 48 Wochen [COMFORT II]) und stammen aus dem Herstellerdossier (Modul 4, Abschnitt 4.4.4) und aus den Studienberichten. Die Ergebnisse basieren, soweit nicht anders angegeben, auf der ITT-Population (in den Studienberichten auch als „Full Analysis Set“ [FAS] bezeichnet). Für die Analyse der Nebenwirkungen wurden diejenigen Patienten eingeschlossen, die mindestens eine Dosis der Interventions- bzw. Kontrollmedikation bekommen haben (im Studienbericht als „Safety Set [SS]“ bezeichnet).

Die Ergebnisse werden gemäß den Kategorien Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen dargestellt. Ergebnisse zu weiteren Endpunkten, die nicht direkt als patientenrelevant betrachtet wurden, werden ergänzend in den angehängten Datenextraktionsbögen zu den Studien COMFORT I und COMFORT II dargestellt.

Tabelle 11: Ergebnisse zur Mortalität (Ruxolitinib vs. Placebo bzw. BAT)

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
	Gesamt N	Ergebnisse N (%)	Gesamt N	Ergebnisse N (%)	HR (95 %-KI)	p-Wert
Gesamtüberleben (Anzahl der Todesfälle [jegliche Ursache] nach 24 [COMFORT I] bzw. 48 [COMFORT II] Wochen)						
COMFORT I ^{1, 2)}	155	10 (6,5) ³⁾	154	14 (9,1)	0,668 (0,30-1,50)	0,3268
COMFORT II ^{4, 5)}	146	6 (4,1)	73	4 (5,5)	0,70 (0,20-2,49)	0,58

- 1) Von den insgesamt 24 Todesfällen in der COMFORT I-Studie, die bis zum Zeitpunkt der Erhebung des primären Endpunkts auftraten, starben 20 Patienten während oder bis zu 28 Tage nach der letzten Dosis, 4 Patienten starben später als 28 Tage nach der letzten Dosis.
- 2) Angabe aus dem Dossier: Vier Monate nach der Erhebung des primären Endpunkts wurde im Rahmen einer weiteren geplanten Sicherheitsanalyse nach einer medianen Beobachtungszeit von 51 Wochen ein deutlicher Überlebensvorteil für Ruxolitinib festgestellt (8,4%) in der Ruxolitinib-Gruppe und 24 Todesfällen (15,6%) in der Placebo-Gruppe betrug die Hazard Ratio 0,5 (0,25-0,98), p = 0,04.
- 3) Ein Patient starb nach Crossover zu Ruxolitinib.
- 4) Von den insgesamt 10 Todesfällen zum Zeitpunkt der Erhebung des primären Endpunkts traten 7 (4 bzw. 3) unter Therapie bzw. bis zu 28 Tage nach der letzten Dosis auf und 3 Patienten starben später als 28 Tage nach der letzten Dosis.
- 5) 2 Monate nach der primären Analyse sind in der Ruxolitinib-Gruppe 5 weitere Patienten verstorben (Ruxolitinib: 11 Todesfälle, Placebo: 4 Todesfälle, HR: 1,01, 95% KI: 0,21; 3,24) (Harrison et al., 2012).

BAT: Best-Available-Therapy, HR: Hazard Ratio, KI: Konfidenzintervall

Für den Endpunkt Gesamtüberleben zeigte sich in den beiden Zulassungsstudien kein signifikanter Unterschied zwischen den beiden Behandlungsgruppen. Die Studie war allerdings auch nicht dafür geeignet, Unterschiede bezüglich dieses Endpunkts aufzudecken.

Ergänzend wurde in der COMFORT I-Studie vier Monate nach der Erhebung des primären Endpunkts im Rahmen einer weiteren Sicherheitsanalyse nach einer medianen Beobachtungszeit von 51 Wochen eine zusätzliche Auswertung des Endpunkts Gesamtüberleben durch-

geführt. Hier zeigte sich ein Überlebensvorteil von Ruxolitinib: Bei 13 Todesfällen (8,4%) in der Ruxolitinib-Gruppe und 24 Todesfällen (15,6%) in der Placebo-Gruppe betrug die Hazard Ratio 0,50 (0,25-0,98); $p = 0,04$ (ITT-Auswertung).

Tabelle 12: Ergebnisse zur Morbidität (Ruxolitinib vs. Placebo bzw. BAT)

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
	Gesamt N	Ergebnisse N (%)	Gesamt N	Ergebnisse N (%)	Effektmaß (95 %-KI)	p-Wert
Leukämiefreies Überleben (interval between randomization and the date of the bone marrow blast count of 20% or greater OR the date of the first peripheral blast count of 20% or greater that was subsequently confirmed to have been sustained for at least 8 weeks OR the date of death from any cause, whichever occurs first)						
COMFORT I ¹⁾	-	-	-	-	-	-
COMFORT II ²⁾	146	6 (4,1)	73	4 (5,5)	0,65 (0,18-2,31)	0,51 ³⁾
1) Das Leukämiefreie Überleben war im Studienprotokoll der COMFORT I-Studie als Langzeitendpunkt geplant, der erst nach Woche 144 ausgewertet werden sollte und für den damit zum Zeitpunkt der Dossiereinreichung noch keine Daten vorlagen. 2) Alle in die Auswertung dieses Endpunkts eingeflossenen Ereignisse waren Todesfälle. 3) Log-Rank-p-Value						
Progressionsfreies Überleben (Progression was defined as increase of spleen volume of at least 25% compared with on-study nadir, splenectomy, splenic irradiation, leukemic transformation or death)						
COMFORT I	-	-	-	-	-	-
COMFORT II	146	44 (30,1) ¹⁾	73	19 (26,0) ¹⁾	0,81 (0,47-1,39)	0,46 ²⁾
1) Es war den vorliegenden Unterlagen nicht zu entnehmen, auf welches Einzelereignis der Progress jeweils zurückzuführen ist. 2) Log-Rank-p-Value						
Milzvolumenreduktion (Anzahl der Patienten mit einem Milzvolumenrückgang $\geq 35\%$ nach 24 bzw. 48 Wochen, gemessen mittels MRT oder CT), ITT						
COMFORT I 24 Wochen ^{1, 2)}	155	65 (41,9)	153	1 (0,7)	OR ²⁾ : 134,4 (17,97-1005)	< 0,001
COMFORT II ⁴⁾ 24 Wochen 48 Wochen ¹⁾	144 144	46 (31,9) 41 (28,5)	72 72	0 (0) 0 (0)	k. A. k. A.	<0,001 ³⁾ <0,001 ³⁾
1) Primärer Endpunkt in der Studie. 2) Die Odds Ratio und das 95% Konfidenzintervall wurden mittels logistischer Regression berechnet. 3) der p-Wert wurde mit dem exakten CMH-Test berechnet. 4) Es wurden nur Patienten eingeschlossen, für die eine Milzgrößenmessung zu Studienbeginn vorlag. BAT: Best-Available-Therapy, CT: Computertomographie, ITT: Intention to treat, KI: Konfidenzintervall, MRT: Magnetresonanztomographie, OR: Odds Ratio						
Milzvolumenreduktion (Prozentuale Veränderung des Milzvolumens gegenüber Studienbeginn nach 24 bzw. 48 Wochen), ohne Crossover-Patienten						
	Gesamt N	Veränderung (%), Median [min; max]	Gesamt N	Veränderung (%), Median [min; max]	Effektmaß (95 %-KI)	p-Wert
COMFORT I 24 Wochen	139	-33 [-75,9; 25,1]	106	8,5 [-46,4; 48,8]	k. A.	-

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
COMFORT II						
24 Wochen	125	-27,5 [-71,2; 36,9]	45	5,1 [-33,3; 29,7]	k. A.	p < 0.0001
48 Wochen	98	-28,3 [-78,6; 14,7]	34	8,5 [-20,0; 32,9]	k. A.	p < 0.0001
k.A.: keine Angaben, KI: Konfidenzintervall, SD: Standardabweichung						
Milzvolumenreduktion (Beginn des Ansprechens [definiert, als anhaltende Milzvolumenreduktion ≥ 35%]), ohne Crossover-Patienten						
	N	Wochen, Median (95% KI)	N	Wochen, Median (95% KI)	Effektmaß	p-Wert
COMFORT I	In der COMFORT I-Studie wurde ein Median nicht separat bestimmt. Es wird allerdings angegeben, dass in dieser Studie nach 12 Wochen bereits 61 von 155 Patienten (39,4%) eine Reduktion des Milzvolumens um mindestens 35% erreicht hatten (Dossier, Modul 4, S. 63).					
COMFORT II	69	12,29 (12,14-14,43)	1	15,43 (n.v. – n.v.)	k. A.	k. A.
k.A.: keine Angaben, n.v. nicht verfügbar, KI: Konfidenzintervall						
Milzvolumenreduktion (Dauer des Ansprechens; definiert, als anhaltende Milzvolumenreduktion [≥ 35%] konsekutiver Messungen bis zum Data Lock) ¹⁾						
COMFORT I	Die Dauer des Ansprechens konnte in der Studie nicht berechnet werden, da das Ansprechen auf die Therapie zum Zeitpunkt des Data Lock bei der Mehrheit der Patienten noch anhielt und für die Berechnung eines medianen Wertes mittels Kaplan-Meier-Analyse zu wenig Ereignisse aufgetreten waren (Dossier, Modul 4, S. 63). Ergänzung aus dem Studienbericht: Zum Zeitpunkt des Data Lock hatten 52 Responder (73,2%) weiterhin eine Milzvolumenreduktion ≥ 35% gegenüber Studienbeginn.					
COMFORT II	In der COMFORT II-Studie wurde die Dauer des Ansprechens auf Ruxolitinib (N=69) mit ca. 48 Wochen berechnet. Eine Ansprechdauer im Kontrollarm (N=1) konnte nicht berechnet werden, da unter BAT nur ein Teilnehmer mindestens einmal über dem Schwellenwert von 35% Milzvolumenrückgang blieb (Dossier, Modul 4, S. 63).					
1) Die Dauer des Ansprechens wurde nur für diejenigen Patienten bestimmt, die bei mindestens einer Messung eine ≥ 35%ige Reduktion der Milzgröße aufwiesen.						
Anteil der Patienten mit einer Verbesserung um ≥ 50% im Total-Symptom-Score nach 24 Wochen ¹⁾ (LOCF)						
	N	Anzahl Responder (%)	N	Anzahl Responder (%)	Odds Ratio (KI)	p-Wert ⁴⁾
COMFORT I	148 ²⁾	68 (45,9)	152 ²⁾	8 (5,3)	15,28 (6,93-33,66)	p < 0,0001
COMFORT II ³⁾	-	-	-	-	-	-
1) Das Einzelsymptom Inaktivität ist nicht Teil des Total-Symptom-Scores (TSS), Begründung des pU: Das siebte MFSAF-Item „Inaktivität“ wurde auf Vorschlag der FDA hin nicht als Symptom gewertet, da Inaktivität eine Folge der eigentlichen Symptome ist (Angabe aus dem Dossier, Modul 4, S. 27). 2) Die fehlenden Patienten waren nicht evaluierbar aufgrund fehlender Werte zu Studienbeginn, bzw. einem Score von 0 zu Studienbeginn und zum Zeitpunkt der Auswertung. 3) Dieser Endpunkt wurde in der Studie nicht erhoben. 4) Chi-Quadrat-Test						
KI: Konfidenzintervall, LOCF: Last Observation Carried Forward						

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
Absolute Veränderung im Total-Symptom-Score gegenüber Studienbeginn (nach 24 Wochen), Range 0-60, ohne Crossover-Patienten						
	N	Studienbeginn (Median); Differenz gegenüber Studien- beginn (Median, [min, max])	N	Studienbeginn (Median); Differenz gegenüber Studien- beginn, (Median, [min, max])	Effektmaß	p-Wert ²⁾
COMFORT I	131	17,8; -6,9 (-35,8, 22,1)	105	14,8; 2,0 (-27,1 , 30,3)	k. A.	p < 0,0001*
COMFORT II ¹⁾	-	-	-	-	-	-
1) Dieser Endpunkt wurde in der Studie nicht erhoben. 2) Wilcoxon-Rangsummentest						
Prozentuale Veränderung im Total-Symptom-Score gegenüber Studienbeginn (nach 24 Wochen), Range: 0-60, ohne Crossover-Patienten						
	N	Prozentuale Verän- derung gegenüber Studienbeginn Median (min, max)	N	Prozentuale Verän- derung gegenüber Studienbeginn Median (min, max)	Effektmaß	p-Wert ²⁾
COMFORT I	129	-56,2 (-100, 108,3)	103	14,6 (-100, 511,6)	k. A.	p < 0,0001
COMFORT II ¹⁾	-	-	-	-	-	-
1) Dieser Endpunkt wurde in der Studie nicht erhoben. 2) Wilcoxon-Rangsummentest						
Veränderung der Einzelsymptome des MFSAF v2.0 gegenüber Studienbeginn nach 24 Wochen (Range der Einzelsymptome: 0 = nicht vorhanden bis 10 = schlimmstmöglich)						
	N	Studienbeginn ¹⁾ (Mittelwert, SD), abs. Veränderung gegenüber Studien- beginn (Mittelwert, SD) ²⁾	N	Studienbeginn ¹⁾ (Mittelwert, SD), abs. Veränderung gegenüber Studien- beginn (Mittelwert, SD) ²⁾	Effektmaß	p-Wert ³⁾
COMFORT I (Range der Einzelsymptome: 0 = nicht vorhanden bis 10 = schlimmstmöglich)						
Oberbauch- beschwerden	131	3,7 (2,2), -1,8 (2,2)	105	3,4 (2,4), 0,6 (2,1)	k. A.	p < 0,0001
Schmerzen unter Rippen	131	2,7 (2,3), -1,4 (2,1)	105	2,5 (2,5), 0,4 (2,3)		p < 0,0001
Völlegefühl	131	3,7 (2,2), -1,8 (2,2)	105	3,3 (2,2), 0,7 (2,3)		p < 0,0001
Nachschweiß	131	2,7 (2,5), -1,3 (2,2)	105	2,4 (2,1) 0,4 (1,9)		p < 0,0001
Juckreiz	131	2,2 (2,1) -1,3 (2,3)	105	2,3 (2,5), 0,4 (2,4)		p < 0,0001
Muskel- und Knochen- schmerzen	131	3,0 (2,5), -1,0 (2,1)	105	2,7 (2,4), 0,7 (1,9)		p < 0,0001

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
Inaktivität ³⁾	131	3,8 (2,6), -1,2 (1,9)	105	3,7 (2,4), 0,6 (2,3)		p < 0,0001
COMFORT II ⁴⁾						
Veränderung der Einzelsymptome des MFSAF v2.0 gegenüber Studienbeginn nach 24 Wochen (Range der Einzelsymptome: 0 = nicht vorhanden bis 10 = schlimmstmöglich)						
1) Bei der Darstellung zum Studienbeginn wurden die Patienten betrachtet, für die auch eine Auswertung zu Woche 24 vorlag. 2) Es wurde (abweichend von der Darstellung im Dossier) die absolute Veränderung der Einzelsymptome gegenüber Studienbeginn dargestellt. 3) Das Einzelsymptom Inaktivität ist nicht Teil des Total-Symptom-Scores (TSS), wurde bei den Einzelsymptomen aber dennoch einzeln ausgewertet und dargestellt. 4) Der Endpunkt wurde in dieser Studie nicht erhoben. 5) Wilcoxon-Rangsummentest						
k.A.: keine Angaben (im Dossier bzw. Studienbericht), SD: Standardabweichung						
Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (post-hoc-Analyse, ohne Crossover-Patienten (N = 126) ¹⁾)						
COMFORT I	Im Rahmen einer post-hoc-Analyse wurde untersucht, inwieweit die drei abdominalen Symptome des MFSAF (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl) mit dem objektiv gemessenen Milzvolumen korrelieren.					
		Besserung der Oberbauchbeschwerden, n (%)				
	Milzvolumen-rückgang, n (%)	< 50%		≥ 50%		Gesamt
	< 35%	34 (51,5)		32 (48,5)		66 (100)
	≥ 35%	22 (36,7)		38 (63,3)		60 (100)
COMFORT II ²⁾						
ECOG Performance-Status (Veränderung des ECOG Performance-Status gegenüber Studienbeginn) ³⁾ , ohne Crossover-Patienten						
COMFORT I ⁴⁾						
Verbesserung	94	42,6	79	29,1	k. A.	k. A.
Verschlechterung		2,1		20,3	k. A.	k. A.
COMFORT II	Auch in der Studie COMFORT II verschlechterten sich im Studienverlauf 60% aller konventionell behandelten Patienten mit ECOG-Wert von 0 zu Studienbeginn, aber nur 34,5% aller Ruxolitinib-Patienten mit einem Ausgangs-ECOG-Wert von 0. In dieser Studie waren die Ausgangsdaten zu Studienbeginn in etwa vergleichbar. Über den gesamten Studienverlauf hielten mehr Patienten den optimalen Wert von 0, wenn sie mit Ruxolitinib behandelt wurden als mit einer anderen Therapie (BAT). Auch einen Ausgangsscore von 1 konnten mehr Patienten im Ruxolitinib-Arm halten oder verbessern (84,4% der Patienten mit Score 1) als im BAT-Arm (57,1%; Dossier, Modul 4, S. 76).					

- 1) Es wurden die mit Ruxolitinib behandelten Patienten eingeschlossen, für die Daten zu Studienbeginn und zu Woche 24 sowohl für das Milzvolumen als auch den TSS vorlagen.
 2) Dieser Endpunkt wurde in der Studie nicht erhoben.
 3) Genauere Angaben zur Veränderung des ECOG Performance-Status im Zeitverlauf finden sich in den jeweiligen Extraktionsbögen zu den Studien.
 4) Angegeben ist die Veränderung gegenüber Studienbeginn nach 24 Wochen.

BAT: Best-Available-Therapy, CT: Computertomographie, ECOG: Eastern Cooperative Oncology Group, MFSAF: Myelofibrosis Symptoms Assessment form, MRT: Magnetresonanztomographie, OR: Odds Ratio

Die Endpunkte Leukämiefreies Überleben und Progressionsfreies Überleben wurden entgegen der Einschätzung des pharmazeutischen Unternehmers als Morbiditätsendpunkte kategorisiert, da es sich um kombinierte Endpunkte handelt, die sich überwiegend aus Endpunkten aus dem Bereich Morbidität zusammensetzen (s. Definition in Tabelle 10). Ergebnisse zu beiden Endpunkten wurden nur für die Studie COMFORT II präsentiert und es konnte jeweils kein statistisch signifikanter Gruppenunterschied gezeigt werden. In der COMFORT I-Studie war das Leukämiefreie Überleben als Langzeitendpunkt mit einer Auswertung nach 144 Wochen geplant. Entsprechende Analysen liegen noch nicht vor. Das Progressionsfreie Überleben wurde in dieser Studie nicht als Endpunkt erhoben.

Der primäre Endpunkt in beiden Studien war die Bestimmung des Anteils an Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ gemessen mittels MRT oder CT nach 24 (COMFORT I) bzw. 48 (COMFORT II) Wochen. Eine Volumenreduktion um $\geq 35\%$ wurde dabei als klinisch relevant angesehen (Tefferi et al., 2006; Barosi et al., 2005) und als Reponsekriterium zur Dichotomisierung der ausgewerteten Patienten definiert. Dieser Endpunkt wurde in beiden Studien erreicht und es konnte für diesen Endpunkt die signifikante ($p < 0,001$) Überlegenheit von Ruxolitinib gegenüber dem jeweiligen Komparator gezeigt werden. In der Studie COMFORT I erreichten nach 24 Wochen 65 Patienten (41,9 %) unter Ruxolitinib den primären Endpunkt gegenüber 1 Patient (0,7 %) in der Placebo-Gruppe. In der COMFORT II-Studie waren es nach 48 Wochen 41 Patienten (28,5 %) in der Ruxolitinib-Gruppe und kein Patient im BAT-Vergleichsarm. Ergänzend dazu wurde in der COMFORT II-Studie die Reponse rate nach 24 Wochen erhoben und auch hier zeigte sich ein statistisch signifikanter Vorteil von Ruxolitinib gegenüber Placebo (31,9 % vs. 0 %, $p < 0,001$). Bei der prozentualen Veränderung des Milzvolumens konnte in beiden Studien zum Zeitpunkt der Erhebung des primären Endpunkts unter Ruxolitinib eine mediane Milzvolumenabnahme um ca. 30 % gegenüber dem Studienbeginn gezeigt werden, wohingegen das Milzvolumen in den Komparator-Armen zunahm (je 8,5 %). Die Verteilung der Milzvolumenveränderung wird auch aus den in den Datenextraktionsbögen ergänzend dargestellten Wasserfall-Plots deutlich. Der Beginn des Ansprechens wurde lediglich in der COMFORT II-Studie bestimmt und mit einem Median von ca. 12 Wochen in der Ruxolitinib-Gruppe angegeben. In der BAT-Gruppe hat nur ein Patient überhaupt angesprochen und zwar nach ca. 15 Wochen. Die Dauer des Ansprechens konnte in der COMFORT I-Studie – obwohl a priori geplant – nicht berechnet werden, da das Ansprechen auf die Therapie zum Zeitpunkt des Data-Lock bei der Mehrheit der Patienten noch anhielt. In der COMFORT II-Studie wurde die Dauer des Ansprechens auf Ruxolitinib (N = 69) mit ca. 48 Wochen berechnet. Eine Ansprechdauer im Kontrollarm konnte nicht berechnet werden, da unter BAT nur ein Teilnehmer mindestens einmal über dem Schwellenwert von 35% Milzvolumenrückgang lag.

Zur Erfassung der krankheitsbedingten Symptome wurde in der Studie COMFORT I der MFSAF v2.0 bzw. ein daraus gebildeter Summenscore (Total-Symptom-Score, TTS) verwendet, der sich aus sechs der sieben krankheitsspezifischen Symptomen des MFSAF v2.0 zusammensetzt (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl, Nachtschweiß, Juckreiz, Muskel- und Knochenschmerzen; ohne das Symptom Inaktivität). Das Einzelsymptom Inaktivität wurde auf Vorschlag der FDA nicht als Symptom gewertet, da Inaktivität eine Folge der eigentlichen Symptome ist. Als klinisch relevante Verbesserung wurde eine Abnahme der Symptomatik im Total-Symptom-Score um $\geq 50\%$ gewertet. Der Anteil an Patienten, die nach 24 Wochen eine $\geq 50\%$ ige Verbesserung der Symptome

aufwies, war in der Ruxolitinib-Gruppe statistisch signifikant größer als in der Placebo-Gruppe (45,9 % vs. 5,3 %, $p < 0,0001$). Des Weiteren wurde auch die absolute Veränderung des TTS gegenüber dem Studienbeginn als ein sekundärer Endpunkt ausgewertet. Hier zeigte sich in der Ruxolitinib-Gruppe eine mediane Verbesserung von 6,9 Punkten gegenüber einer medianen Verschlechterung um 2,0 Punkte in der Placebo-Gruppe. Dies entspricht einer medianen Verbesserung von 56,2 % in der Ruxolitinib-Gruppe und einer Verschlechterung um 14,6 % in der Placebo-Gruppe. Der Gruppenunterschied war statistisch signifikant ($p < 0,0001$ im Wilcoxon-Rangsummentest). Auch bei den sieben einzeln ausgewerteten krankheits-spezifischen Symptomen, die mithilfe des MFSAF v2.0 abgefragt wurden, zeigte sich jeweils ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen (alle p -Werte $< 0,0001$ im Wilcoxon-Rangsummentest). Eine post-hoc-Analyse wurde in dieser Studie durchgeführt, um den Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl) zu untersuchen. Hier zeigte sich kein klarer Zusammenhang zwischen dem Rückgang des Milzvolumens und den untersuchten Oberbauchbeschwerden. Zwar konnten bei 63,3 % der Patienten, die eine $\geq 35\%$ ige Reduktion des Milzvolumens erreicht hatten auch ein Rückgang der Oberbauchbeschwerden um $\geq 50\%$ festgestellt werden; allerdings konnte eine entsprechende Abnahme der Symptomatik bei 36,7 % der Responder hinsichtlich einer Milzvolumenabnahme nicht festgestellt werden. Andererseits war bei ca. der Hälfte der Patienten ein Rückgang der Oberbauchbeschwerden um $\geq 50\%$ nicht mit einer Abnahme des Milzvolumens um $\geq 35\%$ assoziiert.

In beiden Studien wurde der ECOG Performance-Status erhoben und die Veränderung des ECOG Performance-Status im Studienverlauf dargestellt. Bei dem ECOG Performance-Status handelt es sich um ein generisches Messinstrument, um funktionelle Einschränkungen von Krebspatienten in ihrem Alltag zu erfassen und zu klassifizieren. Da dieses Messinstrument daher nicht geeignet ist, um detailliert die Lebensqualität der Patienten zu erheben und therapiebedingte Änderungen der Lebensqualität zu erfassen, wird dieser Endpunkt in der vorliegenden Bewertung in die Endpunktkategorie Morbidität eingestuft.

Die Ergebnisse werden in den Studienberichten deskriptiv – ohne einen direkten Gruppenvergleich – dargestellt. In der COMFORT I-Studie zeigte sich in der Ruxolitinib-Gruppe bei 42,6% der ausgewerteten Patienten eine Verbesserung des ECOG Performance-Status nach 24 Wochen gegenüber Studienbeginn, wohingegen sich nur bei 2,1 % eine Verschlechterung einstellte. In der Placebo-Gruppe kam es bei 20,3 % der ausgewerteten Patienten zu einer Verschlechterung und bei 29,1 % zu einer Verbesserung gegenüber den Werten zu Studienbeginn. Auch in der Studie COMFORT II verschlechterten sich im Studienverlauf mehr Patienten in der BAT-Gruppe als in der Ruxolitinib-Gruppe. Über den gesamten Studienverlauf hielten mehr Patienten den optimalen Wert von 0, wenn sie mit Ruxolitinib behandelt wurden als mit der BAT-Therapie.

Tabelle 13: Ergebnisse zur gesundheitsbezogenen Lebensqualität (Ruxolitinib vs. Placebo bzw. BAT)

Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
QLQ-C30 der EORTC (durchschnittliche absolute Veränderung der Skalen nach 24 [COMFORT I] bzw. 48 [COMFORT II] Wochen gegenüber Studienbeginn, ohne Crossover-Patienten), Range jeder Einzelskala 0-100						
	N	Studienbeginn Mittelwert (SD), abs. Veränderung zu Studienbeginn (SD)	N	Studienbeginn Mittelwert (SD), abs. Veränderung zu Studienbeginn (SD)	Effekt- maß	p-Wert ²⁾
COMFORT I ¹⁾ , absolute Veränderung nach 24 Wochen						
Gesundheit gesamt	136	52,6 (23,1), 12,3 (25,4)	104	54,5 (22,6), -3,4 (21,5)	k. A.	< 0,0001
Körperliche Funktion	137	69,5 (21,1), 9,3 (18,5)	107	67,6 (20,5), -3,5 (17,4)	k. A.	< 0,0001
Rollenfunktion	137	64,6 (29,2), 6,4 (29,7)	105	65,1 (30,6), -11,1 (30,9)	k. A.	< 0,0001
Emotionale Funktion	137	73,7 (23,5), 8,3 (20,7)	105	76,3 (23,5), -1,2 (23,0)	k. A.	< 0,0009
Kognitive Funktion	135	81,1 (19,1), 2,5 (19,0)	103	79,3 (20,1), -1,1 (21,0)	k. A.	0,0587
Soziale Funktion	135	68,0 (26,6), 9,4 (25,5),	106	65,9 (29,8), -9,0 (28,1)	k. A.	< 0,0001
COMFORT II ⁴⁾ , absolute Veränderung nach 24 bzw. 48 Wochen ³⁾						
Gesundheit gesamt	95 66	55,5 (20,8), w 24: 8,4 (18,57) w 48: 9,1 (24,23)	37 27	50,1 (23,3) w 24: 5,2 (23,76) w 48: 3,4 (24,27)	k. A.	k. A.
Rollenfunktion	96 69	67,0 (30,1), w 24: 5,4 (27,52) w 48: 9,9 (28,05)	39 28	64,4 (32,9), w 24: -6,0 (27,96) w 48: -5,4 (26,47)	k. A.	k. A.
Fatigue	98 69	44,4 (27,6), w 24: -8,1 (23,26) w 48: -12,8 (28,57)	39 28	51,2 (28,7), w24: -1,0 (23,57) w 48: 0,4 (25,39)	k. A.	k. A.
Schmerzen	98 69	23,6 (29,3), w 24: -5,4 (24,50) w 48: -1,9 (28,80)	39 28	30,6 (30,8), w 24: 3,0 (28,06) w 48: 3,0 (17,60)	k. A.	k. A.
Atemnot	97 69	34,3 (30,9), w 24: -1,7 (28,61) w 48: -6,3 (30,40)	38 28	38,8 (36,5), w 24: 5,3 (31,51) w 48: 4,8 (34,80)	k. A.	k. A.
Schlaflosigkeit	97 68	38,0 (34,6), w 24: -9,3 (31,08) w 48: -12,3 (35,46)	39 28	24,9 (31,4), w 24: 6,0 (28,48) w 48: 6,0 (24,09)	k. A.	k. A.

Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
Appetit- losigkeit	98 69	16,7 (26,0), w 24: -10,9 (24,77) w 48: -8,2 (27,06)	39 28	21,4 (32,2), w 24: 0,9 (34,61) w 48: 9,5 (33,77)	k. A.	k. A.
Diarrhoe	95 65	17,8 (26,8), w 24: -5,3 (25,87) w 48: -4,1 (23,20)	37 27	14,6 (26,2), w 24: 2,7 (31,80) w 48: -6,2 (24,52)	k. A.	k. A.
1) Im Studienbericht (COMFORT I) wurden nur die funktionalen Skalen dargestellt. 2) p-Werte entstammen einer Kovarianzanalyse. 3) Für die Auswertungen zur Woche 8 und 16 siehe Datenextraktionsbogen. 4) Im Studienbericht (COMFORT II) wurden detailliert nur diejenigen Funktionsbereiche und Symptome dargestellt, bei denen an mindestens einem der Erfassungszeitpunkte ein Unterschied von mindestens 10 Punkten zwischen den Behandlungs-gruppen sichtbar wurde. Dieser Wert wird vom pU als Schwellenwert zur Definition einer klinischen Relevanz deklariert. k.A.: keine Angaben (im Dossier und im Studienbericht)						
FACT-Lym (absolute Veränderung der Skalen gegenüber Studienbeginn, Observed Cases)						
	N	Studienbeginn (SD), Abs. Veränderung zu Studienbeginn (SD) nach 8, 16, 24 und 48 Wochen	N	Studienbeginn (SD), Abs. Veränderung zu Studienbeginn (SD) nach 8, 16, 24 und 48 Wochen	Effekt- maß	p-Wert
COMFORT I ¹⁾						
COMFORT II						
FACT-Lym Total Score (Range 0-168)	141 116 116 100 70	116,2 (24,89), 8 w: 12,1 (22,1) 16 w: 10,5 (22,0) 24 w: 11,9 (20,3) 48 w: 11,3 (25,0)	70 50 42 40 29	113,5 (23,29), 8 w: -0,1 (19,9) 16 w: -3,3 (13,4) 24 w: -0,8 (17,1) 48 w: -0,9 (19,6)	k. A.	k. A.
FACT-TOI (Range 0-116)	141 116 116 100 70	78,6 (19,84), 8 w: 10,1 (17,0) 16 w: 8,7 (15,9) 24 w: 9,0 (14,2) 48 w: 9,1 (18,8)	70 50 42 40 29	76,0 (19,84), 8 w: 0,0 (14,4) 16 w: -2,7 (10,0) 24 w: -0,4 (13,8) 48 w: -0,9 (16,1)	k. A.	k. A.
FACT-G-Total Score (Range 0-108)	141 116 116 100 68	73,5 (17,19), 8 w: 6,3 (15,0) 16 w: 5,3 (16,4) 24 w: 6,3 (17,2) 48 w: 8,9 (15,9)	70 50 42 40 29	72,0 (14,06), 8 w: -0,1 (12,5) 16 w: -2,9 (10,4) 24 w: -1,1 (11,0) 48 w: 0,1 (13,8)	k. A.	k. A.
LymS (Range 0-60)	139 112 113 99 67	43,4 (9,67), 8 w: 5,9 (7,3) 16 w: 5,7 (7,3) 24 w: 6,1 (6,7) 48 w: 6,0 (7,6)	69 49 42 40 28	42,1 (10,41), 8 w: 1,2 (6,3) 16 w: -0,4 (5,99) 24 w: 0,3 (8,4) 48 w: 0,7 (9,1)	k. A.	k. A.

1) Der Endpunkt wurde in dieser Studie nicht erhoben

FACT: Functional Assessment of Cancer Therapy, SD: Standardabweichung, w: Wochen

Die Lebensqualität wurde mit Hilfe verschiedener, für das Krankheitsbild unterschiedlich spezifischer Messinstrumente bestimmt.

Zur Erfassung der Lebensqualität wurde in beiden Studien der EORTC QLQ-C30 verwendet. Dieser 30 Items umfassende Fragebogen dient als generisches Messinstrument zur Erfassung der Lebensqualität von Krebspatienten. Hierbei wird zum einen ein globaler Gesundheitszustand erfasst und zum anderen verschiedene Funktions- und Symptomskalen abgefragt. Nach 24 Wochen hatten sich in der COMFORT I-Studie in der Ruxolitinib-Gruppe die QLQ-C30 Scores im Gesamtbild und 4 der 5 funktionellen Skalen statistisch signifikant gegenüber Placebo verbessert (Ausnahme: kognitive Funktion). Im Studienbericht findet sich die deskriptive Angabe, dass auch für die weiteren Symptome und Skalen des QLQ-C30 (z.B. Fatigue, Dyspnoe und Schmerz) – mit Ausnahme der Obstipation – eine signifikante Überlegenheit von Ruxolitinib gegenüber Placebo gezeigt werden konnte ($p < 0,001$, ANCOVA).

Im Studienbericht zur COMFORT II-Studie wurden detailliert nur diejenigen Funktionsbereiche dargestellt, bei denen an mindestens einem der Erfassungszeitpunkte (8, 16, 24, 48 Wochen) ein Unterschied von mindestens 10 Punkten zwischen den Behandlungsgruppen sichtbar wurde, da dieser Schwellenwert vom pharmazeutischen Unternehmer zur Definition eines klinisch relevanten Unterschiedes herangezogen wird. Ein darüber hinaus gehender statistischer Gruppenvergleich wurde nicht durchgeführt (eine kritische Auseinandersetzung mit diesem Schwellenwert findet sich in Tabelle 11 bzw. in Abschnitt 3.2.2). Beim Gesamtgesundheitszustand und bei der Rollenfunktion konnte ein Gruppenunterschied von > 10 Punkten an mindestens einem Messzeitpunkt gesehen werden. Auch bei den Symptomen Fatigue, Schmerzen, Atemnot, Appetitlosigkeit und Diarrhoe zeigte sich im Beobachtungszeitraum eine Gruppendifferenz von > 10 Punkten. Bei dem Symptom Schlaflosigkeit war der Schwellenwert ebenfalls überschritten, wobei hier bei der Ergebnisbewertung die deutlichen Unterschiede zwischen den Gruppen zu Studienbeginn berücksichtigt werden müssen. Bei allen anderen Unterskalen wurde der Schwellenwert im Studienverlauf zu keinem Zeitpunkt überschritten (Daten nicht separat extrahiert).

In der unverblindeten COMFORT II-Studie wurden darüber hinaus die Fragebögen FACT-G und FACT-LymS bzw. die daraus zusammengesetzten Skalen FACT-Lym Total Score und FACT-TOI verwendet, um spezifischere Einflüsse der Myelofibrose-Symptomatik auf die Lebensqualität zu erfassen. Unter einer Behandlung mit Ruxolitinib verbesserte sich der mittlere Gesamtscore im Vergleich zu Studienbeginn in allen Skalen, während er unter BAT unverändert blieb. Die Unterschiede zeigten sich ab der ersten Erhebung in Woche 8 bis zur Woche 48. Ein statistischer Vergleich wurde nicht durchgeführt.

Tabelle 14: Ergebnisse zu Nebenwirkungen (Ruxolitinib vs. Placebo bzw. BAT), basierend auf der Safety-Population

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
Studie	Gesamt N	Ergebnisse N (%)	Gesamt N	Ergebnisse N (%)	RR (95 %-KI)	p-Wert
Unerwünschte Ereignisse (Patienten mit mind. einem unerwünschten Ereignis)						
COMFORT I	155	151 (97,4)	151	148 (98,0)	-	-
COMFORT II	146	145 (99,3)	73	66 (90,4)	-	-
Schwerwiegende unerwünschte Ereignisse (Patienten mit mind. einem schwerwiegenden unerwünschten Ereignis)						
COMFORT I	155	43 (27,7)	151	53 (35,1)	-	-
COMFORT II	146	44 (30,1)	73	21 (28,8)	-	-
Unerwünschte Ereignisse der CTCAE, Grad 3 und 4						
COMFORT I	155	73 (47,1)	151	67 (44,4)	-	-
COMFORT II	146	63 (43,2)	73	18 (24,7)	-	-
Abbruch aufgrund unerwünschter Ereignisse (Patienten mit Abbruch aufgrund unerwünschter Ereignisse)						
COMFORT I	155	17 (11,0)	151	16 (10,6)	-	-
COMFORT II	146	12 (8,2)	73	6 (8,2) ¹⁾	-	-
Unerwünschte Ereignisse, die zu einer Dosisreduktion oder Unterbrechung der Therapie führten						
COMFORT I	155	62 (40,0) ²⁾	151	14 (9,3) ²⁾	-	-
COMFORT II	146	92 (63,0)	73	11 (15,1)	-	-
Thrombozytopenie (Preferred Term der MedDRA)						
COMFORT I Alle Grade Grad 3 / 4	155	53 (34,2) 13 (8,4)	151	14 (9,3) 3 (2,0)	-	-
COMFORT II Alle Grade Grad 3 / 4	146	65 (44,5) 11 (7,5)	73	6 (9,6) 3 (4,1)	-	-
Anämie (Preferred Term der MedDRA)						
COMFORT I Alle Grade Grad 3 / 4	155	48 (31,0) 24 (15,5)	151	21 (13,9) 87 (4,6)	-	-
COMFORT II Alle Grade Grad 3 / 4	146	60 (41,1) 9 (12,3)	73	9 (12,3) 3 (4,1)	-	-

Endpunkte Studie	Ruxolitinib		Kontrolle (Placebo bzw. BAT)		Intervention vs. Kontrolle	
Studie	Gesamt N	Ergebnisse N (%)	Gesamt N	Ergebnisse N (%)	RR (95 %-KI)	p-Wert
Blutungen (Operationalisiert als Ekchymose [PT] oder Blutungsereignisse)						
COMFORT I Ekchymose Blutungsereignisse ⁴⁾	155	29 (18,7) 17 (11,0) davon 6 SUEs	151	14 (9,3) 13 (8,6) davon 5 SUEs		
COMFORT II Ekchymose Blutungsereignisse ⁵⁾	146	2 (1,4) n.d.	73	0 (0) n.d.	-	-
Infektionen (SOC „Infections and Infestations“)						
COMFORT I ⁶⁾ Alle Grade Grad 3-5	155	59 (38,1) 13 (8,4)	151	64 (42,4) 11 (7,3)	-	-
COMFORT II ⁷⁾ Alle Grade Grad 3 / 4	146	93 (63,7) 19 (13,0)	73	32 (43,8) 3 (4,1)	-	-

- 1) Zu diesem Endpunkt finden sich für die BAT-Gruppe widersprüchliche Angaben im Dossier und im Studienbericht. Es wurden die Angaben aus dem Studienbericht COMFORT II, Table 12-11 (S. 150) extrahiert.
- 2) Zu diesem Endpunkt finden sich widersprüchliche Angaben im Dossier und im Studienbericht. Es wurden die Daten aus dem Studienbericht COMFORT I, Table 26 (S. 150) extrahiert.
- 3) Diese Unerwünschten Ereignisse von besonderem Interesse wurden seitens der EMA als besondere Risiken identifiziert.
- 4) Operationalisiert in COMFORT I als “Preferred Terms containing the words bleeding, hemorrhage or epistaxis”) Diese Auswertung wurde nicht als Endpunkt a priori definiert, sondern nachträglich durch die Gruppierung mehrerer Preferred Terms durchgeführt.
- 5) Eine Gruppierung zum Konzept “Blutungsereignisse” wurde in der COMFORT II-Studie nicht ausgewertet. Eine nachträgliche Gruppierung relevanter Preferred Terms ist aufgrund möglicher Doppelzählungen nicht möglich.
- 6) vgl. Tab. 33 im Studienbericht (S. 163)
- 7) vgl. Tab. 12-5 im Studienbericht (S. 143)

CTCAE: Common Terminology Criteria for Adverse Events, MedDRA: Medical Dictionary for Regulatory Activities, n.d.: Analyse wurde vom pharmazeutischen Unternehmer nicht durchgeführt, PT: Preferred Term, SOC: System Organ Class

Tabelle 15: Deskriptive Auflistung häufiger unerwünschter Ereignisse (Grad 3 und 4), die in > 3% der mit Ruxolitinib behandelten Patienten in mind. 1 Studienarm auftraten (Ruxolitinib vs. Placebo bzw. BAT), Preferred Terms

	Ruxolitinib N=155, Anzahl Patienten mit Ereignis (%)	Placebo N=151, Anzahl Patienten mit Ereignis (%)	Ruxolitinib N=146, Anzahl Patienten mit Ereignis (%)	Placebo N=73, Anzahl Patienten mit Ereignis (%)
Studie	COMFORT I		COMFORT II	
Alle UEs, Grade 3 und 4	73 (47,1)	67 (44,4)	61 (41,8)	18 (24,7)
Anämie	24 (15,5)	7 (4,6)	16 (11,0)	3 (4,1)

	Ruxolitinib N=155, Anzahl Patienten mit Ereignis (%)	Placebo N=151, Anzahl Patienten mit Ereignis (%)	Ruxolitinib N=146, Anzahl Patienten mit Ereignis (%)	Placebo N=73, Anzahl Patienten mit Ereignis (%)
Hämoglobin erniedrigt	15 (9,7)	3 (2,0)	4 (2,7)	2 (2,7)
Thrombozytopenie	13 (8,4)	3 (2,0) ^{b)}	11 (7,5)	3 (4,1)
Ermüdung (Fatigue)	8 (5,2)	10 (6,6)	1 (0,7)	0 (0)
Pneumonie	8 (5,2)	6 (4,0)	2 (1,4)	3 (4,1)
Abdominalschmerz	4 (2,6)	17 (11,3)	5 (3,4)	2 (2,7)
Dyspnoe	2 (1,3)	6 (4,0)	1 (0,7)	3 (4,1)
Hyponatraemie	1 (0,6)	5 (3,3)	-	-
Infarkt der Milz	1 (0,6)	5 (3,3)	0 (0)	0 (0)

Es liegen deskriptive Angaben zu Patientenanteilen mit unerwünschten Ereignissen vor. Ein statistischer Gruppenvergleich wurde nicht durchgeführt. Zur Evaluation der Nebenwirkungen wurde jeweils die Safety-Population verwendet (Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten). Unerwünschte Ereignisse wurden mit Hilfe der MedDRA-Terminologie (Version 12) kodiert. Es werden die unerwünschten Ereignisse unter Therapie unabhängig von ihrem vermuteten Kausalzusammenhang mit der Behandlung dargestellt und diskutiert.

In beiden Studien war der Anteil der Patienten mit mindestens einem unerwünschten Ereignis sehr hoch (COMFORT I: in beiden Gruppen ca. 98%, COMFORT II: 99,3% [Ruxolitinib], 90,4% [BAT]). Schwerwiegende unerwünschte Ereignisse traten in beiden Studien und in den jeweiligen Behandlungsgruppen mit einer vergleichbaren Häufigkeit von ca. 30% auf, sodass sich hier kein Gruppenunterschied aus den Ergebnissen ableiten lässt. In der COMFORT I-Studie war der Anteil der Patienten mit einem unerwünschten Ereignis Grad 3 oder 4 der CTCAE in beiden Studienarmen vergleichbar (ca. 45%). In der COMFORT II-Studie zeigte sich hier jedoch ein deutlicher Unterschied zu Ungunsten von Ruxolitinib (43,2% vs. 24,7%), was insbesondere auf das Auftreten von schweren Anämien (12,3% vs. 4,1%) und Thrombozytämien (7,5% vs. 4,1%) in der Ruxolitinib-Gruppe zurückzuführen ist. In beiden Studien kam es in den Behandlungsgruppen vergleichbar häufig zu einem Studienabbruch aufgrund von unerwünschten Ereignissen (COMFORT I: je ca. 11%, COMFORT II: je ca. 8%). Die meisten spezifischen unerwünschten Ereignisse, die zu einem Studienabbruch führten, kamen nur bei jeweils einem Patienten vor. Insgesamt kam es in beiden Studien in der Ruxolitinib-Gruppe deutlich häufiger zu einer Dosisreduktion aufgrund unerwünschter Ereignisse (COMFORT I: 40% vs. 9,3%; COMFORT II: 63% vs. 15,1%), was auf eine insgesamt schlechtere Verträglichkeit von Ruxolitinib gegenüber den jeweiligen Komparatoren hinweist.

In beiden Studien waren Anämie und Thrombozytopenie die beiden häufigsten hämatologischen unerwünschten Ereignisse (alle Grade und Grad 3 und 4) und traten deutlich häufiger in der Ruxolitinib-Gruppe auf. Diese Nebenwirkung lässt sich auf den Wirkmecha-

nismus von Ruxolitinib als JAK-Inhibitor zurückführen. Thrombozytämie und Anämie haben allerdings nur selten zu einem Studienabbruch geführt ($< 1\%$) und ließen sich gemäß den Angaben des pU in den Studien durch Dosisreduktionen und Bluttransfusionen gut beherrschen. Darüber hinaus wurden als unerwünschte Ereignisse von besonderem Interesse Infektionen (operationalisiert als die Systemorganklasse „Infections and infestations“) und Blutungen (Ekchymosen und Blutungsereignisse) betrachtet. Die Auswertung der Infektionen kommt in beiden Studien zu abweichenden Resultaten. In der COMFORT I-Studie kamen unerwünschte Infektionsereignisse vergleichbar häufig vor (38,1 % vs. 42,4 %), wohingegen in der COMFORT II-Studie Infektionen unter Ruxolitinib deutlich häufiger als unter BAT vorkamen (63,7 % vs. 43,8 %). Dieser Trend zeigt sich in der COMFORT II-Studie auch bei den schweren Infektionen der CTCAE Grade 3 und 4 (13 % vs. 4,1 %).

Bei den Blutungen zeigten sich ebenfalls unterschiedliche Ergebnisse in den Zulassungsstudien: während in der COMFORT I-Studie unter Ruxolitinib deutlich mehr Ekchymosen aufgetreten sind als unter Placebo (18,7 % vs. 9,3 %), trat dieses Ereignis in der COMFORT II-Studie kaum auf (1,4 % vs. 0 %). Bei den Blutungsereignissen konnte in der COMFORT I-Studie kein relevanter Gruppenunterschied gesehen werden (11 % vs. 8,6 %). In der COMFORT II-Studie wurde diese Auswertung nicht durchgeführt.

Weitere nicht-hämatologische unerwünschte Ereignisse traten in der COMFORT I-Studie mit vergleichbarer Häufigkeit in beiden Gruppen auf (überwiegend mit dem Schweregrad 1 und 2; siehe Datenextraktionsbögen zur Auflistung der einzelnen Nebenwirkungen).

Ergebnisse aus Subgruppenanalysen

In beiden Studien waren Subgruppenanalysen für den primären Endpunkt „Anteil der Patienten mit einem Milzvolumenrückgang $\geq 35\%$ “ a priori geplant. In der COMFORT I-Studie wurden zusätzlich (post hoc berechnete) Subgruppenanalysen für den sekundären Endpunkt „Anteil der Patienten mit einer Verbesserung um $\geq 50\%$ im Total-Symptom-Score“ berichtet, die hier jedoch nicht weiter diskutiert werden (aber im Datenextraktionsbogen dargestellt sind).

Für den primären Endpunkt zeigen sich auffällige Unterschiede in der Ruxolitinib-Gruppe hinsichtlich des JAK2-Mutationsstatus. In beiden Studien haben die JAK2V617F-positiven Patienten ein höheres Ansprechen als die Patienten ohne diese Mutation (COMFORT I: 47,8 % vs. 27,5 %, COMFORT II: 32,7 % vs. 14,3 %). Ergänzend zu den Angaben im Studienbericht findet sich in der Primärveröffentlichung der COMFORT I-Studie ein Interaktionstest ($p = 0,07$), was als Hinweis auf eine relevante Heterogenität zwischen den Patienten mit und ohne JAK2V617F-Mutation gedeutet werden kann (Verstovsek et al., 2012). Die Subgruppenanalyse in der COMFORT I-Studie deutet auch auf unterschiedliches Ansprechen auf Ruxolitinib zwischen Männern und Frauen hin (Männer: 25,3 %, Frauen: 59,2 %). Dieser Trend zeigt sich abgeschwächt auch in der COMFORT II-Studie (Männer: 24,7 %, Frauen: 33,3 %). Auch eine höhere Thrombozytenzahl ($> 200.000/\mu\text{l}$ gegenüber $< 200.000/\mu\text{l}$) scheint mit einem besseren Ansprechen einherzugehen (COMFORT I: 53 % vs. 21,8 %, COMFORT II: 35,2 % vs. 17,9 %). In der COMFORT II-Studie zeigte sich auch ein deutlicher Unterschied hinsichtlich des Ansprechens auf den primären Endpunkt in Abhängigkeit von der Indikation (PPV-MF: 41%, PMF: 18,4%, PET-MF: 35%). Dieser deutliche Trend konnte in der COMFORT I-Studie allerdings nicht gesehen werden.

Der pharmazeutische Unternehmer gibt an, dass sich sämtliche untersuchten Vergleichsgruppen signifikant positiv von der Vergleichsbehandlung unterscheiden (Dossier, Modul 4, S. 89) und leitet daraus subgruppenübergreifend einen therapeutisch bedeutsamen (beträchtlichen) Zusatznutzen für alle Patienten im zugelassenen Anwendungsgebiet ab.

2.7. Einschätzung zum Ausmaß des Zusatznutzens durch den pharmazeutischen Unternehmer

Der pharmazeutische Unternehmer leitet aus den Daten der beiden Zulassungsstudien einen beträchtlichen Zusatznutzen für die gesamte Zielpopulation, für die Ruxolitinib zugelassen ist, ab. Dies schließt Erwachsene Patienten mit PMF, PET-MF und PEV-MF ein. Er begründet seine Einschätzung mit

- dem signifikanten Rückgang der Milzvergrößerung,
- der erheblichen Linderung der klinischen Symptomatik,
- der starken Verbesserung der Lebensqualität und Belastbarkeit bei
- insgesamt adäquater Verträglichkeit.

Da es bisher keine zugelassene wirksame Therapie zur Behandlung der Splenomegalie und der verschiedenen Subtypen der Myelofibrose gibt, ist die für einen beträchtlichen Zusatznutzen notwendige Grundvoraussetzung „bisher nicht erreicht“ grundsätzlich erfüllt.

Aus den im Dossier dargelegten Subgruppenanalysen leitet er keine relevanten Interaktionen ab und stellt subgruppenübergreifend einen beträchtlichen Zusatznutzen für alle Patienten im zugelassenen Indikationsgebiet fest.

3. Methodische Anmerkungen und Bewertung der Unterlagen

3.1. Design und Methodik der Studien

Die vorliegende Bewertung beruht auf den beiden Zulassungsstudien COMFORT I und COMFORT II. Es handelt sich bei COMFORT I um eine multizentrische, parallele, doppelblinde, randomisierte, kontrollierte Studie von Ruxolitinib gegenüber Placebo. Der Einsatz eines Placebo-Kontrollarms ist in der eingeschlossenen Patientenpopulation insofern vertretbar, als dass die Patienten unter anderen Therapien resistent oder refraktär sein mussten, diese nicht vertragen haben und – nach der Meinung des behandelnden Arztes – nicht für eine andere verfügbare Therapie infrage kommen durften. COMFORT II ist eine multizentrische, parallele, offene, randomisierte, kontrollierte Studie von Ruxolitinib gegenüber einer vom Prüfarzt gewählten „besten verfügbaren Therapie“ (BAT). Das offene Design in der COMFORT II-Studie ist durch den Einsatz vielfältiger Komparatoren im BAT-Arm erklärbar, muss jedoch bei der Betrachtung subjektiv erhobener Endpunkte berücksichtigt werden. In Absprache mit der EMA diente diese Studie insbesondere dazu, Daten zu Wirksamkeit und Sicherheit über einen längeren Zeitraum (48 Wochen) zu erheben sowie einen direkten Vergleich zu der aktuellen Standardtherapie der Myelofibrose zu generieren. In beiden Studien wurde Ruxolitinib in Tablettenform zweimal täglich in einer initialen Dosierung von jeweils 15 bis 20 mg verwendet (Tagesdosis 30 bis 40 mg), wobei eine im Protokoll spezifizierte Dosisanpassung bei fehlender Wirksamkeit bzw. auftretender Unverträglichkeit erlaubt war. Ein Crossover vom Vergleichs- in den Ruxolitinib-Arm war in beiden Studien auch bereits während der randomisierten Phase erlaubt und ist bei der Betrachtung der Ergebnisse zu berücksichtigen. Wahrscheinlich ist jedoch durch diesen erlaubten Therapiewechsel innerhalb der randomisierten Studienphasen (zumindest bei den Endpunkten zur Erfassung der Wirksamkeit) aufgrund der ITT-Auswertung keine Überschätzung der Therapieeffekte von Ruxolitinib zu erwarten (konservative Auswertung). Kritisch zu sehen ist die für die Indikation vergleichsweise kurze Studiendauer der beiden Zulassungsstudien (24 bzw. 48 Wochen). Ungefähr 75 % der mit Ruxolitinib behandelten Patienten waren über beide Studien hinweg betrachtet zum Zeitpunkt der primären Analyse noch in dem Behandlungsarm, in den sie randomisiert wurden. Dies deutet darauf hin, dass das Medikament in der Praxis über einen deutlich längeren Zeitraum angewendet werden kann und unterstreicht die Notwendigkeit, Langzeitdaten zu Wirksamkeit und Sicherheit zu generieren und zu bewerten. Zwar werden die beiden Studien in Extensionsphasen weitergeführt, jedoch wurden nach der primären Datenanalyse alle eingeschlossenen Patienten auf Wunsch mit Ruxolitinib behandelt. Folglich sind keine weiteren, direkt vergleichenden Therapieeffekte mehr aus den nachfolgenden Analysen zu erwarten.

Die beiden in den Studien untersuchten Patientenpopulationen repräsentieren die gemäß Fachinformation zugelassene Zielpopulation, sie unterscheiden sich jedoch hinsichtlich ihrer Vorbehandlung: Die Patienten in der COMFORT I-Studie mussten unter anderen Therapien resistent oder refraktär sein, durften diese nicht vertragen haben und – nach der Meinung des behandelnden Arztes – nicht für eine andere verfügbare Therapie infrage kommen. Diese Einschränkung galt nicht für die Studie COMFORT II, sondern hier durften die Patienten lediglich nicht für eine Stammzelltransplantation infrage kommen. Dieser Unterschied zusammen mit den verschiedenen Vergleichsarmen in beiden Studien erschwert den direkten Vergleich der Ergebnisse in den Zulassungsstudien und macht eine Meta-Analyse wenig

sinnvoll. Die COMFORT I-Studie ist insbesondere dafür geeignet, den reinen Behandlungseffekt einer Ruxolitinib-Therapie zu beurteilen. Dahingegen ist die COMFORT II-Studie für die Beurteilung des Therapieeffektes gegenüber der gängigen Standardtherapie im Indikationsgebiet geeignet.

Laut öffentlichem Beurteilungsbericht der EMA (European Public Assessment Report, EPAR, EMA, 2012) konnten keine wesentlichen Amendments des Protokolls festgestellt werden, die die Aussagekraft der Ergebnisse und die daraus resultierenden Schlussfolgerungen infrage stellen würden.

Prinzipiell sind die beiden Studien adäquat durchgeführt und können einander ergänzend zur Beurteilung von Ruxolitinib herangezogen werden.

3.2. Endpunkte und Effekte

3.2.1. Primärer Endpunkt

Der primäre Endpunkt in beiden Studien war die Bestimmung des Anteils an Patienten mit einem Rückgang des Milzvolumens um mindestens 35 % gemessen mittels MRT oder CT nach 24 (COMFORT I) bzw. 48 (COMFORT II) Wochen. Eine Volumenreduktion um $\geq 35\%$ wurde dabei als klinisch bedeutsam angesehen (Tefferi et al., 2006; Barosi et al., 2005) und als Responsekriterium zur Dichotomisierung der ausgewerteten Patienten definiert. Da es sich um einen objektiv gemessenen Endpunkt mit verblindeter ITT-Auswertung handelt wird von einem niedrigen Verzerrungspotenzial ausgegangen. Bei der Interpretation der Ergebnisse ist zu berücksichtigen, dass es sich um einen Surrogatendpunkt (unklarer Validität) handelt. Zudem ist das bei diesem Endpunkt angewendete Responsekriterium einer Volumenreduktion um $\geq 35\%$ nicht ausreichend anhand klinischer Daten begründet. Seitens der EMA wird das Responsekriterium unter Berufung auf die vom pharmazeutischen Unternehmer angegebenen Quellen von Tefferi (2006) und Barosi (2005) als bedeutsam eingestuft. Die EMA folgt der Argumentation des pharmazeutischen Unternehmers, dass eine Abnahme der Milzgröße mit einer Abnahme krankheitsbedingter Symptome (z.B. Oberbauchbeschwerden, hypothetisch das Risiko für eine Splenektomie) einhergeht. Im Studienbericht der COMFORT II-Studie gibt der pharmazeutische Unternehmer jedoch auch an, dass die EMA im Rahmen einer Beratung (Scientific Advice) dazu geraten hat, über die Milzvolumenreduktion hinaus weitere (klinisch relevante) Endpunkte zu untersuchen, die direkt patientenrelevant sind (z.B. krankheitsbedingte Symptome, Notwendigkeit von Transfusionen). Es wird im EPAR zudem kritisch angemerkt, dass die Bedeutung der Dauer des klinischen Ansprechens noch nicht ausreichend untersucht ist (EMA, 2012).

Der primäre Endpunkt wurde in beiden Studien erreicht und es konnte eine statistisch signifikante Überlegenheit von Ruxolitinib gegenüber dem jeweiligen Komparator gezeigt werden. In der Studie COMFORT I erreichten nach 24 Wochen 65 Patienten (41,9 %) der Patienten unter Ruxolitinib den primären Endpunkt gegenüber einem Patienten (0,7 %) in der Placebo-Gruppe. In der COMFORT II-Studie waren es nach 48 Wochen 41 Patienten (28,5 %) in der Ruxolitinib-Gruppe und kein Patient im BAT-Vergleichsarm. Ergänzend dazu wurde in der COMFORT II-Studie der Anteil der Responder nach 24 Wochen erhoben und auch hier zeigte sich ein statistisch signifikanter Vorteil von Ruxolitinib gegenüber Placebo (31,9 % vs.

0 %, $p < 0,001$). Beim direkten Vergleich der Ansprechquoten zwischen den Studien fällt auf, dass die Patienten in der COMFORT I-Studie eine höheres Ansprechen unter Ruxolitinib aufweisen (41,9 % vs. 28,5 %). Eine mögliche Erklärung hierfür könnten die unterschiedlichen, eingeschlossenen Patientenpopulationen sein. Allerdings wird die klinische Ansprechquote seitens der EMA in beiden Studien als ausreichend hoch eingeschätzt, um einen klinischen Erfolg der Ruxolitinib-Therapie festmachen zu können. Die Wasserfall-Plots (s. Extraktionsbögen) bestätigen die deutliche Milzvolumenabnahme bei den meisten Patienten unter Ruxolitinib-Therapie gegenüber dem jeweiligen Komparator (Placebo oder BAT). Daten aus der COMFORT II-Studie werden im EPAR hinsichtlich des Beginns (nach ca. 12 Wochen) und der medianen Dauer (ca. 48 Wochen) des klinischen Ansprechens als Anhaltspunkte für eine klinische Relevanz erachtet. Hier sind jedoch weitere Langzeitdaten erforderlich.

3.2.2. Weitere Endpunkte

Mortalität

Prinzipiell waren die beiden Studien nicht dafür geeignet, um Unterschiede hinsichtlich des Gesamtüberlebens aufzudecken. Es konnte zum Zeitpunkt der Erhebung des primären Endpunkts („Primäranalyse“) in keiner der Zulassungsstudien ein statistisch signifikanter Gruppenunterschied hinsichtlich der Gesamtmortalität gesehen werden. Ergänzende Analysen, die in der COMFORT I-Studie 4 Monate nach der Primäranalyse im Rahmen eines Sicherheitsupdates durchgeführt wurden, weisen auf einen Überlebensvorteil hin (Hazard Ratio 0,50 [0,25-0,98]). In der COMFORT II-Studie zeigte sich allerdings ein gegensätzlicher Trend: Hier waren 2 Monate nach der Primäranalyse weitere 5 Todesfälle in der Ruxolitinib-Gruppe aufgetreten, wohingegen kein weiterer Todesfall in der Placebo-Gruppe beobachtet wurde (Harrison et al., 2012). Grundsätzlich ist bei den Analysen nach der initialen, randomisierten Behandlungsphase zu beachten, dass alle Patienten im jeweiligen Vergleichsarm in den Ruxolitinib-Behandlungsarm wechseln durften. Dies kann bei der vorliegenden ITT-Auswertung zu einer möglichen Verzerrung des Behandlungseffektes führen und diese Daten sind entsprechend wenig aussagekräftig.

Zur Untersuchung der Gesamtmortalität sind weitere Langzeituntersuchungen notwendig, um verlässliche Aussagen hinsichtlich der Auswirkungen von Ruxolitinib auf die Überlebensdauer ableiten zu können.

Die kombinierten Endpunkte Leukämiefreies Überleben und Progressionsfreies Überleben wurden in der vorliegenden Bewertung als nicht valide Surrogatparameter für den patienten-relevanten Endpunkt Gesamtmortalität eingestuft. Eine valide Korrelation, die sich für die Kategorisierung eines Zusatznutzens eignet, liegt nicht vor. Ergebnisse zu beiden Endpunkten wurden nur für die Studie COMFORT II präsentiert und es konnte jeweils kein statistisch signifikanter Gruppenunterschied gezeigt werden. In der COMFORT I-Studie war das Leukämiefreie Überleben als Langzeitendpunkt mit einer Auswertung nach 144 Wochen geplant. Entsprechende Analysen liegen noch nicht vor.

Morbidität

Die in den Studien COMFORT I und COMFORT II erhobenen Endpunkte Leukämiefreies Überleben (COMFORT I und COMFORT II) und Progressionsfreies Überleben (COMFORT II)

wurden in der vorliegenden Bewertung in der Endpunktkategorie Morbidität bewertet, sind jedoch aufgrund ihrer Zusammensetzung sowie fehlender Ergebnisdarstellung zu den Einzelkomponenten in ihrer Patientenrelevanz fraglich. Es handelt sich jeweils um kombinierte Endpunkte, die sich aus Endpunkten verschiedener Endpunktkategorien (Mortalität und Morbidität) sehr unterschiedlicher Relevanz und Schwere zusammensetzen (z.B. Anstieg des Milzvolumens um 25% vs. Tod, zur Operationalisierung siehe auch Definition der Endpunkte in Tabelle 10). Die Zusammenfassung zu einem Endpunkt ist daher nicht zur Quantifizierung des Zusatznutzens geeignet. Die Erhebung des Endpunkts Leukämiefreies Überleben wurde in der COMFORT I-Studie als Langzeitendpunkt mit einer Auswertung nach 144 Wochen geplant. Entsprechende Analysen liegen noch nicht vor. Die bisher zur Auswertung des Endpunkts Leukämiefreies Überleben gemessenen Ereignisse waren Todesfälle und stellen somit keinen Erkenntnisgewinn gegenüber der Erhebung der Gesamtmortalität dar.

Neben der Auswertung des Milzvolumens als primärer Endpunkt (Rückgang um $\geq 35\%$) wurde auch die prozentuale Veränderung des Milzvolumens gegenüber Studienbeginn analysiert. Hier konnte in beiden Studien unter Ruxolitinib eine mediane Abnahme um ca. 30% gezeigt werden, wohingegen das Milzvolumen in den Komparator-Armen zunahm (je 8,5% zum Zeitpunkt der primären Analyse). Da es sich bei diesem Endpunkt jedoch – wie zuvor beschrieben – um einen Surrogatparameter unklarer Validität handelt, ist aus den gemessenen Werten nicht klar abzulesen, ob die Veränderung des Milzvolumens auch mit einer entsprechenden Änderung der belastenden Symptome (v. a. Oberbauchbeschwerden) einhergeht. Zudem ist bei diesem Endpunkt von einem hohen Verzerrungspotenzial auszugehen, da insbesondere in den Vergleichsgruppen ein nicht unerheblicher Anteil der Patienten nicht in die Auswertung eingeflossen ist.

Zur Erfassung der krankheitsbedingten Symptome wurde in der Studie COMFORT I der Total-Symptom-Score (TSS) verwendet. Es handelt sich um einen Fragebogen, der speziell die mit der Myelofibrose einhergehenden, belastenden Symptome Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl, Nachtschweiß, Juckreiz und Muskel- und Knochenschmerzen abfragt und in einem Summenscore erfasst (Range 0-60). Die Validierung des Fragebogens – insbesondere der hier angewendeten Fassung – ist nicht eindeutig nachzuvollziehen und wird auch seitens der EMA kritisch gesehen (EMA, 2012; S. 52). Der Anteil an Patienten, die nach 24 Wochen eine $\geq 50\%$ ige Verbesserung der Symptome (gewähltes Responsekriterium im Sinne einer MID) aufweist war in der Ruxolitinib-Gruppe statistisch signifikant größer als in der Placebo-Gruppe (45,9% vs. 5,3%, $p < 0,0001$) und auch bei der absoluten (medianen) Veränderung des Summenscores zeigte sich ein statistisch signifikanter Gruppenunterschied (Ruxolitinib + 6,9 Punkte, Placebo -2,0 Punkte, $p < 0,0001$). Der TSS fragt zwar relevante Symptome (z.B. Fatigue, Oberbauchbeschwerden) ab, jedoch lässt sich die klinische Relevanz einer 50%igen Verbesserung den vorliegenden Unterlagen nicht eindeutig entnehmen und ist nicht abschließend geklärt. Insgesamt kann allerdings festgehalten werden, dass es wichtig ist, die für den Patienten als belastend empfundenen Symptome der Krankheit zu erfassen und auszuwerten.

Des Weiteren wurde eine post-hoc-Analyse in der COMFORT I-Studie durchgeführt, um den Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl) zu untersuchen. Hier zeigte sich kein klarer Zusammenhang zwischen dem Rückgang des Milzvolumens um $\geq 35\%$ und den

untersuchten Oberbauchbeschwerden (Abnahme um $\geq 50\%$), was die Eignung der gewählten Responsekriterien generell infrage stellt.

Der ECOG Performance-Status wird entgegen der Einschätzung des pharmazeutischen Unternehmers der Endpunktkategorie Morbidität zugeordnet, da er als Messinstrument eher den funktionellen Status des Patienten abbildet und nicht zur Erfassung der Lebensqualität geeignet ist. Beim ECOG Performance-Status zeigte sich in beiden Studien, dass mehr Patienten ihren Performance-Status unter einer Ruxolitinib-Therapie erhalten konnten als in der jeweiligen Komparator-Gruppe und sich weniger Patienten verschlechterten. Die Daten wurden in den Studienberichten lediglich deskriptiv, ohne statistischen Vergleich dargestellt. Zu beachten ist in der COMFORT I-Studie, dass der Performance-Status zu Studienbeginn nicht gleich in beiden Gruppen verteilt war und die Patienten in der Placebo-Gruppe tendenziell einen schlechteren Performance-Status zu Studienbeginn hatten. Da allerdings die Veränderung im Studienverlauf erfasst wurde, führt dieser Gruppenunterschied wahrscheinlich nicht zu einer Überschätzung des Therapieeffektes. Auch in der Studie COMFORT II verschlechterten sich im Studienverlauf mehr Patienten in der BAT-Gruppe als in der Ruxolitinib-Gruppe. Über den gesamten Studienverlauf hielten mehr Patienten den besten Wert von 0, wenn sie mit Ruxolitinib behandelt wurden als mit der BAT-Therapie. Bei der Interpretation der Ergebnisse ist jedoch auch zu beachten, dass ein wesentlicher Teil der Patienten nicht in die Betrachtung der Veränderung des ECOG Performance-Status eingeflossen ist und daher von einem hohen Verzerrungspotenzial auszugehen ist.

Lebensqualität

Die Lebensqualität wurde in den beiden Studien mithilfe des generischen Fragebogens QLQ-C30 der EORTC bestimmt. In der COMFORT II-Studie wurde zusätzlich der FACT-Lym-Fragebogen eingesetzt, der krankheitsspezifische Änderungen der Lebensqualität erfassen soll.

Generell ist bezüglich der COMFORT II-Studie bei der Erhebung der Lebensqualität das offene Design zu kritisieren, was zu einer möglichen Verzerrung dieser subjektiv erhobenen Endpunkte führen kann, die die Ergebnisbeurteilung erschwert. Allerdings ist auch bei der COMFORT I-Studie durch das Auftreten charakteristischer Nebenwirkungen (insbesondere Thrombozytämie und Anämie) das Aufrechterhalten der Verblindung schwierig. Dies wird auch seitens der EMA kritisch gesehen (EMA, 2012; S. 60). Zudem wurden bei der Analyse der Lebensqualität nur diejenigen Patienten eingeschlossen, bei denen Werte zu Studienbeginn und zum Analysezeitpunkt vorlagen (Observed Cases); dies führt dazu, dass insbesondere in den Komparator-Armen ein nicht unerheblicher Anteil an Patienten fehlt, die vorzeitig die Studie beendet haben und gegebenenfalls in den Ruxolitinib-Arm gewechselt sind. Die Endpunkte zur Erfassung der Lebensqualität werden wegen der fehlenden / schwer zu erhaltenden Verblindung und der fehlenden ITT-Auswertung in beiden Studien alle als potenziell hoch verzerrt eingestuft.

In beiden Studien wurde der EORTC QLQ-C30 verwendet. Dieser 30 Items umfassende Fragebogen ist ein generisches Messinstrument zur Erfassung der Lebensqualität von Krebspatienten. Es ist zu beachten, dass dieses Instrument nicht speziell für Myelofibrose-Patienten validiert wurde und somit nicht spezifisch die mit der Krankheit einhergehenden Auswirkungen auf die Lebensqualität erfasst. Nichtsdestotrotz kann dieses Messinstrument ergänzend herangezogen werden. Nach 24 Wochen hatten sich in der COMFORT I-Studie in

der Ruxolitinib-Gruppe die QLQ-C30 Scores im Gesamtbild und in 4 der 5 funktionellen Skalen statistisch signifikant gegenüber Placebo verbessert (Ausnahme: kognitive Funktion). Für alle Unterskalen konnte in der Ruxolitinib-Gruppe eine Verbesserung und in der Placebo-Gruppe eine Verschlechterung gegenüber Studienbeginn gesehen werden. Auch in der COMFORT II-Studie zeigte sich eine Verbesserung im Gesamtgesundheitszustand und bei der Rollenfunktion in der Ruxolitinib-Gruppe gegenüber BAT. Der pharmazeutische Unternehmer beansprucht auch für weitere Symptome und Subskalen (Fatigue, Schmerzen, Atemnot, Schlaflosigkeit, Appetitlosigkeit, Diarrhoe) des QLQ-C30 Scores einen klinisch relevanten Vorteil für Ruxolitinib. Die Effekte werden seitens des pharmazeutische Unternehmer als klinisch relevant beschrieben, da die Gruppendifferenz an mindestens einem Messzeitpunkt > 10 Punkte betrug. Der pharmazeutische Unternehmer verweist auf die Publikation von Osoba et al. zur Herleitung dieser Relevanzschwelle: hier wurde jedoch eine Differenz von 10 Punkten auf patientenindividueller Ebene als klinisch relevanter Schwellenwert deklariert. Aussagen zur Patientenrelevanz die auf einem Gruppenunterschied auf Basis von aggregierten Patientendaten getroffen werden, sind folglich wenig aussagekräftig. Zu kritisieren sind zudem teilweise erhebliche Gruppenunterschiede zu Studienbeginn (z.B. beim Symptom Schlaflosigkeit, Fatigue), die die Beurteilung eines Therapieeffektes erschweren.

In der COMFORT II-Studie wurden darüber hinaus die Fragebögen FACT-G und FACT-LymS bzw. die daraus zusammengesetzten Skalen FACT-Lym Total-Score und FACT-TOI verwendet, um spezifischere Einflüsse der Myelofibrose-Symptomatik auf die Lebensqualität zu erfassen. Unter einer Behandlung mit Ruxolitinib verbesserten sich der mittlere Gesamtscore (FACT-Lym Total Score) und alle ausgewerteten Subskalen im Vergleich zu Studienbeginn, während er unter BAT unverändert blieb. Die Unterschiede zeigten sich ab der ersten Erhebung in Woche 8 bis zur Woche 48. Die angegebenen minimalen klinisch relevanten Differenzen (s. Tabelle 9) werden zwar bei nahezu allen Auswertungszeitpunkten und Subskalen erreicht, allerdings sind die methodischen Angaben zur Ermittlung der MIDs nicht ausführlich beschrieben und somit nicht abschließend zu beurteilen. Darüber hinaus ist zu kritisieren, dass dieser spezifischere Fragebogen nur in der unverblindeten COMFORT II-Studie eingesetzt wurde und die Ergebnisse dieses subjektiv erhobenen Endpunkts daher potenziell hoch verzerrt sind.

Insgesamt zeigen die Instrumente zur Erfassung der Lebensqualität eine Verbesserung der Lebensqualität unter Ruxolitinib gegenüber dem jeweiligen Komparator. Kritikpunkte sind die fehlende Verblindung in COMFORT II, die nicht durchgeführte ITT-Auswertung und die fehlende Krankheitsspezifität der QLQ-C30 der EORTC, die auch die EMA im EPAR bemängelt (EPAR, S. 52 ff).

Nebenwirkungen

Es liegen deskriptive Angaben zu der Anzahl der Patienten, bei denen mindestens ein unerwünschtes Ereignis aufgetreten ist, vor. Ein statistischer Gruppenvergleich wurde nicht durchgeführt. Die unerwünschten Ereignisse wurden unter Verwendung des MedDRA (Version 12) erhoben und klassifiziert. Es kann insgesamt von einem niedrigen Verzerrungspotenzial bei der Darstellung der unerwünschten Ereignisse ausgegangen werden. In der Bewertung wurden die unerwünschten Ereignisse unabhängig von ihrem vermuteten Kausalzusammenhang mit der Medikation dargestellt. Insbesondere das offene Design der Studie COMFORT II hätte bei Darstellung der unerwünschten Ereignisse, bei denen ein

Kausalzusammenhang mit der Medikation vermutet wird, zu einer Verzerrung des Behandlungseffektes geführt.

In beiden Studien zeigte sich ein hoher Anteil an Patienten mit unerwünschten Ereignissen in jeweils beiden Studienarmen (> 90 %), die sich teilweise auch durch die mit der Krankheit einhergehenden Symptome ergibt. Schwerwiegende unerwünschte Ereignisse traten in den Studien bei 27,7 % bis 35,1 % der Patienten auf, wobei ein formaler statistischer Vergleich weder im Dossier noch in den Studienberichten durchgeführt wurde. Es kann allerdings festgehalten werden, dass sich diese Nebenwirkungen in den jeweiligen Studienarmen nicht wesentlich voneinander unterscheiden. Vergleichbare Ergebnisse zwischen den Studienarmen ergaben sich auch bei dem Endpunkt „Abbruch aufgrund unerwünschter Ereignisse“ (COMFORT I je ca. 11 %; COMFORT II je ca. 8 %).

Ein unterschiedliches Ergebnis zeigte sich jedoch in den beiden Studien hinsichtlich des Endpunkts „Unerwünschte Ereignisse der CTCAE Grade 3 und 4“. Während es in der COMFORT I-Studie keinen bedeutsamen Unterschied zwischen den Behandlungsarmen gab (47,1 % bzw. 44,4 %) zeigte sich in der COMFORT II-Studie eine auffällige Differenz zuungunsten von Ruxolitinib (Ruxolitinib-Gruppe: 43,2 % vs. BAT-Gruppe: 24,7%). Weitere Gruppenunterschiede zuungunsten von Ruxolitinib zeigten sich in beiden Studien für den Endpunkt „Unerwünschte Ereignisse, die zu einer Dosisreduktion führten“ (COMFORT I: 40 % vs. 9,3 %; COMFORT II: 63 % vs. 15,1 %). Dies spricht für eine insgesamt schlechtere Verträglichkeit von Ruxolitinib gegenüber dem jeweiligen Komparator-Arm.

In beiden Studien waren Anämie und Thrombozytopenie die beiden häufigsten hämatologischen unerwünschten Ereignisse (alle Grade und Grad 3 und 4) und traten bei einem deutlich höheren Anteil an Patienten in der jeweiligen Interventionsgruppe auf. Diese Nebenwirkungen lassen sich auf den Wirkmechanismus von Ruxolitinib als JAK-Inhibitor zurückführen. Thrombozytämie und Anämie haben allerdings nur selten zu einem Studienabbruch geführt (< 1 %) und ließen sich gemäß den Angaben des pharmazeutischen Unternehmers in den Studien durch Dosisreduktionen und Bluttransfusionen gut beherrschen. Allerdings ist hierbei auch zu beachten, dass diese Nebenwirkungen für den Patienten durchaus belastend sind und die damit einhergehenden Bluttransfusionen auch mit Nebenwirkungen und Beeinträchtigungen verbunden sind. Die Patienten sollten – wie in der Fachinformation beschrieben – hinsichtlich der hämatologischen Parameter engmaschig überwacht werden. Der weitaus überwiegende Teil der häufigsten Nebenwirkungen war von milder bis moderater Natur und trat in beiden Gruppen jeweils vergleichbar häufig auf (s. Datenextraktionsbögen).

Als weitere Nebenwirkungen von besonderem Interesse wurden Blutungen und Infektionen identifiziert. Ergänzend zu den Auswertungen in den Studienberichten wird an dieser Stelle auf die Auswertungen der EMA in ihrem öffentlichen Beurteilungsbericht verwiesen (European Public Assessment Report, EPAR, EMA, 2012). Im EPAR findet sich eine gepoolte Auswertung der beiden Zulassungsstudien (Tabelle 17). Bei der Interpretation der Tabelle ist zu berücksichtigen, dass lediglich die unerwünschten Ereignisse in der Ruxolitinib-Gruppe – ohne einen direkten Gruppenvergleich – aufgezeigt werden (Angaben zum Gruppenvergleich finden sich teilweise ergänzend im EPAR). Es ist nicht eindeutig nachzuvollziehen, welche Preferred-Terms der MedDRA jeweils in die Auswertung eingeflossen sind.

Infektionen (SOC „Infections and Infestations“) kamen in der COMFORT II-Studie in der Ruxolitinib-Gruppe häufiger als im Vergleichsarm vor (63,7 % vs. 42,5 %). In der COMFORT I-

Studie wurden dagegen weniger Infektionen unter Ruxolitinib als unter Placebo beobachtet (38,1 % vs. 41,7 %). Mit einer gepoolten Häufigkeit von 12,3 % kommen Infektionen des Urogenitaltraktes sehr häufig in beiden Studien unter der Behandlung mit Ruxolitinib vor. Des Weiteren wurde eine Herpes-Zoster-Infektion häufiger unter Ruxolitinib beobachtet (gepoolte Auswertung: 3,3 % vs. 0,45 %). Ein vermehrtes Auftreten von Infektionen lässt sich durch den Wirkmechanismus der Januskinase-Inhibition erklären und sollte bei der Anwendung von Ruxolitinib hinsichtlich des Risikos der Entwicklung schwerwiegender (myko)bakterieller, fungaler und viraler Infektionen beobachtet werden.

Blutungen werden als gruppierte Auswertung im EPAR besonders erwähnt (EMA, 2012, S. 65). Es kam studienübergreifend bei 32,6 % der Patienten unter Ruxolitinib-Behandlung zu einem Blutungsereignis gegenüber 23,4 % der Patienten in den gepoolten Vergleichsarmen (Placebo oder BAT). Schwerwiegende Blutungen traten in den Behandlungsarmen vergleichbar häufig auf (4,7 % vs. 3,1 %). Es kam zudem zu drei tödlich verlaufenden Blutungen in den Ruxolitinib-Armen und zu einem blutungsbedingten Todesfall in den Vergleichsarmen. Die Mehrzahl der Patienten mit Blutungen während der Behandlung berichtete über Blutergüsse (21,3 % vs. 11,6 %).

Das Auftreten von Infektionen und Blutungsereignissen ist in zukünftigen Studien weiter zu beobachten und genauer zu untersuchen.

Tabelle 16: Ergänzende Auswertung der Nebenwirkungen unter Ruxolitinib durch die EMA, gepoolte Analyse aus COMFORT I / II (Quelle: EPAR, S. 63)

Adverse drug reaction	Ruxolitinib – myelofibrosis patients N=301*		
	All CTCAE grades ^c (%)	CTCAE grade 3/4 ^c (%)	Frequency category
Infections and infestations			
Urinary tract infections ^{a,d}	12.3	1.0	Very common
Herpes zoster ^{a,d}	4.3	0.3	Common
Blood and lymphatic system disorders^{b,d}			
Anaemia	82.4	42.5	Very common
Thrombocytopenia	69.8	11.3	Very common
Neutropenia	15.6	6.6	Very common
Bleeding (any bleeding including intracranial, and gastrointestinal bleeding, bruising and other bleeding)	32.6	4.7	Very common
Intracranial bleeding	1.0	1.0	Common
Gastrointestinal bleeding	5.0	1.3	Common
Bruising	21.3	0.3	Very common
Other bleeding (including epistaxis, post-procedural haemorrhage and hematuria)	13.3	2.3	Very Common
Metabolism and nutrition disorders			
Weight gain ^a	10.0	1.3	Very common
Hypercholesterolaemia ^b	16.6	0	Very common
Nervous system disorders			
Dizziness ^a	15.0	0.3	Very common
Headache ^a	13.9	0.5	Very common
Gastrointestinal disorders			
Flatulence ^a	2.9	0	Common
Hepatobiliary disorders			
Raised alanine aminotransferase ^b	26.9	1.3	Very common
Raised aspartate aminotransferase ^b	19.3	0	Very common
* Myelofibrosis patients randomised to and treated with ruxolitinib from the phase 3 pivotal COMFORT-I and COMFORT-II studies ^a Frequency is based on adverse event data. A subject with multiple occurrence of an adverse drug reaction (ADR) is counted only once in that ADR category. ADRs reported are on treatment or up to 28 days post treatment end date. ^b Frequency is based on laboratory values. A subject with multiple occurrences of an ADR is counted only once in that ADR category. ADRs reported are on treatment or up to 28 days post treatment end date. ^c Common Terminology Criteria for Adverse Events (CTCAE) version 3.0; grade 1 = mild, grade 2 = moderate, grade 3 = severe, grade 4 = life-threatening ^d These ADRs are discussed in the text.			

Insgesamt bleibt bei der Betrachtung der Nebenwirkungen festzuhalten, dass es sich in beiden Studien um einen relativ kurzen Beobachtungszeitraum handelt und dass die Datenlage zur Sicherheit von Ruxolitinib noch durch weitere Untersuchungen verbessert werden muss. Dies ist insbesondere vor dem Hintergrund wichtig, dass es sich bei Ruxolitinib um einen Vertreter einer neuen Wirkstoffklasse handelt, zu der noch keine vergleichbaren Langzeitsicherheitsdaten zur Verfügung stehen. Da zum Zeitpunkt der primären Analyse noch ca. 75% der mit Ruxolitinib behandelten Patienten in der Gruppe waren, zu der sie randomisiert wurden, ist in der Praxis mit einem längeren Einsatz der Substanz zu rechnen, der (deutlich) über die in den Studien untersuchten 24 bzw. 48 Wochen hinausgeht. Dieser Aspekt wird auch explizit im EPAR kritisiert und entsprechende Langzeitdaten werden gefordert (EMA, 2012). Die EMA verweist in diesem Zusammenhang auf die Extensionsphasen der beiden Zulassungsstudien und auf eine Langzeit-Sicherheitsstudie (CINC424AIC01T, Post Authorization Safety Study), die bereits durchgeführt wird und deren Ergebnisse im Februar 2019 erwartet werden. Insbesondere Untersuchungen hinsichtlich der Auswirkungen einer Therapie mit Ruxolitinib auf die Mortalität sind hier erforderlichlich.

Subgruppenergebnisse

Die Ergebnisse aus den Subgruppenanalysen in den Studien COMFORT I und COMFORT II werden im Dossier und in den Studienberichten lediglich deskriptiv dargestellt (siehe jeweiliger Datenextraktionsbogen zur Darstellung der durchgeführten Subgruppenanalysen). A priori geplant wurden Subgruppenanalysen jeweils nur für den primären Endpunkt der Milzvolumenreduktion um $\geq 35\%$. Da in beiden Studien im Kontrollarm kein ausreichendes klinisches Ansprechen erreicht wurde, konnten die Subgruppenanalysen nur innerhalb des Ruxolitinib-Arms durchgeführt werden. Als Folge fehlender Interaktionstests im Dossier und in den Studienberichten können lediglich auffällige Unterschiede beschrieben und diskutiert werden, ohne allerdings konkrete Schlüsse ableiten zu können.

Die Subgruppenergebnisse für den primären Endpunkt deuten darauf hin, dass Patienten mit JAK2V617F-Mutation studienübergreifend mehr von einer Ruxolitinib-Behandlung profitieren als Patienten ohne diese Mutation. Die Subgruppenanalyse in der COMFORT I-Studie lässt auch ein besseres Ansprechen von Frauen gegenüber Männern unter Ruxolitinib vermuten; ein Trend, der sich abgeschwächt auch in der COMFORT II-Studie zeigt. Auch eine höhere Thrombozytenzahl (entsprechend einer höheren Ruxolitinib-Anfangsdosierung) scheint (in beiden Studien) mit einem besseren Ansprechen einherzugehen. In der COMFORT II-Studie zeigte sich auch ein deutlicher Unterschied hinsichtlich des Ansprechens auf den primären Endpunkt in Abhängigkeit von der Indikation (PPV-MF: 41 %, PMF: 18,4 %, PET-MF: 35 %). Dieser Trend konnte in der COMFORT I-Studie allerdings nicht gesehen werden.

Insgesamt kommt die EMA zu dem Schluss, dass bei Patienten mit V617F-Mutation der Nutzen von Ruxolitinib den Schaden deutlich überwiegt. Auch bei V617F-negativen Patienten sieht die EMA noch ein positives Nutzen-Risiko-Verhältnis, wenn auch insgesamt ein kleinerer Therapieeffekt (zumindest hinsichtlich des primären Endpunkts) zugunsten von Ruxolitinib in beiden Studien gezeigt werden konnte (EMA; 2012).

3.3. Zusammenfassende Darstellung der Studienergebnisse

Die Resultate aus den beiden für die Bewertung herangezogenen Zulassungsstudien COMFORT I und COFORT II werden in der nachfolgenden Tabelle – unter Berücksichtigung des Verzerrungspotenzials – deskriptiv zusammengefasst.

Es wird dargestellt, ob sich unter Ruxolitinib in den Studien ein positiver Effekt gezeigt hat bei hohem (↑) bzw. niedrigem (⇕) Verzerrungspotenzial, ob sich ein negativer Effekt gezeigt hat bei hohem (↓) bzw. niedrigem (⇓) Verzerrungspotenzial, oder ob kein Unterschied nachgewiesen werden konnte bei hohem (↔) bzw. niedrigem Verzerrungspotenzial (⇔). Bei den Nebenwirkungen werden die absoluten Häufigkeiten zum deskriptiven Gruppenvergleich dargestellt.

	COMFORT I Ruxolitinib vs. Placebo (nach 24 Wochen)	COMFORT II Ruxolitinib vs. BAT (nach 48 Wochen)
Mortalität		
Gesamtüberleben	↔	↔
Morbidität		
Leukämiefreies / Progressionsfreies Überleben ¹	- ²⁾	↔
Milzvolumenreduktion (≥ 35%) ³	⇕	⇕
Symptomreduktion um ≥ 50% (gemessen mit dem TSS) ⁴	⇕	- ²⁾
ECOG Performance-Status	↑ ⁵⁾	↑ ⁵⁾
Gesundheitsbezogene Lebensqualität		
QLQ-C30 der EORTC	↑	↑ ⁵⁾
FACT-Lym	- ²⁾	↑ ⁵⁾
Nebenwirkungen (Angabe der Häufigkeiten [%], Ruxolitinib vs. Placebo bzw. BAT)		
UE	97,4 vs. 98,0	99,3 vs. 90,4
SUE	27,7 vs. 35,1	30,1 vs. 28,8
UE der CTCAE Grad 3 / 4	47,1 vs. 44,4	43,2 vs. 24,7
Abbruch aufgrund von UE	11,0 vs. 10,6	8,2 vs. 8,2
UE, die zur Dosisreduktion führten	40,0 vs. 9,3	63,0 vs. 15,1
Thrombozytopenie (alle Grade)	34,2 vs. 9,3	44,5 vs. 9,6
Anämie (alle Grade)	31,0 vs. 13,9	41,1 vs. 12,3
Blutungen (Ekchymosen)	18,7 vs. 9,3	1,4 vs. 0
Infektionen	38,1 vs. 42,4	63,7 vs. 43,8

1) Diese Endpunkte wurden in der vorliegenden Bewertung in der Endpunktkategorie Morbidität bewertet, sind jedoch aufgrund ihrer Zusammensetzung sowie fehlender Ergebnisdarstellung zu den Einzelkomponenten in ihrer Patientenrelevanz fraglich.

2) Diese Endpunkte wurden in der Studie nicht erhoben.

3) Dieser Endpunkt war in beiden Studien der primäre Endpunkt.

4) Dieser Summenscore (TSS) fasst folgende Symptome zusammen: Oberbauchbeschwerden, Schmerzen unter den Rippen, Völlegefühl, Nachtschweiß, Juckreiz und Muskel- und Knochenschmerzen.

5) Bei diesen Endpunkten wurden zwar Effekte zugunsten von Ruxolitinib beobachtet, es wurde jedoch keine statistisch signifikante oder (methodisch adäquat begründete) klinisch relevante Überlegenheit von Ruxolitinib nachgewiesen.

BAT: Best-Available-Therapy, SUE: Schwere unerwünschte Ereignisse, TSS: Total-Symptom-Score,

UE: Unerwünschte Ereignisse

4. Anforderungen an eine qualitätsgesicherte Anwendung

Die Vorgaben der Fachinformation sind zu berücksichtigen.

Vor Beginn der Therapie mit Ruxolitinib und zur Dosiseinstellung ist ein großes Blutbild, einschließlich eines Differentialblutbildes der weißen Blutkörperchen, durchzuführen und danach alle 2 bis 4 Wochen zur Kontrolle der hämatologischen Parameter (insbesondere der Thrombozytenzahl, des Hämoglobinwertes und der Zahl der neutrophilen Granulozyten), oder öfter, wenn klinisch erforderlich.

Die Patienten sind über das erhöhte Risiko für Blutungen und für Infektionen zu informieren und sorgfältig auf Anzeichen von Blutungen oder von Infektionen hin zu beobachten.

Die Behandlung ist nach 6 Monaten zu beenden, falls es zu keiner Reduktion der Milzgröße oder Verbesserung der Symptome seit dem Beginn der Therapie gekommen ist. Für Patienten, die eine gewisse Verbesserung der klinischen Symptomatik zeigen, wird empfohlen, die Therapie mit Ruxolitinib zu beenden, wenn sie eine Vergrößerung der Milzlänge von 40% im Vergleich zur Ausgangsgröße (entspricht ungefähr einer 25%igen Zunahme des Milzvolumens) erleiden und nicht länger eine spürbare Verbesserung der krankheitsbedingten Symptome festzustellen ist.

Referenzen

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85 (5): 365-76.

Abdel-Wahab OI, Levine RL. Primary myelofibrosis: update on definition, pathogenesis, and treatment. *Annu Rev Med* 2009; 60: 233-45.

Barbui T, Barosi G, Birgegard G, Cervantes F, Finazzi G, Griesshammer M, et al. Philadelphia-negative classical myeloproliferative neoplasms: critical concepts and management recommendations from European LeukemiaNet. *J Clin Oncol* 2011; 29 (6): 761-70.

Barosi G, Bordessoule D, Briere J, Cervantes F, Demory JL, Dupriez B, et al. Response criteria for myelofibrosis with myeloid metaplasia: results of an initiative of the European Myelofibrosis Network (EUMNET). *Blood* 2005; 106 (8): 2849-53.

Carter GC, Liepa AM, Zimmermann AH, Morschhauser F. Validation of the functional Assessment of Cancer Therapy-Lymphoma (FACT-LYM) in Patients with Relapsed/Refractory Mantle Cell Lymphoma. *ASH Annual Meeting Abstracts* 2008; 112: 2376.

Cella D, Webster K, Cashy J, Kutikova L, Burgess MF, Lin BK, et.al. Development of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma Clinical Research: The Functional Assessment of Cancer Therapy - Lymphoma (FACT-Lym). *ASH Annual Meeting Abstracts* 2005; 106: 750.

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993 Mar; 11 (3): 570-9.

Cervantes F, Dupriez B, Pereira A, Passamonti F, Reilly JT, Morra E, et al. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 2009; 113 (13): 2895-901.

Cervantes F. How I treat splenomegaly in myelofibrosis. *Blood Cancer J* 2011; 1 (e37): 1-6.

European Medicines Agency (EMA). CHMP assessment report (EPAR): Ruxolitinib (Jakavi). Stand Oktober 2012. London: EMA, 2012. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002464/WC500133226.pdf. Zugriff am 30.11.2012.

Fachinformation JAKAVI, Stand: August 2012.

Ghoreschi K, Laurence A, O'Shea JJ. Janus kinases in immune cell signaling. *Immunol Rev* 2009; 228 (1): 273-87.

Gräfe KA und Siebenand S. Neu auf dem Markt - Perampanel und Ruxolitinib. Pharmazeutische Zeitung 2012; 157 (40). <http://www.pharmazeutische-zeitung.de/index.php?id=43647>. Zugriff am: 30.11.2012.

Griesshammer M, Petrides PE, Lengfelder E. DGHO Leitlinie: Primäre Myelofibrose (OMF) Leitlinie. Empfehlungen der Fachgesellschaft zur Diagnostik und Therapie hämatologischer und onkologischer Erkrankungen. Stand März 2010. Berlin: DGHO, 2010.

Harrison C, Kiladjan JJ, Al-Ali HK, Gisslinger H, Waltzman R, Stalbovskaya V, et al. JAK inhibition with ruxolitinib versus Best-Available-Therapy for myelofibrosis. N Engl J Med 2012; 366 (9): 787-98.

Martinez-Trillos A, Gaya A, Maffioli M, Arellano-Rodrigo E, Calvo X, Diaz-Beya M, et al. Efficacy and tolerability of hydroxyurea in the treatment of the hyperproliferative manifestations of myelofibrosis: results in 40 patients. Ann Hematol 2010; 89 (12): 1233-37.

Mesa R, Gotlib J, Gupta V, DiPersio JF, Catalano J, Deininger MW, et al. Associations Between Improvements in Myelofibrosis (MF) Symptoms and Quality of Life Measures With Splenomegaly Reduction in COMFORT-I: A Randomized, Double-Blind, Phase III Trial of the JAK1 and JAK2 Inhibitor Ruxolitinib Versus Placebo in Patients With MF. ASH Annual Meeting Abstracts 2011; 118: 3842.

Mesa RA, Kantarjian H, Tefferi A, Dueck A, Levy R, Vaddi K, et al. Evaluating the serial use of the Myelofibrosis Symptom Assessment Form for measuring symptomatic improvement: performance in 87 myelofibrosis patients on a JAK1 and JAK2 inhibitor (INCB018424) clinical trial. Cancer 2011; 117 (21): 4869-77.

Mesa RA, Niblack J, Wadleigh M, Verstovsek S, Camoriano J, Barnes S, et al. The burden of fatigue and quality of life in myeloproliferative disorders (MPDs): an international Internet-based survey of 1179 MPD patients. Cancer 2007; 109 (1): 68-76.

Mesa RA, Schwager S, Radia D, Cheville A, Hussein K, Niblack J, et al. The Myelofibrosis Symptom Assessment Form (MFSAF): an evidence-based brief inventory to measure quality of life and symptomatic response to treatment in myelofibrosis. Leuk Res 2009; 33 (9): 1199-203.

Novartis Pharma. Studienprotokoll und Studienbericht COMFORT I (INCB18424-351): A Randomized, Double-Blind, Placebo-Controlled Study of the JAK Inhibitor INCB018424 Tablets Administered Orally to Subjects With Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF) or Post-Essential Thrombocythemia-Myelo-fibrosis (PET-MF). 2011.

Novartis Pharma. Studienprotokoll und Studienbericht COMFORT II (CINC424A2352): A Randomized Study of the JAK Inhibitor INCB018424 Tablets Compared to Best-Available-Therapy in Subjects with Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF) or Post-Essential Thrombocythemia Myelofibrosis (PET-MF). 2011.

Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982; 5 (6): 649-55.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; 16 (1): 139-44.

Sirhan S, Lasho TL, Hanson CA, Mesa RA, Pardanani A, Tefferi A. The presence of JAK2V617F in primary myelofibrosis or its allele burden in polycythemia vera predicts chemosensitivity to hydroxyurea. *Am J Hematol* 2008; 83 (5): 363-5.

Tefferi A, Barosi G, Mesa RA, Cervantes F, Deeg HJ, Reilly JT, et al. International Working Group (IWG) consensus criteria for treatment response in myelofibrosis with myeloid metaplasia, for the IWG for Myelofibrosis Research and Treatment (IWG-MRT). *Blood* 2006; 108 (5): 1497-503.

Tefferi A, Vainchenker W. Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol* 2011; 29 (5): 573-82.

Verstovsek S, Kantarjian H, Mesa RA, Pardanani AD, Cortes-Franco J, Thomas DA, et al. Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N Engl J Med* 2010; 363 (12): 1117-27.

Verstovsek S, Mesa RA, Gotlib J, Levy RS, Gupta V, DiPersio JF, et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. *N Engl J Med* 2012; 366 (9): 799-807.

Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003; 1: 79.

Anhang

i. Datenextraktionsbogen der pivotalen Studie COMFORT I (INCB 18242-351)



Indikation: Myelofibrose

Indikation: Myelofibrose

Datenextraktion und Bewertung der Studie

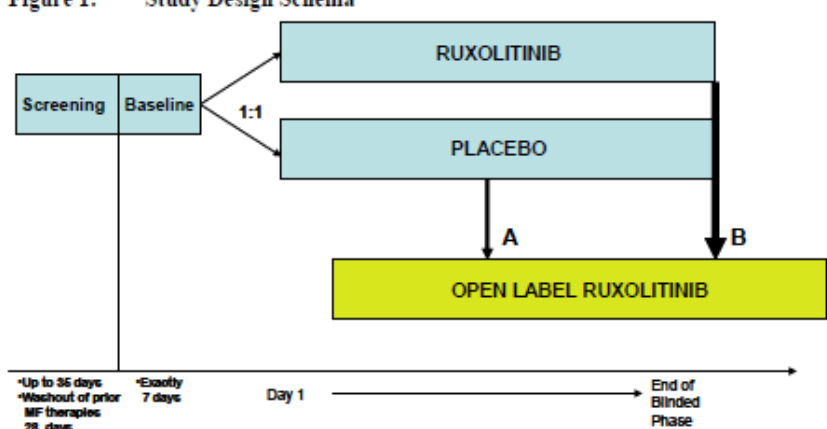
Datenextraktion und Bewertung der Studie

INCB 18424-351 (COMFORT I)

Nr.	Feld	Allgemeine Angaben
1	Quelle	Studienbericht inkl Studienprotokoll (Protocol No. INCB 18424-351) : "A Randomized, Double-Blind, Placebo-Controlled Study of the JAK Inhibitor INCB018424 Tablets Administered Orally to Subjects With Primary Myelofibrosis (PMF), Post-Polycythemia-vera-Myelofibrosis (PPV-MF), Or Post-Essential Thrombocythemia-Myelofibrosis (PET-MF)" (34) Primärpublikation, Verstoveck et al., 2012 (15) European Public Assessment Report: Ruxolitinib Jakavi (8) Herstellerdossier zu Ruxolitinib
2	Studien- typ	Zuordnung zu einem der folgenden Studientypen (bitte ankreuzen): ☒ Therapiestudie mit randomisierter Vergleichsgruppe ☐ Therapiestudie mit nicht-randomisierter Vergleichsgruppe ☐ Therapiestudie mit Vergleichen aus früherer Zeit und / oder anderem Ort (z. B. historische Kontrollen) ☐ Fall-Kontroll-Studie ☐ Therapiestudie ohne (unabhängige) Vergleichsgruppe (inkl. „Vorher-Nachher-Studien“) ☐ Fallserie ☐ Fallbericht / Kasuistik (case report) ☐ Nicht eindeutig zuzuordnen (ggf. Beschreibung mittels Originaltext): Details zum Studiendesign: Es handelt sich bei der COMFORT I-Studie um eine parallele, doppelblinde, Placebo-kontrollierte und randomisierte Studie im Verhältnis 1:1 zum Vergleich der Wirksamkeit und Sicherheit von Ruxolitinib gegenüber Placebo. Die Erhebung des primären Endpunkts („Primäranalyse“) wurde nach 24 Wochen durchgeführt.
3	Einord- nung in die Evidenz- kategorie gemäß VerfO	Hier folgt eine formale Zuordnung zu den Evidenzstufen nach Verfahrensordnung (bitte ankreuzen): ☒ Ib: Randomisierte klinische Studien ☐ IIb: Prospektiv vergleichende Kohortenstudien ☐ III: Retrospektiv vergleichende Studien ☐ IV: Fallserien und nicht-vergleichende Studien ☐ V: Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte u. a.; nicht mit Studien belegte Meinungen anerkannter Experten, Bericht von Expertenkomitees und Konsenskonferenzen.
4	Kontext der Studie / Interes- sen- konflikte	Sponsor: Novartis Pharma GmbH
5	Indikation	Ruxolitinib ist angezeigt für die Behandlung von krankheitsbedingter Splenomegalie oder Symptomen bei Erwachsenen mit primärer Myelofibrose (auch bekannt als chronische idiopathische Myelofibrose), Post-Polycythaemia-vera-Myelofibrose oder Post-Essentieller-Thrombozythämie-Myelofibrose.

6	Ziel- setzung	Primäres Studienziel: • Untersuchung der Wirksamkeit, Sicherheit und Verträglichkeit von Ruxolitinib bei zweimal täglicher Gabe im Vergleich zu Placebo bei Patienten mit PMF, PPV-MF und PET-MF. Sekundäre Studienziele: • Untersuchung der Auswirkung von Ruxolitinib auf die MF-Symptomatik aus Sicht der Patienten • Populations-Pharmakokinetik von Ruxolitinib
Methodik		
7	Ein- und Aus- schluss- kriterien	Key inclusion criteria 1. 18 through 85 years of age. 2. Life expectancy of at least 6 months. 3. Diagnosed with PMF, PPV-MF or PET-MF, according to the 2008 WHO criteria irrespective of JAK2 mutation status. 4. Subjects with myelofibrosis requiring therapy must be classified as high risk (3 or more prognostic factors) OR intermediate risk level 2 (2 or more prognostic factors) if they have also been diagnosed with PMF, PPV-MF, or PET-MF at least 24 months prior to study entry. 5. ECOG Performance-Status of 0, 1, 2 or 3. 6. Must have palpable spleen measuring 5 cm or greater below the costal margin. 7. Have adequate bone marrow reserve as demonstrated by: ANC > 1000/μL and platelet count > 100,000/μL without the assistance of growth factors, thrombopoietic factors or platelet transfusions. Subjects may not have received growth factors for at least one month prior to receiving the first dose of study medication. 8. Total bilirubin $\leq$ 2.0 mg/dL; Alanine aminotransferase (ALT) $\leq$ 2.5x institutional ULN; Creatinine $\leq$ 2.0 mg/dL. Key exclusion criteria 1. Females who are pregnant or are currently breastfeeding. 2. Subjects with clinically significant bacterial or viral infection which require therapy 3. Subjects with NYHA Class III or IV impairments. 4. Treatment within the last 12 weeks prior to Screening with androgens to treat any MF symptoms. 5. Subjects with a known history of platelet counts < 50,000/μL or ANC < 500/μL except during treatment with cytotoxic therapy or irradiation. 6. Subjects who have had splenic irradiation within 3 months prior to Screening. 7. Subjects with cardiac disease which in the Investigator's opinion may jeopardize the safety of the subject or the compliance with the protocol.
8	Prüfinter- vention	Ruxolitinib 2 x täglich oral. Die Dosierung war abhängig von der Anzahl der Blutplättchen zu Studienbeginn. • Patienten mit einer Thrombozytenzahl von > 200,000/μL fingen mit einer Dosis von 2 x täglich 20 mg (je 4 Tabletten) an • Patienten mit einer Thrombozytenzahl von 100.000-200.000/μL fingen mit einer Dosis von 2 x täglich 15 mg (je 3 Tabletten) an

		Dosisanpassungen waren je nach Wirksamkeit und Verträglichkeit erlaubt: <u>Dosiserhöhung um 5 mg 2 x täglich (Maximaldosis: 25 mg 2 x täglich)</u> • bei unzureichender Wirksamkeit nach 4 Wochen (weniger als 40% Milzlängenabnahme) und • Thrombozytenzahl > 150.000/μl seit Studienbeginn und • Neutrophilenzahl ≥ 1000 μl seit Studienbeginn <u>Dosissenkung</u> • bei Blutungsneigung • Bei Neutrophilenzahl < 500/μl: Therapiepause • Bei Thrombozytenzahl < 50.000/μl
9	Ver- gleichs- inter- ven- tion(en)	Placebo 2 x täglich 3 oder 4 Tabletten
10	Prüfzen- tren	68 Zentren in den USA, 6 Zentren in Kanada und 15 Zentren in Australien.
11	Randomi- sierung	Die Randomisierungsliste wurde im Vorfeld der Studie automatisiert durch ein Computersystem mit Zufallsgenerator erstellt. Nach dem Screening wurden geeignete und willige Patienten durch den Prüfarzt aufgenommen. Die Zuteilung erfolgte durch den Anruf des Prüfarztes beim interaktiven Spracherkennungssystem zentral. Randomisierung 1:1 (keine Stratifizierung)
12	Conceal- ment („Maskie- rung“ der Randomi- sierung)	Die Patienten wurden zentral durch ein interaktives Spracherkennungssystem randomisiert, dass lediglich eine Nummer ausgab, die der Prüfarzt in den elektronischen Erfassungsbogen eintrug und mit der die Medikationsflasche für den Patienten gekennzeichnet war. Die behandelnden Ärzte hatten keinen Zugang zur Randomisierungsliste. Verum und Placebo waren äußerlich identisch. Zudem erfolgte eine zentral verblindete Auswertung der MRT-Bilder als Grundlage für den primären Studienendpunkt „Milzvolumenreduktion“. Dabei wurden die Aufnahmen eines Patienten immer vom selben Untersucher beurteilt, welchem nur eine Patientennummer vorlag.
13	Verblin- dung der Behand- lung	Erfolgte eine Verblindung der Behandlung (bitte ankreuzen)? ☐ Nein, offene Behandlung ☒ Patienten verblindet ☒ Behandler verblindet ☒ Patienten und Behandler verblindet* ☒ Auswerter verblindet ☐ Verblindung unklar * bis zur Analyse der primären Daten nach 24 Wochen; danach erfolgte die Entblindung und ein Crossover von Placebo zu Ruxolitinib war möglich
14	Beobacht- ungs- dauer	• 24.08.2009: erster Patient eingeschlossen • 02.11.2010: cutoff-Date (last patient completed war “ongoing”) <i>“When ALL subjects had completed Week 24 (or discontinued) and half of the subjects remaining in the study had completed Week 36, the database was frozen, primary analyses were conducted, and the study would be unblinded so that all subjects could continue on open-</i>

		<i>label ruxolitinib."</i> Figure 1: Study Design Schema  A: Subjects meeting protocol-defined criteria could be unblinded and if found to be randomized to placebo, could cross over to receive open-label ruxolitinib prior to primary analysis (early crossover). B: When ALL subjects had completed Week 24 (or discontinued) and half of the subjects remaining in the study had completed Week 36, the database was frozen, primary analyses were conducted, and the study would be unblinded so that all subjects could continue on open-label ruxolitinib.
15	Erhebung der primären Zielkriterien	<u>Primäres Zielkriterium:</u> Anzahl der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ nach 24 Wochen in MRT oder CT
16	Erhebung der sekundären Zielkriterien	<u>Weitere Zielkriterien (die in der Bewertung berücksichtigt wurden):</u> • Gesamtüberleben • Prozentuale Veränderung des Milzvolumens gegenüber Studienbeginn • Anteil der Patienten mit $\geq 50\%$ Rückgang im Total-Symptom-Score (TSS) nach 24 Wochen • (Milzvolumenreduktion [Beginn / Dauer des Ansprechens]) • Absolute / prozentuale Veränderung im Total-Symptom-Score TSS gegenüber Studienbeginn nach 24 Wochen • Veränderung der Einzelsymptome des MFSAFv2.0 • Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden • ECOG Performance-Status • QLQ-C30 der EORTC • Unerwünschte Ereignisse • Schwerwiegende unerwünschte Ereignisse • Unerwünschte Ereignisse Grad 3-4 der CTCAE • Abbruch aufgrund von unerwünschten Ereignissen • Dosisreduktion aufgrund von unerwünschten Ereignissen • Häufigste (schwere) unerwünschte Ereignisse (gesonderte Betrachtung der UEs Thrombozytämie, Anämie, Blutungen und Infektionen als UEs von besonderem Interesse) <u>Weitere Zielkriterien (die nicht in der Bewertung berücksichtigt wurden, allerdings im Datenextraktionsbogen dargestellt werden, siehe dazu auch Tabelle 6):</u> • Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$

		% im Zeitverlauf • Notwendigkeit von Erythrozytenkonzentraten • Leukämische Transformation • Anzahl der Patienten mit einer tastbaren Milzgrößenreduktion um > 50% • PROMIS Fatigue-Scale • Patient Global Impression of Change • JAK2-Allellast <u>Hinweis: Die Ergebnisse der folgenden Zielkriterien werden im Datenextraktionsbogen in Feld 20 nicht dargestellt (siehe dazu auch Tabelle 6)</u> • Fibrosegrad • Plasma PD-Marker • Weitere Operationalisierungen des Total-Symptom-Score • Körpergewicht • Weiter Auswertungen der Kategorie Nebenwirkungen • Laborparameter • Vitalzeichen
17	Anzahl der einzuschließenden Patienten pro Gruppe	Fallzahl- / Powerberechnung durchgeführt? ☒ ja ☐ nein ☐ keine Angabe Angaben aus dem statistischen Analyseplan (Abschnitt 3.2): „Based on the data from ongoing Study INCB 18424-251, it was assumed that at least 30% of the active subjects would achieve a $\geq 35\%$ reduction from Baseline to Week 24, and that rate for the placebo subjects would be no more than 10%. Under this assumption, a sample size of 240 subjects (120 per group) would provide a 97% power to detect a treatment difference in the primary endpoint at two-sided alpha level of 0.05 using the chi-square test.“
18	Statistische Auswertung	Angaben zur statistischen Auswertung: ☒ ITT-Analyse ☒ per-Protokoll-Analyse (zusätzlich, hier im Bogen nicht dargestellt) ☒ direkter Vergleich der Behandlungsgruppen ☒ Umgang mit fehlenden Daten (missings) beschrieben^a ☒ präspezifizierte Subgruppenanalysen ☐ Interaktionstests ☒ vorgesehene statistische Auswertung hinreichend beschrieben a: Patienten mit fehlenden Baseline-Werten w urden nicht in die ITT-Analyse eingeschlossen. Studienabbrecher oder Patienten mit einem Crossover zu Ruxolitinib w urden in der Primäranalyse als Non-Responder für die Endpunkte „Milzvolumenrückgang $\geq 35\%$“ und „Verbesserung im TSS $\geq 50\%$“. Darüber hinaus w urde keine Ersetzungsstrategie für fehlende Werte verwendet.

		Ergebnisse																																																																								
19	Patientencharakteristika und -fluss	Patientencharakteristika und Vergleichbarkeit der Behandlungsgruppen: X Charakteristika der Behandlungsgruppen dargestellt X relevante Unterschiede zwischen den Gruppen beschrieben <u>Wesentliche Patientencharakteristika</u> <table> <tr> <th></th><th>Ruxolitinib (N = 155)</th><th>Placebo (N = 154)</th></tr> <tr> <td>Alter</td><td></td><td></td></tr> <tr> <td>Median (min-max)</td><td>66 (43-91)</td><td>70 (40-86)</td></tr> <tr> <td>≤ 65 Jahre (%)</td><td>45,2</td><td>33,8</td></tr> <tr> <td>> 65 Jahre (%)</td><td>54,8</td><td>66,2</td></tr> <tr> <td>Geschlecht</td><td></td><td></td></tr> <tr> <td>Männer, N (%)</td><td>79 (51)</td><td>88 (57,1)</td></tr> <tr> <td>Frauen, N (%)</td><td>76 (49,0)</td><td>65 (42,2)</td></tr> <tr> <td>Krankheitstyp</td><td></td><td></td></tr> <tr> <td>PMF, N (%)</td><td>70 (45,2)</td><td>84 (54,5)</td></tr> <tr> <td>PPV-MF, N (%)</td><td>50 (32,2)</td><td>47 (30,5)</td></tr> <tr> <td>PET-MF, N (%)</td><td>35 (22,6)</td><td>22 (14,3)</td></tr> <tr> <td>missing</td><td>-</td><td>1 (0,6)</td></tr> <tr> <td>ECOG Performance-Status, N (%)</td><td></td><td></td></tr> <tr> <td>0</td><td>47 (31,1)</td><td>38 (25,5)</td></tr> <tr> <td>1</td><td>87 (57,6)</td><td>82 (55,0)</td></tr> <tr> <td>2</td><td>14 (9,3)</td><td>25 (16,8)</td></tr> <tr> <td>3</td><td>3 (2,0)</td><td>4 (2,7)</td></tr> <tr> <td>Vorangegangene Therapie mit Hydroxyurea, N (%)</td><td>104 (67,1)</td><td>87 (56,5)</td></tr> <tr> <td>Tastbare Milzlänge, cm (min-max)</td><td>16 (0-33)</td><td>16 (5-34)</td></tr> <tr> <td>IWG-MRT-Risikostufe</td><td></td><td></td></tr> <tr> <td>3 (hoch), %</td><td>58,1</td><td>64,3</td></tr> <tr> <td>2 (intermediär 2), %</td><td>41,3*</td><td>35,1</td></tr> <tr> <td>V617F-Mutation**, %</td><td>72,9</td><td>79,9</td></tr> </table> *: Daten von 1 Patient fehlen/unbekannt, da alle Unteralkalen beim Umzug des Prüfzentrums verloren gingen ** : Daten von mehreren Patienten fehlen aus Datenschutzgründen Relevante Unterschiede zwischen den Gruppen: Angaben aus dem Studienbericht (S. 69): „In general, the treatment groups were well balanced with regard to Baseline demographic characteristics with some small differences noted in age, with a higher proportion of subjects in the ruxolitinib group being ≥ 65 years (54.8% versus 66.2%).”		Ruxolitinib (N = 155)	Placebo (N = 154)	Alter			Median (min-max)	66 (43-91)	70 (40-86)	≤ 65 Jahre (%)	45,2	33,8	> 65 Jahre (%)	54,8	66,2	Geschlecht			Männer, N (%)	79 (51)	88 (57,1)	Frauen, N (%)	76 (49,0)	65 (42,2)	Krankheitstyp			PMF, N (%)	70 (45,2)	84 (54,5)	PPV-MF, N (%)	50 (32,2)	47 (30,5)	PET-MF, N (%)	35 (22,6)	22 (14,3)	missing	-	1 (0,6)	ECOG Performance-Status, N (%)			0	47 (31,1)	38 (25,5)	1	87 (57,6)	82 (55,0)	2	14 (9,3)	25 (16,8)	3	3 (2,0)	4 (2,7)	Vorangegangene Therapie mit Hydroxyurea, N (%)	104 (67,1)	87 (56,5)	Tastbare Milzlänge, cm (min-max)	16 (0-33)	16 (5-34)	IWG-MRT-Risikostufe			3 (hoch), %	58,1	64,3	2 (intermediär 2), %	41,3*	35,1	V617F-Mutation**, %	72,9	79,9
	Ruxolitinib (N = 155)	Placebo (N = 154)																																																																								
Alter																																																																										
Median (min-max)	66 (43-91)	70 (40-86)																																																																								
≤ 65 Jahre (%)	45,2	33,8																																																																								
> 65 Jahre (%)	54,8	66,2																																																																								
Geschlecht																																																																										
Männer, N (%)	79 (51)	88 (57,1)																																																																								
Frauen, N (%)	76 (49,0)	65 (42,2)																																																																								
Krankheitstyp																																																																										
PMF, N (%)	70 (45,2)	84 (54,5)																																																																								
PPV-MF, N (%)	50 (32,2)	47 (30,5)																																																																								
PET-MF, N (%)	35 (22,6)	22 (14,3)																																																																								
missing	-	1 (0,6)																																																																								
ECOG Performance-Status, N (%)																																																																										
0	47 (31,1)	38 (25,5)																																																																								
1	87 (57,6)	82 (55,0)																																																																								
2	14 (9,3)	25 (16,8)																																																																								
3	3 (2,0)	4 (2,7)																																																																								
Vorangegangene Therapie mit Hydroxyurea, N (%)	104 (67,1)	87 (56,5)																																																																								
Tastbare Milzlänge, cm (min-max)	16 (0-33)	16 (5-34)																																																																								
IWG-MRT-Risikostufe																																																																										
3 (hoch), %	58,1	64,3																																																																								
2 (intermediär 2), %	41,3*	35,1																																																																								
V617F-Mutation**, %	72,9	79,9																																																																								

Patientenfluss und Studienausscheider („drop outs“):

- x Studienausscheider (nach Randomisierung) dokumentiert bzw. plausibel beschrieben
- x Umgang mit Studienausscheidern in der Analyse begründet
- ☐ bei fehlender ITT-Analyse: Hinweise auf systematische und relevante Unterschiede zwischen den Studienausscheidern und den gemäß Studienprotokoll behandelten Patienten

Analysenpopulationen

	Ruxolitinib N = 155	Placebo N = 154
Eingeschlossenen und randomisierte Patienten, N (%)	155 (100)	154 (100)
ITT-Population, N (%)	155 (100)	154 (100)
Safetypopulation, N (%)	155 (100)	151 (98,1)
Per Protocol-Population, N (%)	133 (85,8)	123 (79,9)

Patientenfluss (Safety-Population)

	Ruxolitinib N = 155	Placebo N = 154
Patienten mit randomisierter Therapie in der Primäranalyse nach 24 Wochen (02.11.2010) [#] , N (%)	134 (86,5)	78 (51,7)
Patienten mit Crossover, N (%)		36 (23,8)*
Studienabbrecher, N (%)	21 (13,5)	37 (24,5)
Gründe für einen Abbruch (Patienten ohne Crossover), N (%)		
Tod	9 (5,8)	9 (6,0)
Unerwünschtes Ereignis	8 (5,2)	8 (5,3)
Krankheitsprogression	3 (1,9)	12 (7,9)
Einwilligung zurückgezogen	1 (0,6)	5 (3,3)
Andere	0	3 (2,0)
[#] : Dies beinhaltet alle Patienten, die nicht ausgeschiedenen waren, keinen Crossover zu Ruxolitinib hatten und noch am Leben waren.		
[*] : Patienten mit einem Milzgrößenwachstum $\geq 25\%$ gegenüber Studienbeginn konnten vorzeitig entblindet werden und in den Ruxolitinib-Arm wechseln		

20

Ergebnisse*

Primärer Endpunkt: Anteil der Patienten mit einem Rückgang des Milzvolumens um ≥ 35 % nach 24 Wochen (gemessen mit MRT oder CT). ITT-Auswertung

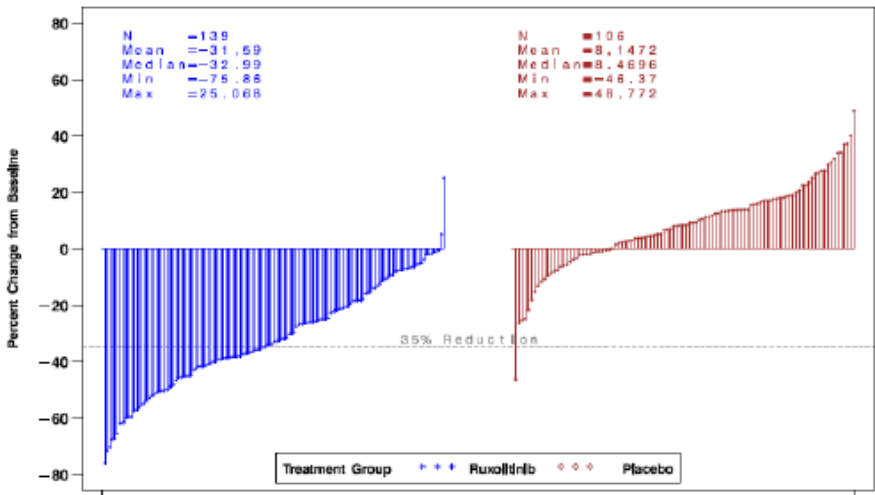
	Gesamt N	Ereignisse	Gesamt N	Ereignisse	OR* (KI)	p-Wert
COMFORT I	155	65 (41,9%)	153	1 (0,7%)	134,4 (17,97; 1005)	p < 0,001

“Odds ratio and 95% confidence interval was calculated using a logistic regression with Baseline spleen volume, sex, Baseline tumor type, previous hydroxyurea use in the 3 months before Screening, and treatment as the model effects.”

Prozentuale Veränderung des Milzvolumens nach 24 Wochen gegenüber Studienbeginn (%) inkl Wasserfall-Plot

	Ruxolitinib	Placebo
N	139	106
Mittelwert	-31,59	+8,1472
Median	-32,99	+8,4696
Min	-75,86	-46,37
Max	+25,068	+48,772

Figure 5: Waterfall Plot of Percent Change From Baseline in Spleen Volume at Week 24 (Observed Cases)



Anteil der Patienten mit einem Rückgang des Milzvolumens um ≥ 35 % im Zeitverlauf (Studienbericht, S. 92)

Zeitpunkt der Visite	Ruxolitinib Gesamt, N*	Ruxolitinib, Patienten mit Response, %	Placebo ¹ Gesamt, N ²	Placebo, Patienten mit Response, %

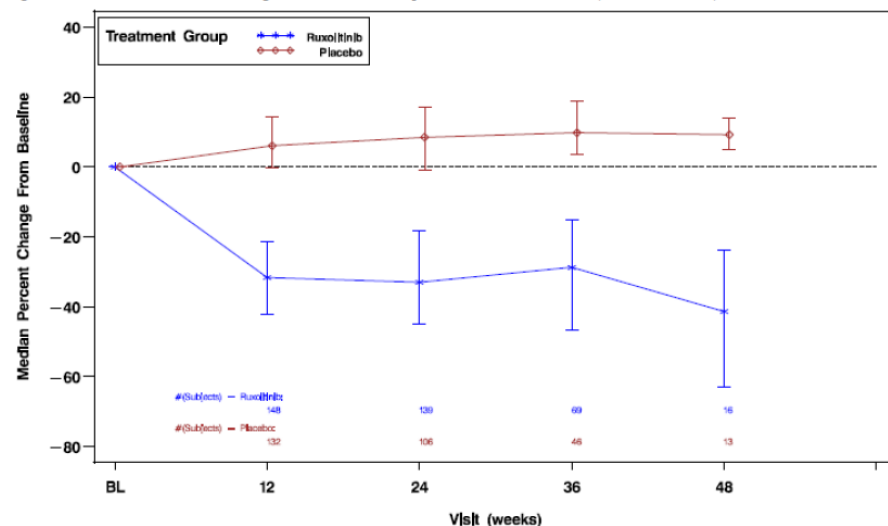
12 Wochen	155	39,4	153	0
24 Wochen	155	41,9	153	0,7
36 Wochen	89	31,5	114	0
48 Wochen	38	23,7	88	0

1: There are substantially higher numbers of evaluable subjects in the placebo group at Weeks 36 and 48 compared with the Ruxolitinib group, because there were a larger number of drop outs in the placebo group who remained in the analysis as non-responders. In addition, placebo subjects who crossed over to Ruxolitinib were included in the analysis as non-responders.

2: The decrease in the number of responders at Weeks 36 and 48 is largely because subjects had not yet reached the visit, rather than because they had lost their response.

Mediane prozentuale Milzvolumenreduktion im Zeitverlauf Observed Cases (ohne Crossover-Patienten)

Figure 9: Median Percent Change from Baseline in Spleen Volume Over Time (Observed Cases)



Note: This analysis excludes subjects with missing data or who withdrew from the study prior to the visit.
Note: Bars represent the 25th and 75th percentiles.

Milzvolumenreduktion (Beginn des Ansprechens). Angaben aus dem Studienbericht, S. 94:

"This study was not designed to rigorously assess the time to response, as the first MRI spleen assessment was not performed until Week 12. The majority of subjects who were responders at Week 24 achieved this response at approximately Week 12 (93.8%). Although spleen volume assessment by MRI did not occur until Week 12, spleen length was measured by palpation at Week 4. As early as Week 4, 38 (24.5%) subjects in the ruxolitinib group achieved a $\geq 50\%$ reduction in palpable spleen length compared with 3 (2.0%) subjects in the placebo group."

Milzvolumenreduktion (Dauer des Ansprechens). Angaben aus dem Studienbericht, S. 94:

"Seventy-one subjects were included in the analysis. The median duration of response was not reached as the majority of subjects were still responding at the time of the data cutoff (52 subjects, 73.2%). Nineteen (26.8%) subjects lost the response prior to the data cutoff date."

Anzahl der Patienten mit einer tastbaren Milzlängenreduktion um $\geq 50\%$ (post-hoc-Analyse). Angaben aus dem Studienbericht, S. 102

“A post hoc analysis was conducted to determine the proportions of subjects who achieved a $\geq 50\%$ reduction from Baseline in palpable spleen length over time. Reduction in palpable spleen length $\geq 50\%$ is part of the IWG response criteria and is discussed in. At Week 24, 60 (39.0%) subjects in the ruxolitinib group had a $\geq 50\%$ reduction from Baseline in palpable spleen length compared with 6 (3.9%) subjects in the placebo group ($p < 0.0001$). The differences between the groups was evident at the first assessment (Week 4), with 38 (24.5%) subjects in the ruxolitinib group and 3 (2.0%) subjects in the placebo group achieving a $\geq 50\%$ reduction from Baseline in palpable spleen length“

Zusammenhang zwischen der Milzvolumenreduktion und der tastbaren Milzlängenreduktion. Angaben aus dem Studienbericht, S. 103 f.

„Reduction in palpable spleen length and reduction in spleen volume as measured by MRI were directionally correlated, however, on an individual subject basis, reduction in spleen volume and reduction in palpable spleen length were not well correlated. However, the proportions of subjects who achieved each of these criteria at Week 24 (primary endpoint) were similar. The proportion of subjects who achieved a $\geq 35\%$ reduction in spleen volume at Week 24 (41.9%) was similar to the proportion of subjects who achieve a $\geq 50\%$ reduction in palpable spleen length from Baseline at Week 24 (39.0%). In addition, the median percent change from Baseline in spleen volume at Week 24 was -33% and the median percent change from Baseline in palpable spleen length at Week 24 was -41%.“

Anteil der Patienten mit einer Verbesserung um $\geq 50\%$ im Total-Symptom-Score nach 24 Wochen. Angaben aus dem Studienbericht, S. 103 ff.:

„A significantly larger proportion of subjects in the ruxolitinib group achieved a $\geq 50\%$ improvement from Baseline in Week 24 Total-Symptom-Score compared with the placebo group (45,9% and 5,3%, respectively, $p < 0,0001$ from Chi-Square test).“

Table 18: Percentage of Subjects With $\geq 50\%$ Improvement From Baseline in Week 24 Total Symptom Score (ITT Subjects)

	Ruxolitinib (N = 155)	Placebo (N = 154)
Number of evaluable subjects	148 (95.5)	152 (98.7)
Subjects achieving a $\geq 50\%$ improvement from Baseline in total symptom score, n (%)		
Yes	68 (45.9)	8 (5.3)
No	80 (54.1)	144 (94.7)
p-value ^a	< 0.0001	
Subjects not evaluable, n (%)		
Missing value at Baseline	6 (3.9)	2 (1.3)
0 at both Baseline and the visit	1 (0.6)	0
Disposition of subjects not meeting the endpoint:		
Crossed over to ruxolitinib prior to the Week 24 visit	0	16 (10.4)
Less than 50% improvement on Week 24 total symptom score	61 (39.4)	94 (61.0)
Discontinued prior to the Week 24 visit	11 (7.1)	25 (16.2)
Missing value at the Week 24 visit	7 (4.5)	7 (4.5)
0 at Baseline and non-zero at the Visit	1 (0.6)	2 (1.3)

^a p-value from Chi Square test.

Source: Table 14.2.4.1.1

Absolute / Prozentuale Veränderung im Total-Symptom-Score gegenüber Studienbeginn (nach 24 Wochen), Observed Cases (ohne Crossover-Patienten)

Table 19: Baseline, Change, and Percent Change From Baseline in Week 24 Total Symptom Score (Observed Cases)

	Ruxolitinib (N = 155)	Placebo (N = 154)
Baseline total symptom score		
N	131	105
Mean (SD)	18.0 (10.9)	16.5 (11.5)
Median	17.8	14.8
Min, Max	0.0, 50.1	0.0, 52.7
Week 24 total symptom score		
Mean (SD)	9.4 (9.7)	19.7 (13.7)
Median	6.6	17.6
Min, Max	0.0, 49.1	0.0, 59.7
Change from Baseline to Week 24^a		
N	131	105
Mean (SD)	-8.6 (10.0)	3.2 (9.4)
Median	-6.9	2.0
Min, Max	-35.8, 22.1	-27.1, 30.3
p-value ^b	<0.0001	
Percent Change from Baseline to Week 24^c		
N	129	103
Mean (SD)	-46.1 (48.6)	41.8 (99.3)
Median	-56.2	14.6
Min, Max	-100.0, 108.3	-100.0, 511.6
p-value ^b	<0.0001	

^a This analysis only includes subjects who had a non-missing change from Baseline to Week 24. Data collected after the date of treatment crossover were not included in this analysis.

^b The p-value was calculated using the Wilcoxon Rank-Sum test.

^c This analysis only includes subjects who had a non-missing percent change from Baseline to Week 24. Data collected after the date of treatment crossover were not included in this analysis.

Source: Table 14.2.5.1 and 14.2.6.1

Veränderung des Total-Symptom-Score gegenüber Studienbeginn im Zeitverlauf (post-hoc-Analyse). Angaben aus dem Studienbericht, S. 112:

“At each of the weeks before the first spleen assessment, the ruxolitinib group had a significantly larger median improvement in Total-Symptom-Scores compared with the placebo group ($p < 0.0001$ at all time points). The placebo group showed no improvement, while the ruxolitinib group showed increasing improvement across the weeks.”

Veränderung der Einzelsymptome des MFSAFv2.0 gegenüber Studienbeginn nach 24 Wochen. Angaben aus dem Studienbericht, S. 114, ff.:

„Although 6 individual symptoms comprised the Total-Symptom-Score, responses for the 6 individual symptoms as well as inactivity in the modified MFSAF v2.0 diary were collected and assessed.

At Week 24, the media percent improvement from Baseline in the Ruxolitinib group was significantly higher than in the placebo group for all 7 of the individual symptoms ($p < 0,0001$ for all comparisons).”

Table 20: Baseline, Change, and Percent Change From Baseline at Week 24 for Individual Symptom Scores (Observed Cases)

	Change From Baseline		Percent Change From Baseline	
	Ruxolitinib (N = 155)	Placebo (N = 154)	Ruxolitinib (N = 155)	Placebo (N = 154)
Night sweats				
Overall Baseline				
N	149	152		
Mean (SD)	2.8 (2.5)	2.4 (2.2)		
Median	2.4	1.9		
Min, Max	0.0, 10.0	0.0, 8.9		
Baseline for subjects included in Week 24 analyses				
N	131	105	103 ^a	90
Mean (SD)	2.7 (2.5)	2.4 (2.1)	3.5 (2.3)	2.8 (2.0)
Median	2.4	2.1	3.1	2.4
Min, Max	0.0, 10.0	0.0, 8.9	0.1, 10.0	0.1, 8.9
Change from Baseline at Week 24^b				
N	131	105	103	90
Mean (SD)	-1.3 (2.2)	0.4 (1.9)	-42.1 (95.5)	51.0 (139.0)
Median	-1.0	0.4	-75.4	14.9
Min, Max	-8.8, 5.0	-4.6, 5.1	-100.0, 573.1	-100.0, 700.0
Itching				
Overall Baseline				
N	149	152		
Mean (SD)	2.3 (2.1)	2.3 (2.6)		
Median	1.9	1.4		
Min, Max	0.0, 9.1	0.0, 9.6		
Baseline for subjects included in Week 24 analyses				
N	131	105	98 ^a	77
Mean (SD)	2.2 (2.1)	2.3 (2.5)	3.0 (2.0)	3.1 (2.5)
Median	1.7	1.4	2.7	2.3
Min, Max	0.0, 9.1	0.0, 9.6	0.1, 9.1	0.1, 9.6
Change from Baseline at Week 24^b				
N	131	105	98	77
Mean (SD)	-1.3 (2.3)	0.4 (2.4)	-42.8 (111.0)	111.9 (573.5)
Median	-1.0	0.0	-85.3	-3.7
Min, Max	-9.1, 6.0	-5.8, 10.0	-100.0, 596.4	-100.0, 4571.4
Abdominal discomfort				
Overall Baseline				
N	149	152		
Mean (SD)	3.8 (2.3)	3.5 (2.4)		
Median	3.7	3.3		
Min, Max	0.0, 9.9	0.0, 10.0		
Baseline for subjects included in Week 24 analyses				
N	131	105	124 ^a	100
Mean (SD)	3.7 (2.2)	3.4 (2.4)	3.9 (2.2)	3.6 (2.3)
Median	3.7	3.3	3.8	3.3
Min, Max	0.0, 9.9	0.0, 10.0	0.1, 9.9	0.1, 10.0
Change from Baseline at Week 24^b				
N	131	105	124	100
Mean (SD)	-1.8 (2.2)	0.6 (2.1)	-29.4 (125.8)	55.7 (181.0)
Median	-1.6	0.3	-55.6	7.7
Min, Max	-8.3, 4.6	-5.4, 7.0	-100.0, 850.0	-100.0, 1075.0

Pain under ribs on left				
Overall Baseline				
N	149	152		
Mean (SD)	2.6 (2.3)	2.5 (2.5)		
Median	2.3	1.6		
Min, Max	0.0, 9.3	0.0, 9.4		
Baseline for subjects included in the Week 24 analyses				
N	131	105	109 ^a	86
Mean (SD)	2.7 (2.3)	2.5 (2.5)	3.2 (2.1)	3.0 (2.4)
Median	2.3	1.6	3.0	2.2
Min, Max	0.0, 9.3	0.0, 9.4	0.1, 9.3	0.1, 9.4
Change from Baseline at Week 24 ^b				
N	131	105	109	86
Mean (SD)	-1.4 (2.1)	0.4 (2.3)	-46.9 (111.5)	62.8 (224.6)
Median	-1.0	0.0	-75.0	-7.7
Min, Max	-7.9, 5.4	-7.3, 8.4	-100.0, 950.0	-100.0, 988.9
Feeling of fullness (early satiety)				
Overall Baseline				
N	149	152		
Mean (SD)	3.7 (2.2)	3.5 (2.4)		
Median	4.0	3.2		
Min, Max	0.0, 8.9	0.0, 9.4		
Baseline for subjects included in Week 24 analyses				
N	131	105	122 ^a	96
Mean (SD)	3.7 (2.2)	3.3 (2.2)	3.9 (2.0)	3.6 (2.1)
Median	4.0	3.0	4.0	3.3
Min, Max	0.0, 8.9	0.0, 9.4	0.1, 8.9	0.1, 9.4
Change from Baseline at Week 24 ^b				
N	131	105	122	96
Mean (SD)	-1.8 (2.2)	0.7 (2.3)	-43.0 (70.9)	70.3 (309.2)
Median	-1.7	0.6	-62.0	18.9
Min, Max	-8.2, 4.1	-5.2, 7.1	-100.0, 458.3	-100.0, 2855.6
Bone/muscle pain				
	Change From Baseline		Percent Change From Baseline	
	Ruxolitinib (N = 155)	Placebo (N = 154)	Ruxolitinib (N = 155)	Placebo (N = 154)
Overall Baseline				
N	149	152		
Mean (SD)	3.0 (2.5)	2.6 (2.4)		
Median	2.6	2.0		
Min, Max	0.0, 10.0	0.0, 9.8		
Baseline for subjects included in Week 24 analyses				
N	131	105	108 ^a	89
Mean (SD)	3.0 (2.5)	2.7 (2.4)	3.6 (2.3)	3.2 (2.3)
Median	2.6	2.1	3.5	2.4
Min, Max	0.0, 10.0	0.0, 9.8	0.1, 10.0	0.1, 9.8
Change from Baseline at Week 24 ^b				
N	131	105	108	89
Mean (SD)	-1.0 (2.1)	0.7 (1.9)	-21.2 (125.1)	37.7 (107.0)
Median	-0.6	0.5	-46.6	12.3
Min, Max	-6.5, 4.6	-4.2, 7.0	-100.0, 1025.0	-100.0, 521.4

Degree of inactivity				
Overall Baseline				
N	149	152		
Mean (SD)	3.9 (2.6)	3.8 (2.5)		
Median	3.9	3.7		
Min, Max	0.0, 9.9	0.0, 9.3		
Baseline for subjects included in Week 24 analyses				
N	131	105	119 ^a	97
Mean (SD)	3.8 (2.6)	3.7 (2.4)	4.2 (2.4)	4.0 (2.2)
Median	3.7	3.7	4.3	4.0
Min, Max	0.0, 9.9	0.0, 8.9	0.1, 9.9	0.1, 8.9
Change from Baseline at Week 24 ^b				
N	131	105	119	97
Mean (SD)	-1.2 (1.9)	0.6 (2.3)	-32.1 (69.4)	60.0 (238.7)
Median	-0.8	0.6	-37.1	15.2
Min, Max	-8.0, 5.1	-5.1, 8.0	-100.0, 460.0	-100.0, 1896.3

^a These values are the Baseline scores for the subjects who were included in the percent change from Baseline analyses. Mean and median Baseline scores are higher than the overall group or the Baseline for subjects included in the change from Baseline analyses, because subjects with a 0 at Baseline are excluded from the percent change from Baseline analyses.

^b $p < 0.0001$ using Wilcoxon Rank-Sum Test.

Note: Degree of inactivity is not part of the total symptom score.

Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (post-hoc-Analyse, ohne Crossover-Patienten), Angaben aus dem

Studienbericht, S. 119

"The majority of subjects who achieved at least a $\geq 35\%$ reduction from Baseline in spleen volume also achieved a $\geq 50\%$ improvement in abdominal symptom score (38/60, 63.3%). Among the 66 subjects who had a less than 35% reduction from Baseline in spleen volume, 32 (48.5%) achieved a $\geq 50\%$ improvement in abdominal symptom score. This analysis shows that a substantial proportion of subjects who did not meet the responder criteria for spleen volume obtained benefit in abdominal symptoms; however the proportion who achieved a $\geq 50\%$ improvement in abdominal symptom score was greater in those who met the primary endpoint definition of $\geq 35\%$ reduction from Baseline in spleen volume."

Milzvolumen- rückgang, N (%)	Besserung der Oberbauch- beschwerden		Gesamt
	< 50%	$\geq 50\%$	
< 35%	36 (51,5)	32 (48,5)	66 (100)
$\geq 35\%$	22 (36,7)	38 (63,3)	60 (100)

Gesamtüberleben (ITT), Angaben aus dem Studienbericht S. 119

"Ten (6.5%) subjects in the ruxolitinib group and 14 (9.1%) subjects in the placebo group died. This includes 20 subjects who died during randomized treatment, after crossover, and up to 28 days after study withdrawal, and 4 subjects who died more than 28 days after withdrawal from the study and represents a hazard ratio of 0.668 ($p = 0.3268$). This study was not designed or powered for a robust analysis of overall survival, and, in fact, the number of events in each treatment group was too low for a meaningful statistical comparison."

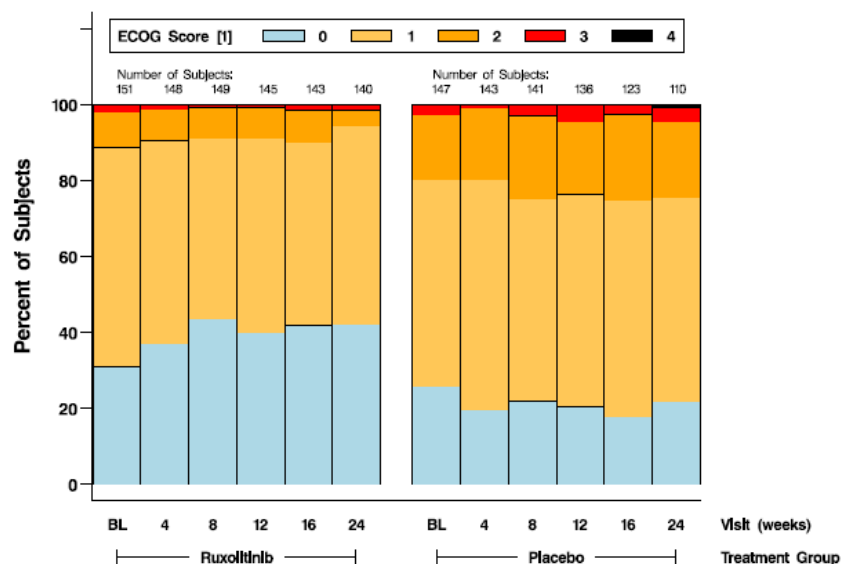
Leukämiefreies Überleben

Das Leukämiefreie Überleben war im Studienprotokoll der COMFORT I-Studie als Langzeitendpunkt geplant, der erst nach Woche 144 ausgewertet werden sollte.

Veränderung des ECOG Performance-Status, Studienbericht, S. 127:

„At Baseline 47 (31.1%) subjects in the ruxolitinib group and 38 (25.9%) subjects in the placebo group had an ECOG score of 0 (fully active, able to carry on all pre-disease performance without restriction); 87 (57.6%) subjects in the ruxolitinib and 80 (54.4%) in the placebo group had an ECOG score of 1 (restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature); 14 (9.3%) subjects in the ruxolitinib group and 25 (17.0%) in the placebo group had an ECOG score of 2 (ambulatory and capable of all self-care but unable to carry out any work activities, up and about more than 50% of waking hours); and 3 (2.0%) subjects in the ruxolitinib group and 4 (2.7%) subjects in the placebo group had an ECOG score of 3 (capable of only limited self-care, confined to bed or chair more than 50% of waking hours).

In general, the proportion of subjects in the ruxolitinib group who had an improvement in ECOG status increased across the visits, and the proportions of subjects in the placebo group who had a worsening in ECOG status increased across the visits.“



EORTC QLQ-C30 (Observed Data), Studienbericht, S. 129 ff.

„At Baseline, there were no clinically meaningful differences between Ruxolitinib and placebo groups for the individual subscales of the EORTC QLQ-C30.

At Week 24, the Ruxolitinib group showed significant improvement in the global health status and functional subscales of the EORTC QLQ-C30 compared with the placebo group with exception of the cognitive functioning.

Comparisons between the treatment groups for the individual subscales of the EORTC QLQ-C30 of fatigue, nausea and vomiting, pain, dyspnea, insomnia, appetite loss, diarrhea, and financial difficulties showed significant greater improvement in the ruxolitinib group compared with the placebo group at Week 24 ($p < 0.01$ for all comparisons using the ANCOVA). The individual subscale of constipation did not reach statistical significance; however, the majority of subjects did not have constipation as shown by the median of 0 at Baseline and Week 24 in both treatment groups.“

Table 24: Mean EORTC QLQ-C30 Global Health Status and Functional Scale Results at Baseline and Week 24 (Observed Data)

Scale Time point	Ruxolitinib (N = 155)	Placebo (N = 154)
Global Health Status/Quality of Life		
N	136	104
Baseline, Mean (SD)	52.6 (23.1)	54.5 (22.6)
Change from Baseline at Week 24, Mean (SD)	12.3 (25.4)	-3.4 (21.5)
p-value ^a	< 0.0001	
Physical Functioning		
N	137	107
Baseline, Mean (SD)	69.5(21.1)	67.6 (20.5)
Change from Baseline at Week 24, Mean (SD)	9.3 (18.5)	-3.5 (17.4)
p-value ^a	< 0.0001	
Role Functioning		
N	137	105
Baseline, Mean (SD)	64.6 (29.2)	65.1 (30.6)
Change from Baseline at Week 24, Mean (SD)	6.4 (29.7)	-11.1 (30.9)
p-value ^a	< 0.0001	
Emotional Functioning		
N	137	105
Baseline, Mean (SD)	73.7 (23.5)	76.3 (23.5)
Change from Baseline at Week 24, Mean (SD)	8.3 (20.7)	-1.2 (23.0)
p-value ^a	0.0009	
Cognitive Functioning		
N	136	103
Baseline, Mean (SD)	81.1 (19.1)	79.3 (20.1)
Change from Baseline at Week 24, Mean (SD)	2.5 (19.0)	-1.1 (21.0)
p-value ^a	0.0587	
Social Functioning		
N	135	106
Baseline, Mean (SD)	68.0 (26.6)	65.9 (29.8)
Change from Baseline at Week 24, Mean (SD)	9.4 (25.5)	-9.0 (28.1)
p-value ^a	<0.0001	

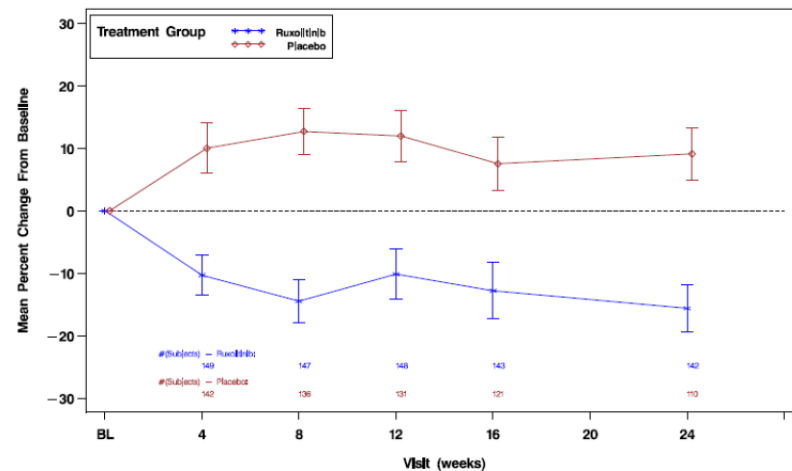
^a p-value from analysis of covariance.

Source: Tables 14.2.15.1, 14.2.15.2, 14.2.15.3, 14.2.15.4, 14.2.5.1, and 14.2.15.6

PROMIS Fatigue-Scale, Studienbericht, S. 130 f.

“At baseline, both groups had similar mean scores on the PROMIS Fatigue-Scale (43.7 in the Ruxolitinib group and 43.3 in the placebo group on a 100 point scale, where 100 presents always fatigued and 0 represents never fatigued. At week 24, the ruxolitinib group showed significantly larger mean percent improvement from Baseline in fatigue compared with the placebo group (15.6% improvement in the ruxolitinib group and 9.1% worsening in the placebo group, $p < 0.0001$) This difference was evident as early as Week 4, with the ruxolitinib group having a mean percent improvement from Baseline of 10.3% and the placebo group having a 10.0% worsening from Baseline this difference between the treatment groups was seen at all visits through Week 24.”

Figure 19: Mean Percent Change From Baseline in PROMIS Fatigue Score Over Time (Observed Cases)

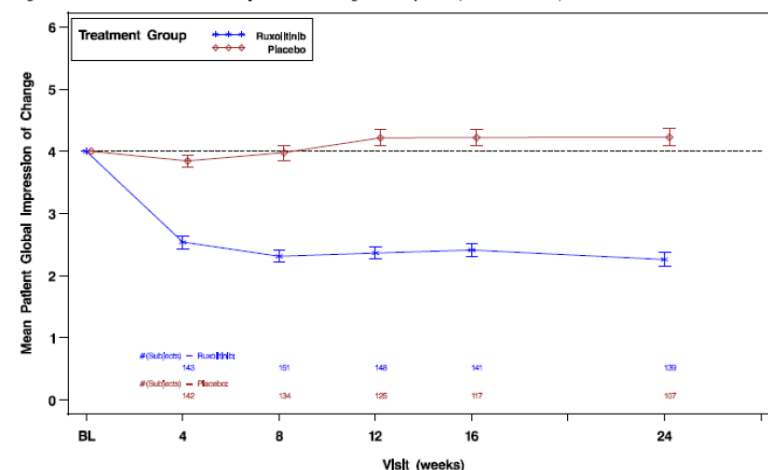


Note: Bars indicate the 25th and 75th percentile at each visit.

Patient Global Impression of Change (Observed Cases), Angaben aus dem Studienbericht, S. 133 f. :

“The PGIC was used to evaluate a subjects overall sense of treatment effect using a 7-point scale. The specific wording was: Since the start of the treatment you’ve received in this study, your myelofibrosis 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. “At each time point measured, there were significant differences in PGIC scores between the treatment groups ($p < 0.0001$ at all time points). From Week 4 through Week 24, the median PGIC score was 2.0 (much improved) in the ruxolitinib group and 4.0 (no change) in the placebo group. Over time, the ruxolitinib group maintained a mean PGIC score associated with improvement; whereas, the placebo group maintained a mean PGIC score associated with no change”

Figure 20: Mean Patient Global Impression of Change Score by Visit (Observed Cases)



Note: Vertical line indicates the standard error at each visit.

Note: Although PGIC was not assessed at Baseline, a PGIC score of 4 (no change) was applied to all subjects at Baseline for clarity in results for change from Baseline.

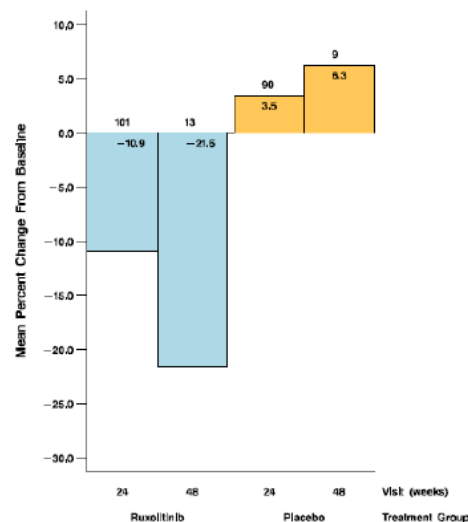
JAK2 V617F (allele burden was measured in whole blood at Baseline and Weeks 24 and 48), Angaben aus dem Studienbericht, S. 136 f.:

„Recent studies have shown that an activating mutation resulting in a valine to phenylalanine substitution at residue 617 (V617F) in the pseudokinase

domain of JAK2 enzyme is present in a high proportion of patients with myeloproliferative neoplasms.”

“Given the importance of the JAK2 V617F mutation in MPNs, the percentage of JAK2 V617F mutant allele relative to total JAK2 (wild type JAK2 plus JAK2 V617F, referred to as JAK2V617F allele burden) was assessed in this study to determine what effect, if any, JAK 1/2 inhibition had on JAK2 V617F allele burden. JAK2 V617F allele burden was measured in whole blood at Baseline and Weeks 24 and 48.”

Figure 22: Mean Percent Change From Baseline in Percent V617F at Weeks 24 and 48 (Observed Cases)



Notwendigkeit von Erythrozytenkonzentraten (Studienbericht, S. 191)

“Forty-three (27.7%) subjects in the ruxolitinib group and 44 (28.6%) subjects in the placebo group entered the study with a history of blood component transfusions. (...)”

As a simple objective measure of transfusion requirements over the course of the study the total units of red blood cells per unit time on study was calculated for each treatment group. The mean number of red blood cell transfusions was 0.96 units per subject per month in the ruxolitinib group and 0.77 units per subject per month in the placebo group.”

Table 38: Summary of Red Blood Cell Transfusion Dependency Based on SAP-Defined Criteria (Safety Subjects)

Dependency Change	Ruxolitinib		Placebo	
	N	N (%)	N	n (%)
Dependent to Independent				
8 weeks before Screening to final 8 weeks	32	6 (18.8)	32	10 (31.3)
Independent to Dependent				
8 weeks before Screening to final 8 weeks	123	33 (26.8)	118	17 (14.4)

Source: Table 14.2.2.11

Ergebnisse aus Subgruppenanalysen. Angaben aus dem Studienbericht, S. 137 ff.:

Folgende Subgruppeneigenschaften waren im statistischen Analyseplan prädefiniert:

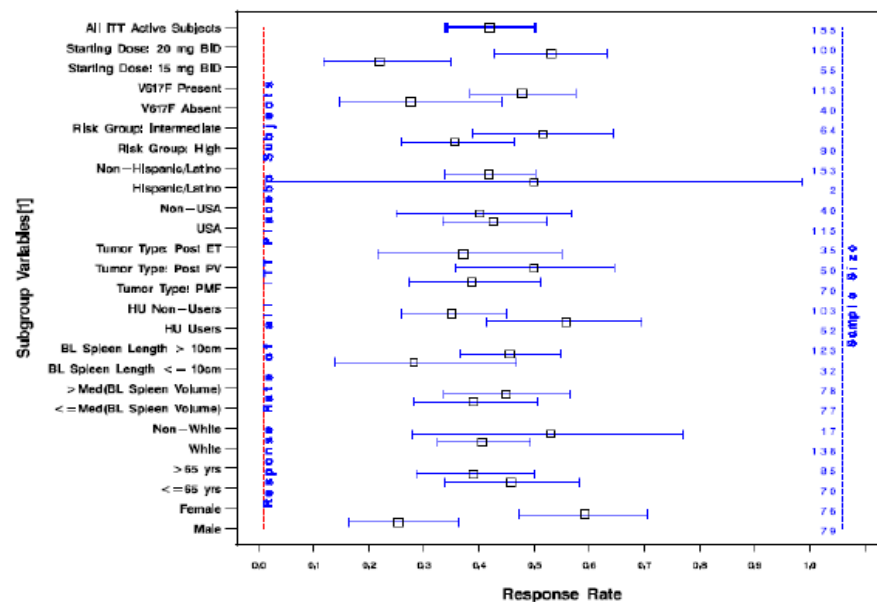
- Geschlecht, Alter, Myelofibrose-Typ, Hydroxyurea-Gabe (in den zurückliegenden 3 Monaten), Vorliegen der V617F-Mutation, Milzvolumen zu Studienbeginn, Tastbare Milzlänge, Land, Rasse,

Ethnizität

Untersuchte Endpunkte:

- Anteil der Patienten mit einem Rückgang des Milzvolumens um $\geq 35\%$ nach 24 Wochen (Figure 23), a priori geplant
- Anteil der Patienten mit einer Verbesserung um $\geq 50\%$ im Total-Symptom-Score (Figure 24), (post hoc berechnet)

Figure 23: Proportions of Subjects in Each Subgroup Who Achieved a $\geq 35\%$ Reduction From Baseline in Spleen Volume at Week 24 (Ruxolitinib-Treated ITT Subjects)



Note: The vertical line to the left indicates the response rate of all ITT placebo subjects.

Note: The box represents the point estimate and the error bars represent the 95% confidence intervals based upon binomial exact method.

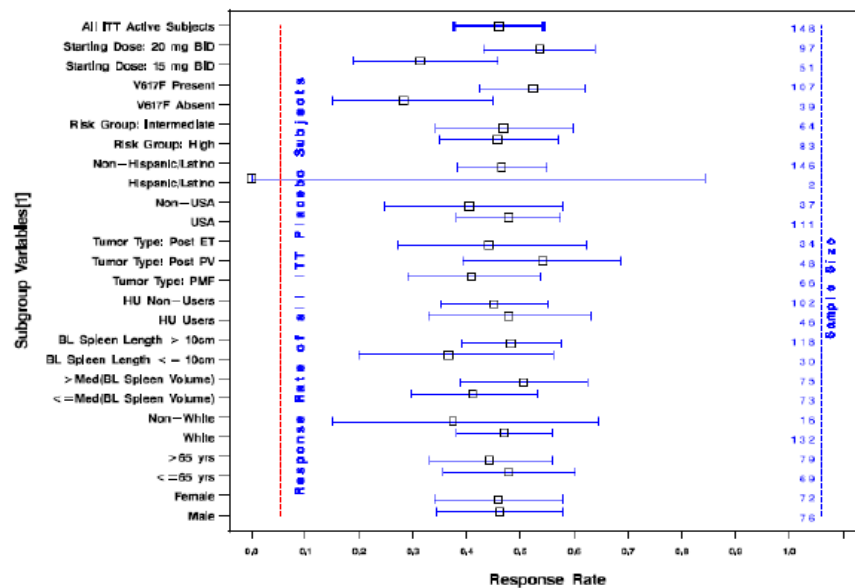
BID = twice daily; BL = Baseline; HU = hydroxyurea use; Med = median

Source: Figure 14.4.2.13

“For all subgroups, a significantly higher proportion of subjects in the ruxolitinib group achieved $\geq 35\%$ reduction from Baseline in spleen volume (see Figure 23). Placebo data are not presented, because there was only 1 subject who achieved a $\geq 35\%$ reduction from Baseline in spleen volume; therefore, a response rate could not be calculated in the majority of subgroups for the placebo group.

Subgroups of ruxolitinib-treated subjects that showed the largest deviation from the point estimate of response seen for the entire population were further investigated. These included starting dose, sex, and V617F mutation at Baseline.”

Figure 24: Proportions of Subjects in Each Subgroup Who Achieved a $\geq 50\%$ Reduction From Baseline in Total Symptom Score at Week 24 (Ruxolitinib-Treated ITT Subjects)



Note: The vertical line to the left indicates the response rate of all ITT placebo subjects.
 Note: The box represents the point estimate and the error bars represent the 95% confidence intervals based upon binomial exact method.
 BID = twice daily; BL = Baseline; HU = hydroxyurea use; Med = median
 Source: Figure 14.4.25.13

“For all subgroups in which there was an adequate number of evaluable subjects, a significantly higher proportion of subjects in the ruxolitinib group achieved $\geq 50\%$ improvement from Baseline in Week 24 Total-Symptom-Score.

**21 Unerwünschte
Therapie-
wirkungen**

Angaben aus dem Studienbericht, S. 148

„The safety population includes all 155 subjects randomized to the ruxolitinib group and 151 of the 154 subjects randomized to the placebo group. Two subjects randomized to the placebo group never received study medication. One subject randomized to the placebo group was not included because the subject's primary source data was irretrievable by the site after they moved their physical location.

Safety data for the 36 subjects randomized to the placebo group who crossed over to ruxolitinib are included in the safety tables in the placebo group up until the time of crossover.

Any adverse event (AE) that occurred or worsened after the time of the first dose through 28 days after the last dose was considered a treatment-emergent AE (TEAE).”

Zusammenfassung der unerwünschten Ereignisse (Studienbericht, S. 150)

	Ruxolitinib (N = 155), Anzahl der Patienten (%)	Placebo (N = 151), Anzahl der Patienten (%)
Subjects who died	9 (5,8)	10 (6,6)

	during study*		
	Subjects who had an SAE	43 (27,7)	53 (35,1)
	Subjects who had an TEAE Grad 3 or higher	73 (47,1)	67 (44,4)
	Subjects who had a TEAE	151 (97,4)	148 (98,0)
	Subjects who had an treatment-related TEAE	115 (74,2)	84 (55,6)
	Subjects who withdrew from the study because of an AE	17 (11,0)	16 (10,6)
	Subjects who had study medication discontinued because of an AE	37 (23,9)	32 (21,2)
	Subjects who had study medication decrease because of an AE	62 (40,0)	14 (9,3)
	Note: TEAEs that occurred after the date of treatment crossover are not included in this table. * The number of deaths discussed for safety differs from the number of deaths included in the overall survival analysis, because additional data beyond the subjects participation in the study (28 days after last dose) were obtained for the overall survival analysis.		

Unerwünschte Ereignisse alle Grade

Table 27: Summary of Treatment-Emergent Adverse Events Occurring in $\geq 5\%$ of Ruxolitinib-Treated Subjects (Safety Population)

	Ruxolitinib (N = 155)	Placebo (N = 151)
Subjects with any treatment-emergent adverse events, n (%)	151 (97.4)	148 (98.0)
MedDRA Preferred Term		
Thrombocytopenia ^a	53 (34.2)	14 (9.3)
Anemia ^b	48 (31.0)	21 (13.9)
Fatigue	39 (25.2)	51 (33.8)
Diarrhea	36 (23.2)	32 (21.2)
Ecchymosis ^c	29 (18.7)	14 (9.3)
Edema peripheral	29 (18.7)	34 (22.5)
Dyspnea	27 (17.4)	26 (17.2)
Dizziness	23 (14.8)	10 (6.6)
Headache	23 (14.8)	8 (5.3)
Nausea	23 (14.8)	29 (19.2)
Hemoglobin decreased ^b	22 (14.2)	6 (4.0)
Constipation	20 (12.9)	18 (11.9)
Pain in extremity	19 (12.3)	15 (9.9)
Vomiting	19 (12.3)	15 (9.9)
Insomnia	18 (11.6)	15 (9.9)
Arthralgia	17 (11.0)	13 (8.6)
Pyrexia	17 (11.0)	11 (7.3)
Abdominal pain	16 (10.3)	62 (41.1)
Cough	15 (9.7)	13 (8.6)
Platelet count decreased ^b	15 (9.7)	4 (2.6)
Abdominal distension	13 (8.4)	16 (10.6)
Pneumonia	13 (8.4)	9 (6.0)
Back pain	11 (7.1)	13 (8.6)
Cardiac murmur	11 (7.1)	5 (3.3)
Urinary tract infection	11 (7.1)	7 (4.6)
Muscle spasms	10 (6.5)	11 (7.3)
Night sweats	10 (6.5)	18 (11.9)
Weight increased	10 (6.5)	2 (1.3)
Abdominal pain upper	9 (5.8)	13 (8.6)
Anorexia	9 (5.8)	10 (6.6)

	Ruxolitinib (N = 155)	Placebo (N = 151)
Dyspepsia	9 (5.8)	8 (5.3)
Upper respiratory tract infection	9 (5.8)	14 (9.3)
Asthenia	8 (5.2)	12 (7.9)
Chills	8 (5.2)	3 (2.0)
Flatulence	8 (5.2)	1 (0.7)
Pallor	8 (5.2)	9 (6.0)

^a New onset or worsening thrombocytopenia/platelet count decreased was reported in 65 (41.9%) subjects in the ruxolitinib group and 16 (10.6%) subjects in the placebo group.

^b New onset or worsening anemia/Hgb decreased was reported in 67 (43.2%) subjects in the ruxolitinib group and 26 (17.2%) subjects in the placebo group.

^c Verbatim terms of bruising where a traumatic cause was not specified were coded to the preferred term of ecchymosis, so events would code to the skin system-organ class, rather than the injury system-organ class. Verbatim terms of bruising where a traumatic cause was specified were coded to contusion and therefore the injury system-organ class, as the MedDRA convention would specify.

Note: Adverse event listings were searched for the terms of thrombocytopenia or platelet count decreased. Subjects with both terms reported were only counted once. This was also done for anemia and Hgb decreased.

Note: TEAEs that occurred after the date of treatment crossover are not included in this table.

Source: Table 14.3.2.5

"The most frequently reported TEAEs in the ruxolitinib group were thrombocytopenia and anemia, both of which occurred more frequently in the ruxolitinib group compared with the placebo group. (...)

*Other events that were reported $\geq 50\%$ more frequently in the **ruxolitinib group** (presented as ruxolitinib and placebo) were ecchymosis (18.7% and 9.3%), dizziness (14.8% and 6.6%), headache (14.8% and 5.3%), pyrexia (11.0% and 7.3%), urinary tract infection (7.1% and 4.6%), cardiac murmur*

(7.1% and 3.3%), weight increased (6.5% and 1.3%), chills (5.2% and 2.0%), and flatulence (5.2% and 0.7%).

Other events that were reported $\geq 50\%$ more frequently in the **placebo group** (presented as ruxolitinib and placebo) were abdominal pain (10.3% and 41.1%), night sweats (6.5% and 11.9%), upper respiratory tract infection (5.8% and 9.3%), and asthenia (5.2% and 7.9%). (...)

Abdominal pain and fatigue were reported less frequently in the ruxolitinib group; this is consistent with the efficacy findings. More subjects in the ruxolitinib group had urinary tract infections, and more subjects in the placebo group had upper respiratory tract infections.”

Unerwünschte Ereignisse der Grade 3 und 4 (Studienbericht, S. 152)

Table 28: Summary of Treatment-Emergent Adverse Events of Grade 3 or Higher Severity Occurring in $\geq 3\%$ of Subjects in Either Treatment Group (Safety Population)

	Ruxolitinib (N = 155)	Placebo (N = 151)
Subjects with a treatment-emergent adverse event of Grade 3 or higher, n (%)	73 (47.1)	67 (44.4)
MedDRA Preferred Term		
Anemia ^a	24 (15.5)	7 (4.6)
Hemoglobin decreased	15 (9.7)	3 (2.0)
Thrombocytopenia ^b	13 (8.4)	3 (2.0)
Fatigue	8 (5.2)	10 (6.6)
Pneumonia	8 (5.2)	6 (4.0)
Abdominal pain	4 (2.6)	17 (11.3)
Dyspnea	2 (1.3)	6 (4.0)
Hyponatremia	1 (0.6)	5 (3.3)
Splenic infarction	1 (0.6)	5 (3.3)

^a New onset or worsening anemia/Hgb decreased of Grade 3 or higher was reported in 38 (24.5%) subjects in the ruxolitinib group and 9 (6.0%) subjects in the placebo group.

^b New onset or worsening thrombocytopenia/platelet count decreased of Grade 3 or higher was reported in 10 (6.5%) subjects in the ruxolitinib group and 3 (2.0%) subjects in the placebo group.

Note: Adverse event listings were searched for the terms of thrombocytopenia or platelet count decreased. Subjects with both terms reported were only counted once. This was also done for anemia and Hgb decreased.

“Similar proportions of subjects in the ruxolitinib and placebo groups had a TEAE of Grade 3 or higher severity (47,1 and 44,4% respectively”

Ekchymosen (Studienbericht, S. 155)

„The incidences of ecchymosis were higher with ruxolitinib (n=29), compared with placebo (n=14). Three episodes of ecchymosis in ruxolitinib-treated subjects may possibly have been temporally associated with decreased platelets (Grades 2, 3, and 4 platelet counts each n=1) and 26 episodes were associated with normal or Grade 1 platelet counts (see Listing 16.2.4.8.1 for platelet count by subject).”

Blutungsergebnisse (Studienbericht S. 155)

“Bleeding events (defined by preferred terms containing the words bleeding, hemorrhage, or epistaxis) occurred slightly more frequently with ruxolitinib (n=17), compared with placebo (n=13). Eleven of these bleeding events were SAEs (6 in the ruxolitinib group and 5 in the placebo group).”

Infektionen (Studienbericht S. 162)

“59 (38,1%) subjects in the ruxolitinib group and 64 (42,4%) subjects in the

		<i>placebo group had a TEAE in the infections and investigations system organ class. The most frequently reported infection was pneumonia, which occurred in similar proportions of subjects in the Ruxolitinib and placebo groups (8,4% and 6,0%).</i> <i>Serious AE in the infections and investigations system organ class occurred in 14 (9,0%) subjects in the Ruxolitinib group and 14 (9,3%) subjects in the placebo group. Pneumonia was reported in 6,5% of subjects in the Ruxolitinib group compared with 3,3% of subjects in the placebo group"</i> <u>Leukämische Transformation</u> (Studienbericht S. 163) <i>"Two subjects in the ruxolitinib group and no subjects in the placebo group underwent leukemic transformation during the study."</i>
22	Fazit der Autoren	-

* Die Ergebnisse werden mit den im Studienbericht verwendeten Effektmaßen dargestellt.

Verzerrungspotenzial auf Studien- und Endpunktebene

Verzerrungspotenzial auf Studienebene (COMFORT I)

RCT mit dem zu bewertenden Arzneimittel (Erhebung des primären Endpunkts nach 24 Wochen)

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial auf Studienebene
			Patient	Behandler			
COMFORT I	Ja	Ja	Ja	Ja	Nein	Nein	niedrig

Verzerrungspotenzial auf Endpunktebene

Bewertung des Verzerrungspotenzials für den Endpunkt Gesamtüberleben

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Ja	Nein	Nein	Niedrig

Bewertung des Verzerrungspotenzials für den Endpunkt Rückgang des Milzvolumens um $\geq 35\%$ nach 24 Wochen in MRT oder CT

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Ja	Nein	Nein	Niedrig

Bewertung des Verzerrungspotenzials für den Endpunkt Milzvolumenreduktion (prozentuale Veränderung gegenüber Studienbeginn)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Nein ¹	Nein	Nein	Hoch
1: Für diesen Endpunkt wurde eine Bewertung nur für diejenigen Patienten vorgenommen, für die ein Wert zu Studienbeginn und nach 24 Wochen vorlag (Observed Cases = ohne Crossover-Patienten), sodass ein Teil der Patienten – insbesondere in der Placebo-Gruppe – nicht in die Auswertung eingegangen ist.						

Bewertung des Verzerrungspotenzials für den Endpunkt TSS (Total-Symptom-Score); Anteil der Patienten mit Rückgang um $\geq 50\%$

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	ja	Nein	Nein	Niedrig

Bewertung des Verzerrungspotenzials für den Endpunkt TSS (absolute bzw. prozentuale Veränderung gegenüber Studienbeginn nach 24 Wochen)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Nein ¹	Nein	Nein	Hoch
1: Für diesen Endpunkt wurde eine Bewertung der Observed Cases (ohne Crossover-Patienten) vorgenommen, sodass ein Teil der Patienten – insbesondere in der Placebo-Gruppe – nicht in die Auswertung eingegangen ist.						

Bewertung des Verzerrungspotenzials für den Endpunkt Zusammenhang zwischen Milzvolumen und Oberbauchbeschwerden (post-hoc-Analyse, ohne Crossover-Patienten)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Nein ¹	Ja ²	Nein	Hoch
1: Für diesen Endpunkt wurde eine Bewertung der Observed Cases (ohne Crossover-Patienten) vorgenommen. 2: Es handelt sich nicht um eine a priori geplante Auswertung.						

Bewertung des Verzerrungspotenzials für den Endpunkt ECOG Performance-Status

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Niedrig	Ja	Nein	Ja ¹	Nein	Hoch
1: Die Verteilung der ECOG-Scores zu Studienbeginn war nicht in beiden Behandlungsarmen gleich.						

Bewertung des Verzerrungspotenzials für den Endpunkt QLQ-C30 der EORTC (absolute Veränderung einzelner Subskalen gegenüber Studienbeginn)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Nein	Ja	Nein ¹	Nein	Nein	Hoch
1: In die Auswertung gingen nur die Patienten ein, für die zum Zeitpunkt der Analyse ein ausgefüllter Fragebogen vorlag						

Bewertung des Verzerrungspotenzials für den Endpunkt unerwünschte Ereignisse unter Behandlung

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT I	Nein	Ja	Ja ¹	Nein	Nein	Niedrig
1: Die Auswertungen basieren auf der Safety-Population, die alle Patienten einschloss, die mindestens 1 Dosis der Studienmedikation erhalten hatten.						

**ii. Datenextraktionsbogen der pivotalen Studie COMFORT II
(CINC424A2352)**



Indikation: Myelofibrose

**Datenextraktion und Bewertung der Studie
CINC424A2352 (COMFORT II)**

Nr.	Feld	Allgemeine Angaben
1	Quelle	• Herstellerdossier zu Ruxolitinib • Studienbericht inkl. Studienprotokoll (Protocol INCB 18424-352): A Randomized Study of the JAK Inhibitor INCB018424 Tablets Compared to Best-Available-Therapy in Subjects with Primary Myelofibrosis (PMF), Post-Polycythemia Vera-Myelofibrosis (PPV-MF) or Post-Essential Thrombocythemia Myelofibrosis (PET-MF) (35) • Primärveröffentlichung (17) • Europeam Public Assessment Report: Jakavi (8)
2	Studien- typ	Zuordnung zu einem der folgenden Studientypen (bitte ankreuzen): ☒ Therapiestudie mit randomisierter Vergleichsgruppe ☐ Therapiestudie mit nicht-randomisierter Vergleichsgruppe ☐ Therapiestudie mit Vergleichen aus früherer Zeit und / oder anderem Ort (z. B. historische Kontrollen) ☐ Fall-Kontroll-Studie ☐ Therapiestudie ohne (unabhängige) Vergleichsgruppe(inkl. „Vorher Nachher-Studien“) ☐ Fallserie ☐ Fallbericht / Kasuistik (case report) ☐ Nicht eindeutig zuzuordnen (ggf. Beschreibung mittels Originaltext): Details zum Studiendesign: Es handelt sich bei der COMFORT II-Studie um eine parallele, multizentrische, kontrollierte, offene, randomisierte Studie im Verhältnis Ruxolitinib zu BAT (Best-Available-Therapy) 2:1, stratifiziert nach Risikoprofil (Intermediär 2 [2 Risikofaktoren] oder Hochrisiko [3 oder mehr Risikofaktoren]) Erhebung des primären Endpunkts („Primäre Datenaanalyse, Primär-analyse“) nachdem alle Patienten 48 Wochen beendet haben bzw. ein Krankheitsprogress erreicht wurde. Eine Behandlung war bis zu 144 Wochen möglich (Nachbeobachtung auf Wunsch länger).
3	Einord- nung in die Evidenz- kategorie gemäß VerfO	Hier folgt eine formale Zuordnung zu den Evidenzstufen nach Verfahrensordnung (bitte ankreuzen): ☒ Ib: Randomisierte klinische Studien ☐ IIb: Prospektiv vergleichende Kohortenstudien ☐ III: Retrospektiv vergleichende Studien ☐ IV: Fallserien und nicht-vergleichende Studien ☐ V: Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte u. a.; nicht mit Studien belegte Meinungen anerkannter Experten, Bericht von Expertenkomitees und Konsenskonferenzen.
4	Kontext der Studie	<i>“This study was initiated by Incyte Corporation (Wilmington, DE, United States); due to an inlicense agreement, the change of study sponsorship was</i>

	I Interessen-konflikte	<i>submitted in February 2010. Novartis assumed responsibility for the study on 15-Mar-2010“</i>
5	Indikation	Ruxolitinib ist angezeigt für die Behandlung von krankheitsbedingter Splenomegalie oder Symptomen bei Erwachsenen mit primärer Myelofibrose (auch bekannt als chronische idiopatische Myelofibrose), Post-Polycythaemia-vera-Myelofibrose oder Post-Essentieller-Thrombozythämie-Myelofibrose.
6	Zielsetzung	<u>Primäres Studienziel:</u> <i>„To compare the efficacy, safety, and tolerability of Ruxolitinib given twice daily compared to BAT in subjects with PMF, PPV-MF or PET-MF“</i> <u>Sekundäre Studienziele:</u> <i>“To evaluate the population pharmacokinetics of Ruxolitinib”</i>
Methodik		
7	Ein- und Ausschlusskriterien	<u>Key inclusion criteria</u> 1. Subjects 18 years of age or older. 2. Subjects must have received the diagnosis of PMF, PPV-MF or PET-MF, according to the 2008 WHO criteria, regardless of JAK2 mutation status. 3. Subjects with MF requiring therapy must be classified as high risk (3 or more prognostic factors) OR Intermediate risk level 2 (2 or more prognostic factors). The prognostic factors, defined by the International Working Group are: • Age > 65 years • Presence of constitutional symptoms (weight loss, fever, night sweats) • Marked anaemia (Hemoglobin < 10g/dL)* • Leukocytosis (history of WBC > 25 x10 ⁹ /L) • Circulating blasts ≥ 1%. * For the purpose of evaluating risk factors, subjects receiving regular transfusions of packed red blood cells were considered to have hemoglobin < 10 g/dL. 4. Subjects with peripheral blood blast count of < 10%. 5. ECOG Performance-Status of 0, 1, 2 or 3. 6. Subjects with a palpable spleen measuring 5 cm or greater from the costal margin to the point of greatest splenic protrusion. 7. Subjects with a stable therapeutic regimen for at least 2 weeks before screening and no less than 4 weeks prior to baseline. <u>Key exclusion criteria</u> 1. Subjects with a life expectancy of less than 6 months. 2. Females who are pregnant or are currently breastfeeding. 3. Subjects with inadequate bone marrow reserve as demonstrated by: a) Absolute neutrophil count (ANC) ≤ 1000/μL. b) Platelet count < 100,000/μL without the assistance of growth factors, thrombopoietic factors or platelet transfusions. 4. Subjects with any history of platelet count < 50,000/μL or ANC < 500/μL except during treatment for a myeloproliferative disorder or treatment with cytotoxic therapy for any other reason. 5. Direct bilirubin > 2.0 x upper limit of normal (ULN). Alanine aminotransferase (ALT) > 2.5x institutional ULN. Creatinine > 2.0 mg/dL. 6. Subjects with clinically significant bacterial, fungal, parasitic or viral infection which require therapy 7. Subjects currently without the option of stem cell transplantation, either because they were not a candidate, or because a suitable donor was not available.

		8. Subjects with cardiac disease which in the Investigator's opinion could jeopardize the safety of the subject or the compliance with the protocol. 9. Subjects for whom the dose or dose regimen of any therapy used to treat MF was modified at any time from 2 weeks prior to the start of screening through the beginning of baseline evaluations. 10. Subjects who had splenic irradiation within 12 months prior to Screening. 11. Subjects undergoing treatment with hematopoietic growth factor receptor agonists (i.e. Epo, G-CSF, romiplostim, eltrombopag) at any time within 2 weeks prior to Screening or 4 weeks prior to baseline.
8	Prüfinter- vention	Ruxolitinib Tabletten (N = 146) 2 x täglich 15 mg (3 Tabletten) oder 2 x täglich 20 mg (4 Tabletten), Dosisanpassung war nach 4 Wochen möglich: Dosiserhöhung um 5 mg 2 x täglich bei unzureichender Wirksamkeit nach 4 Wochen bei • weniger als 40% Milzlängenabnahme und • Thrombozytenzahl > 150.000/µl und • ≥ 1000 µl seit Baseline Dosissenkung bei • Blutungsneigung: • Bei Neutrophilenzahl < 500/µl: Therapiepause • Bei Thrombozytenzahl < 50.000/µl: Therapiepause; 50.000-125.000/µl Dosisanpassung nach Thrombozytenzahl und Dosishöhe
9	Ver- gleichs- interven- tion(en)	Best-Available-Therapy (BAT, N=73): die nach dem Ermessen des jeweiligen Prüfarztes beste verfügbare Therapie zur Behandlung der Erkrankung und der damit einhergehenden Symptome, wobei hier auch ein Therapiewechsel, eine Kombinationen oder eine Nichtbehandlung möglich waren. Verboten war der Einsatz weiterer Prüfsubstanzen. Folgende Medikationen kamen infrage • Hydroxycarbamide, anagrelide (anti-neoplastic agents) • Prednisone, methylprednisone, prednisolone (glucocorticoids) • Epoetin alfa, an anti-anemic preparation • Thalidomide, lenalidomide (Immunosuppressants) • Mercaptopurine, thioguanine (purine analogs) • Danazol, an anti-gonadotropins or similar agent • Pegasys and Interferon-α (Interferons) • Melphalan (nitrogen mustard analog) • Acetylsalicylic acid (platelet aggregation inhibitor) • Colchicine, a preparation with no effect on uric acid metabolism
10	Prüfzen- tren	56 Zentren in 9 europäischen Ländern (Belgien, Deutschland, Frankreich, Großbritannien, Spanien, Italien, Niederlande, Österreich, Schweden)
11	Randomi- sierung	Die Randomisierungsliste wurde im Vorfeld der Studie automatisiert durch ein Computersystem mit Zufallsgenerator erstellt. Blockrandomisierung in 6er-Blöcken
12	Conceal- ment („Maskier- ung“ der	Die Patienten wurden zentral durch ein interaktives Spracherkennungssystem randomisiert, das die Behandlungsgruppe und die Patientennummer ausgab, mit der die Medikationsflaschen für die Ruxolitinib-Patienten gekennzeichnet war.

	Randomisierung)	„Sites will contact an Interactive Voice Response System (IVRS) to obtain the subject study drug assignment. The investigator or designee will select the assigned bottles from their stock that corresponds to the number provided by IVRS and dispense the medication. The investigator will enter the bottle numbers in the eCRF.“
13	Verblindung der Behandlung	Erfolgte eine Verblindung der Behandlung (bitte ankreuzen)? ☒ Nein, offene Behandlung ☐ Patienten verblindet ☐ Behandler verblindet ☐ Patienten und Behandler verblindet (bis zur Analyse der primären Daten, danach war ein Crossover von Placebo zu Ruxolitinib möglich) ☒ Auswerter verblindet (gilt für die Auswertung der MRT- und CT-Bilder als Grundlage für den primären Endpunkt „Milzvolumenreduktion“) ☐ Verblindung unklar
14	Beobachtungsdauer	- Study initiation date: 01-Jul-2009 (first subject first visit) - Data cut-off date: 04-Jan-2011 (last subject last visit) Figure 9-1 Study design * Dose was reassessed throughout for safety and at specific time points for both safety and efficacy with potential dose adjustments; see Study INC424A2352 Protocol – Section 5.3. ** Implemented with Protocol Amendment 3. A qualifying disease progression event (25% increase in spleen volume from on-study nadir or splenectomy) permits participation in the Extension Phase.
15	Erhebung der primären Zielkriterien	<u>Primäres Zielkriterium:</u> Anteil der Patienten mit einer Milzvolumenreduktion um $\geq 35\%$ nach 48 Wochen gemessen mittels MRT oder CT
16	Erhebung der sekundären Zielkriterien	<u>Weitere Zielkriterien:</u> - Anteil der Patienten mit einer Milzvolumenreduktion um $\geq 35\%$ nach 24 Wochen gemessen mittels MRT oder CT - Beginn bzw. Dauer des Ansprechens (Milzvolumenreduktion um $\geq 35\%$)

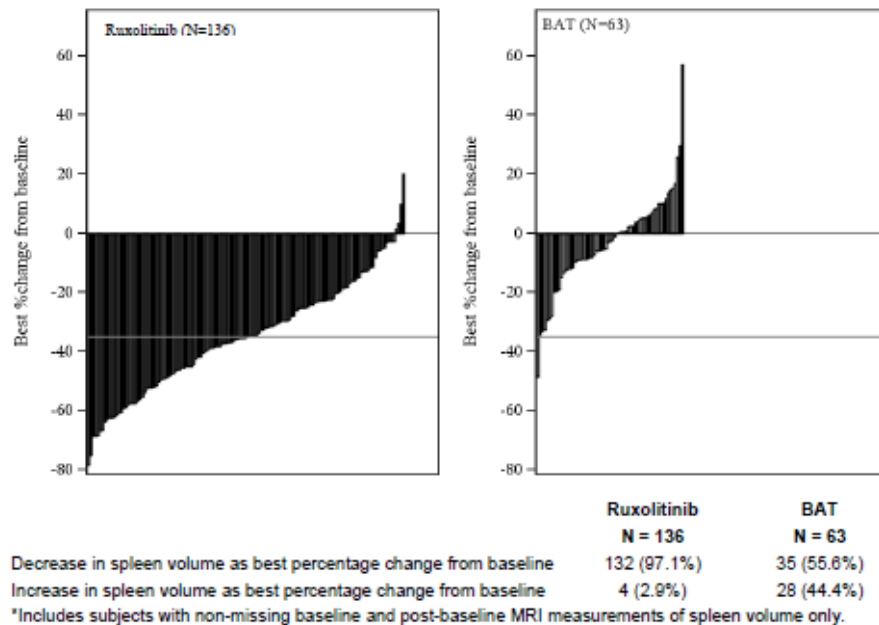
		• Prozentuale Milzvolumenveränderung • Gesamtüberleben • Leukämiefreies Überleben • Progressionsfreies Überleben • ECOG Performance-Status • QLQ-C30 der EORTC • Fact-Lym • Unerwünschte Ereignisse • Schwerwiegende unerwünschte Ereignisse • Unerwünschte Ereignisse Grad 3-4 der CTCAE • Abbruch aufgrund von unerwünschten Ereignissen • Dosisreduktion aufgrund von unerwünschten Ereignissen • Häufigste (schwere) unerwünschte Ereignisse (gesonderte Betrachtung der UEs Thrombozytämie, Anämie, Ekchymosen und Infektionen als UEs von besonderem Interesse) <u>Weitere Zielkriterien, die im Extraktionsbogen dargestellt werden, aber nicht explizit in die Bewertung eingehen</u> • Notwendigkeit von Thrombozyten- bzw. Erythrozytenkonzentraten • JAK2-Allellast <u>Hinweis: Die Ergebnisse der folgenden Zielkriterien werden im Datenextraktionsbogen in Feld 20 nicht dargestellt (s. dazu Tabelle 6) in der Nutzenbewertung</u> • Fibrosegrad • Plasma PD Marker • Körpergewicht • Weitere Auswertungen zu Nebenwirkungen • Laborparameter • Vitalzeichen • Pharmakokinetische Ergebnisse
17	Anzahl der einzuschließen den Patienten pro Gruppe	Fallzahl- / Powerberechnung durchgeführt? ☒ ja ☐ nein ☐ keine Angabe Unter Annahme einer Ansprechrate von 35% vs. 10% in der Vergleichsgruppe ergibt eine Fallzahl von 150 Patienten (100 + 50) im zweiseitigen Chi²-Test bei einem alpha von 0,05 eine Power von 90%.
18	Statistische Auswertung	Angaben zur statistischen Auswertung: ☒ ITT-Analyse (primärer Endpunkt) ☐ per-Protokoll-Analyse (zusätzlich, hier im Bogen nicht dargestellt) ☒ direkter Vergleich der Behandlungsgruppen ☒ Umgang mit fehlenden Daten (missings) beschrieben* ☒ präspezifizierte Subgruppenanalysen (für den primären

		Studienendpunkt) ☐ Interaktionstests x vorgesehene statistische Auswertung hinreichend beschrieben *Angaben aus dem Studienprotokoll: <i>"For subjects who dropped out due to lack of efficacy or treatment related adverse events,</i> <i>• Reduction of spleen volume will be considered as less than 35%.</i> <i>• No last observation carried forward will be used for primary and secondary efficacy analyses.</i> <i>Handling of QoL values will be according to the QLQ-C30 and FACT-lym manuals, and is detailed in the Report and Analysis Plan (RAP). For subjects who missed one scheduled visit that included a spleen volume assessment (MRI or CT), data from subsequent visits will be included in assessment of progression free survival (PFS). For subjects missing more than one consecutive visit in which assessment of spleen volume was scheduled, data handling will be described in Section 12.2.3."</i>																											
		Ergebnisse																											
19	Patientencharakteristika und -fluss	Patientencharakteristika und Vergleichbarkeit der Behandlungsgruppen: x Charakteristika der Behandlungsgruppen dargestellt ☐ relevante Unterschiede zwischen den Gruppen beschrieben Wesentlich Patientencharakteristika <table border="1"> <thead> <tr> <th></th><th>Ruxolitinib (N = 146)</th><th>Placebo (N = 73)</th></tr> </thead> <tbody> <tr> <td>Alter Median (min-max)</td><td>66 (35-83)</td><td>66 (35-85)</td></tr> <tr> <td>≤ 65 Jahre (%)</td><td>47,2</td><td>49,3</td></tr> <tr> <td>> 65 Jahre (%)</td><td>52,8</td><td>50,7</td></tr> <tr> <td>Geschlecht Männer (%) Frauen (%)</td><td>56,8 43,2</td><td>57,5 42,5</td></tr> <tr> <td>Krankheitstyp PMF (%) PPV-MF (%) PET-MF (%)</td><td>2,7 32,9 14,4</td><td>53,4 27,4 19,2</td></tr> <tr> <td>ECOG Performance-Status, N (%) 0 1 2 3</td><td>58 (39,7) 77 (52,7) 10 (6,8) 1 (0,7)</td><td>26 (35,6) 37 (50,7) 9 (12,3) 1 (1,4)</td></tr> <tr> <td>Vorangegangene Therapie mit Hydroxyurea (%)</td><td>75,3</td><td>68,5</td></tr> <tr> <td>Tastbare Milzlänge, cm (min-max)</td><td>14 (5-30)</td><td>15 (5-37)</td></tr> </tbody> </table>		Ruxolitinib (N = 146)	Placebo (N = 73)	Alter Median (min-max)	66 (35-83)	66 (35-85)	≤ 65 Jahre (%)	47,2	49,3	> 65 Jahre (%)	52,8	50,7	Geschlecht Männer (%) Frauen (%)	56,8 43,2	57,5 42,5	Krankheitstyp PMF (%) PPV-MF (%) PET-MF (%)	2,7 32,9 14,4	53,4 27,4 19,2	ECOG Performance-Status, N (%) 0 1 2 3	58 (39,7) 77 (52,7) 10 (6,8) 1 (0,7)	26 (35,6) 37 (50,7) 9 (12,3) 1 (1,4)	Vorangegangene Therapie mit Hydroxyurea (%)	75,3	68,5	Tastbare Milzlänge, cm (min-max)	14 (5-30)	15 (5-37)
	Ruxolitinib (N = 146)	Placebo (N = 73)																											
Alter Median (min-max)	66 (35-83)	66 (35-85)																											
≤ 65 Jahre (%)	47,2	49,3																											
> 65 Jahre (%)	52,8	50,7																											
Geschlecht Männer (%) Frauen (%)	56,8 43,2	57,5 42,5																											
Krankheitstyp PMF (%) PPV-MF (%) PET-MF (%)	2,7 32,9 14,4	53,4 27,4 19,2																											
ECOG Performance-Status, N (%) 0 1 2 3	58 (39,7) 77 (52,7) 10 (6,8) 1 (0,7)	26 (35,6) 37 (50,7) 9 (12,3) 1 (1,4)																											
Vorangegangene Therapie mit Hydroxyurea (%)	75,3	68,5																											
Tastbare Milzlänge, cm (min-max)	14 (5-30)	15 (5-37)																											

		IWG-MRT-Risikostufe													
		3 (hoch), %	60,3	58,9											
		2 (intermediär 2), %	39,7	39,7											
		*: Daten von mehreren Patienten fehlen aus Datenschutzgründen bzw. gingen verloren													
		<i>“The study population was well balanced with respect to demographics. The median age at the time of randomization was 66 years and was similar between treatment arms. Overall, PMF, PPV-MF and PET-MF were diagnosed in 53.0%, 31.1%, and 16.0% of subjects, respectively. The frequency of PMF was similar between ruxolitinib and BAT arms (52.7% and 53.4% in the ruxolitinib and BAT arms, respectively). In addition, the frequencies of PPV-MF (ruxolitinib: 32.9%; BAT: 27.4%) and PET-MF (ruxolitinib: 14.4%; BAT: 19.2%) were also similar between both treatment arms. There were no remarkable differences between ruxolitinib- and BAT-treated subjects with respect to baseline spleen size and volume. Prior use of HU was similar between treatment arms. Prior use of HU was similar between treatment arms.”</i>													
		Relevante Unterschiede zwischen den Gruppen:													
		Es gab keine wesentlich / relevanten Unterschiede zwischen den Gruppen.													
		Patientenfluss und Studienausscheider („drop outs“):													
		x	Studienausscheider (nach Randomisierung) dokumentiert bzw. plausibel beschrieben												
		x	Umgang mit Studienausscheidern in der Analyse begründet												
	☐ bei fehlender ITT-Analyse: Hinweise auf systematische und relevante Unterschiede zwischen den Studienausscheidern und den gemäß Studienprotokoll behandelten Patienten														
		Analysenpopulationen													
		<table><tr><td>Analysenpopulation</td><td>Ruxolitinib (N = 146), N (%)</td><td>BAT (N = 73) N (%)</td></tr><tr><td>Full Analysis Set (FAS)*</td><td>146 (100)</td><td>73 (100)</td></tr><tr><td>Per Protocoll Set</td><td>137 (93,8)</td><td>65 (89,0)</td></tr><tr><td>Safety Set</td><td>146 (100)</td><td>73 (100)</td></tr></table>		Analysenpopulation	Ruxolitinib (N = 146), N (%)	BAT (N = 73) N (%)	Full Analysis Set (FAS)*	146 (100)	73 (100)	Per Protocoll Set	137 (93,8)	65 (89,0)	Safety Set	146 (100)	73 (100)
Analysenpopulation	Ruxolitinib (N = 146), N (%)	BAT (N = 73) N (%)													
Full Analysis Set (FAS)*	146 (100)	73 (100)													
Per Protocoll Set	137 (93,8)	65 (89,0)													
Safety Set	146 (100)	73 (100)													
		* In der vorliegenden Bewertung werden die Begriffe FAS und ITT synonym verwendet.													
		Patientenfluss													
		<table><tr><td></td><td>Ruxolitinib (N = 146), N (%)</td><td>BAT (N = 73), N (%)</td></tr><tr><td>In der randomisierten Studie oder Eintritt in die Extentionsphase</td><td>120 (82,2)</td><td>49 (67,1)</td></tr><tr><td>In der randomisierten Studie</td><td>91 (62,3)</td><td>31 (42,5)</td></tr><tr><td>Nicht mehr in der randomisierten Studie</td><td>55 (37,7)</td><td>42 (57,5)</td></tr></table>			Ruxolitinib (N = 146), N (%)	BAT (N = 73), N (%)	In der randomisierten Studie oder Eintritt in die Extentionsphase	120 (82,2)	49 (67,1)	In der randomisierten Studie	91 (62,3)	31 (42,5)	Nicht mehr in der randomisierten Studie	55 (37,7)	42 (57,5)
	Ruxolitinib (N = 146), N (%)	BAT (N = 73), N (%)													
In der randomisierten Studie oder Eintritt in die Extentionsphase	120 (82,2)	49 (67,1)													
In der randomisierten Studie	91 (62,3)	31 (42,5)													
Nicht mehr in der randomisierten Studie	55 (37,7)	42 (57,5)													
		Gründe für einen Abbruch der randomisierten Behandlungsphase													

		<table> <tr> <td>Eintritt in die Extensionsphase*</td><td>29 (19,9)</td><td>18 (24,7)</td></tr> <tr> <td>Unerwünschte Ereignisse</td><td>12 (8,2)</td><td>4 (5,5)</td></tr> <tr> <td>Einwilligung zurückgezogen</td><td>2 (1,4)</td><td>9 (12,3)</td></tr> <tr> <td>Krankheitsprogression</td><td>1 (0,7)</td><td>3 (4,1)</td></tr> <tr> <td>Protokollverletzung</td><td>4 (2,8)</td><td>0</td></tr> <tr> <td>Andere</td><td>7 (4,8)</td><td>8 (11,0)</td></tr> </table> * Das Auftreten eines im Studienprotokoll definierten Krankheitsprogresses qualifizierte die Patienten für den Eintritt in die Extensionsphase	Eintritt in die Extensionsphase*	29 (19,9)	18 (24,7)	Unerwünschte Ereignisse	12 (8,2)	4 (5,5)	Einwilligung zurückgezogen	2 (1,4)	9 (12,3)	Krankheitsprogression	1 (0,7)	3 (4,1)	Protokollverletzung	4 (2,8)	0	Andere	7 (4,8)	8 (11,0)						
Eintritt in die Extensionsphase*	29 (19,9)	18 (24,7)																								
Unerwünschte Ereignisse	12 (8,2)	4 (5,5)																								
Einwilligung zurückgezogen	2 (1,4)	9 (12,3)																								
Krankheitsprogression	1 (0,7)	3 (4,1)																								
Protokollverletzung	4 (2,8)	0																								
Andere	7 (4,8)	8 (11,0)																								
20	Ergebnisse	Primärer Endpunkt: Anzahl der Patienten mit einem Milzvolumenrückgang $\geq 35\%$ gegenüber Studienbeginn nach 48 Wochen (ITT) <table> <tr> <th></th><th>Ruxolitinib (N = 144)</th><th>BAT (N = 72)</th></tr> <tr> <td>Anzahl der Responder, N (%)</td><td>41 (28,5)</td><td>0</td></tr> <tr> <td>95% KI</td><td>21,3-36,6</td><td>0-5</td></tr> <tr> <td>p-Wert *</td><td colspan="2">p < 0,0001</td></tr> </table> * Berechnet mit dem CMH-Test <u>Sekundäre Endpunkte:</u> <u>Anzahl der Patienten mit einem Milzvolumenrückgang $\geq 35\%$ gegenüber Studienbeginn nach 24 Wochen</u> <table> <tr> <th></th><th>Ruxolitinib (N = 144)</th><th>BAT (N = 72)</th></tr> <tr> <td>Anzahl der Responder, N (%)</td><td>46 (31,9)</td><td>0</td></tr> <tr> <td>95% KI</td><td>24,4-40,2</td><td>0-5</td></tr> <tr> <td>p-Wert *</td><td colspan="2">p < 0,0001</td></tr> </table> * Berechnet mit dem CMH-Test <u>Milzvolumenveränderung (dargestellt ist die jeweils beste Veränderung gegenüber Studienbeginn gemessen mit MRT oder CT)</u>		Ruxolitinib (N = 144)	BAT (N = 72)	Anzahl der Responder, N (%)	41 (28,5)	0	95% KI	21,3-36,6	0-5	p-Wert *	p < 0,0001			Ruxolitinib (N = 144)	BAT (N = 72)	Anzahl der Responder, N (%)	46 (31,9)	0	95% KI	24,4-40,2	0-5	p-Wert *	p < 0,0001	
	Ruxolitinib (N = 144)	BAT (N = 72)																								
Anzahl der Responder, N (%)	41 (28,5)	0																								
95% KI	21,3-36,6	0-5																								
p-Wert *	p < 0,0001																									
	Ruxolitinib (N = 144)	BAT (N = 72)																								
Anzahl der Responder, N (%)	46 (31,9)	0																								
95% KI	24,4-40,2	0-5																								
p-Wert *	p < 0,0001																									

Figure 11-1 Best percentage change from baseline in spleen volume measured by MRI/CT (FAS)*



Weitere sekundäre Endpunkte:

Dauer des Ansprechens (Studienbericht, S. 125)

„The median duration of maintenance (i.e. the period beyond the point at which a $\geq 35\%$ reduction was observed) of spleen volume reduction in ruxolitinib-treated subjects was 48 weeks; the Kaplan-Meier estimate of DoMSR at the 24-week time point for the ruxolitinib treatment arm was 0.67. In contrast, on the BAT arm, since only one subject was reported with a spleen volume measurement at least once below threshold levels and was censored, neither the calculation of the median duration nor the Kaplan-Meier estimates for duration of maintenance could be performed.”

Table 11-7 Duration of maintenance of at least a 35% reduction in spleen volume from baseline (FAS)

	Ruxolitinib N = 69	BAT N = 1
No. of events	26 (37.7%)	0
No. censored	43 (62.3%)	1 (100%)*
Median (95% CI), weeks	48.0 (35.86 - N/A)	N/A
Kaplan-Meier estimates of duration of maintenance (95% CI)		
12 weeks	0.78 (0.66, 0.87)	N/A
24 weeks	0.67 (0.54, 0.77)	N/A
36 weeks	0.58 (0.44, 0.70)	N/A
48 weeks	0.39 (0.10, 0.67)	N/A

* One subject in BAT arm had one post-baseline MRI showing $\geq 35\%$ reduction in spleen volume and no MRIs thereafter.

Duration of maintenance of spleen volume reduction is defined as the longest duration of consecutive measurements of $\geq 35\%$ reduction observed prior to the time of database freeze for subjects who have at least one measured $\geq 35\%$ reduction.

Source: Table 14.2-2.4

Beginn des Ansprechens (Studienbericht, S. 126)

Table 11-9 Time to first at least 35% reduction in spleen volume from baseline by treatment (FAS)

	Ruxolitinib N= 69	BAT N= 1
Median (95% CI), weeks	12.29 (12.14 - 14.43)	15.43 (N/A - N/A)
Minimum, weeks	11.00	15.43
Maximum, weeks	49.14	15.43
Kaplan-Meier estimates of time to first 35% reduction (95% CI)		
12 Weeks	0.23 (0.14, 0.34)	0.00 (N/A, N/A)
24 Weeks	0.67 (0.54, 0.76)	1.00 (N/A, N/A)
36 Weeks	0.87 (0.76, 0.93)	1.00 (N/A, N/A)
48 Weeks	0.97 (0.89, 0.99)	1.00 (N/A, N/A)
N/A: Not available		
Source: Table 14.2-2.5		

Weitere exploratorische Endpunkte:

Prozentuale Veränderung der Milzlänge und des Milzvolumens (Studienbericht, S. 129)

„Treatment with ruxolitinib resulted in a larger reduction in spleen volume than what was observed with BAT with a mean percent change from baseline at **Week 48** of $-30.05 \pm 22.12\%$ with ruxolitinib vs. $7.34 \pm 12.99\%$ with BAT (Table 14.2-1.2). The difference in change in spleen volume was statistically significant ($p < 0.0001$).

Treatment with ruxolitinib resulted in a larger reduction in spleen volume than what was observed with BAT with a mean percent change from baseline at **Week 24** of $-29.23 \pm 19.48\%$ with ruxolitinib vs. $2.70 \pm 14.12\%$ with BAT (Table 14.2-2.2). The difference in change in spleen volume was statistically significant ($p < 0.0001$).

Notable differences were also observed with respect to the change in palpable spleen length below the left costal margin. Ruxolitinib-treated subjects exhibited a decrease in spleen length below the left costal margin at each time point between Weeks 4 and 48 (-5.7 ± 4.01 cm and -7.3 ± 5.01 cm, respectively). This decrease was not observed in the BAT arm, for which negligible change in spleen size below the left costal margin was reported from Weeks 4 to 48 (Week 4: -0.2 ± 4.30 cm; Week 48: 0.6 ± 3.47 cm).“

JAK2-Allellast (Studienbericht, S. 129)

„There was a modest decrease in allele burden (percent V617F) with ruxolitinib treatment which was not seen in the BAT arm, with median changes in allele burden of -6.0% ($N = 76$) and -7.0% ($N = 60$) at Weeks 24 and 48 in the ruxolitinib arm versus 0.5% ($N = 22$) and 0.0% ($N = 22$) with BAT.“

Notwendigkeit von Transfusionen (Studienbericht, S. 129)

„Subjects in both ruxolitinib and BAT arms received more packed red blood cell transfusions than platelet transfusions (Table 11-12). More subjects in the BAT treatment arm received red blood cell transfusions during the first 4 weeks compared to the ruxolitinib arm; after Week 4, more ruxolitinib-treated subjects received red blood cell transfusions. The frequency of subjects receiving any red blood cell transfusions or two or more transfusions per month varied during each time period; however, there are approximately 5% more subjects on the ruxolitinib arm requiring red cell transfusions during most time periods relative to the BAT arm.“

Table 11-12 Number (%) of subjects who received any blood component transfusions by time and treatment (FAS)

Blood component/Time period	Ruxolitinib N = 146		BAT N = 73	
	Any units n (%)	≥2 units n (%)	Any units n (%)	≥ 2 units n (%)
Packed red blood cells				
> 0 weeks - ≤ 4 weeks	20 (13.7)	17 (11.6)	20 (27.4)	19 (26.0)
> 4 weeks - ≤ 8 weeks	49 (33.6)	47 (32.2)	21 (28.8)	21 (28.8)
> 8 weeks - ≤ 12 weeks	40 (27.4)	36 (24.7)	16 (21.9)	16 (21.9)
> 12 weeks - ≤ 16 weeks	50 (34.2)	47 (32.2)	20 (27.4)	20 (27.4)
> 16 weeks - ≤ 24 weeks	51 (34.9)	47 (32.2)	19 (26.0)	19 (26.0)
> 24 weeks - ≤ 36 weeks	46 (31.5)	42 (28.8)	17 (23.3)	15 (20.5)
> 36 weeks - ≤ 48 weeks	32 (21.9)	32 (21.9)	10 (13.7)	9 (12.3)
> 48 weeks - ≤ 60 weeks	21 (14.4)	20 (13.7)	8 (11.0)	7 (9.6)
> 60 weeks	3 (2.1)	3 (2.1)	0	0
Platelets				
> 0 weeks - ≤ 4 weeks	0	0	1 (1.4)	1 (1.4)
> 4 weeks - ≤ 8 weeks	1 (0.7)	1 (0.7)	1 (1.4)	1 (1.4)
> 8 weeks - ≤ 12 weeks	0	0	1 (1.4)	1 (1.4)
> 12 weeks - ≤ 16 weeks	1 (0.7)	1 (0.7)	2 (2.7)	2 (2.7)
> 16 weeks - ≤ 24 weeks	2 (1.4)	2 (1.4)	1 (1.4)	1 (1.4)
> 24 weeks - ≤ 36 weeks	0	0	2 (2.7)	1 (1.4)
> 36 weeks - ≤ 48 weeks	2 (1.4)	2 (1.4)	0	0
> 48 weeks - ≤ 60 weeks	0	0	0	0
> 60 weeks	0	0	0	0

EORTC QLQ-C30 (Studienbericht, S. 132)

“At baseline, there were no clinically meaningful differences between the ruxolitinib and BAT groups for the individual subscales of the EORTC QLQ-C30 with the exception of insomnia, which had a greater than 10-point difference between ruxolitinib vs. BAT, 38.0 vs. 24.9 respectively. A 10 point difference is equivalent to 10% of the scale (0 - 100) and is considered clinically meaningful (Osoba, et al 1998).

Based on descriptive analysis of the EORTC QLQ-C30 outcomes, in general, subjects in the ruxolitinib arm, relative to the BAT arm, had more improvement in overall QoL, functioning, and symptoms. For global health status/QoL, role functioning, fatigue, pain, dyspnea, insomnia, appetite loss, and diarrhea subscales, the difference in mean change from baseline score between the ruxolitinib and BAT arms was more than 10 points (10% of the scale range) in favor of ruxolitinib for at least one time point between Weeks 8 and 48.”

Table 11-15 Summary of Change from Baseline Scores for Selected* EORTC QLQ-C30 Subscales (FAS)

		Ruxolitinib N=146		BAT N=73	
		n	Mean (SD)	n	Mean (SD)
Global health status/QoL	Week 8	109	8.5 (20.57)	46	4.5 (20.77)
	Week 16	107	8.5 (20.02)	36	-3.0 (21.84)
	Week 24	95	8.4 (18.57)	37	5.2 (23.76)
	Week 48	66	9.1 (24.23)	27	3.4 (24.27)
Role Functioning	Week 8	115	8.0 (25.97)	48	-5.2 (22.59)
	Week 16	112	5.4 (28.05)	39	-10.3 (24.96)
	Week 24	96	5.4 (27.52)	39	-6.0 (27.96)
	Week 48	69	9.9 (28.05)	28	-5.4 (26.47)
Fatigue	Week 8	115	-9.7 (24.40)	48	-2.1 (21.46)
	Week 16	112	-5.9 (22.86)	40	3.8 (20.65)
	Week 24	98	-8.1 (23.26)	39	-1.0 (23.57)
	Week 48	69	-12.8 (28.57)	28	0.4 (25.39)
Pain	Week 8	113	-6.6 (25.25)	48	-3.5 (22.27)
	Week 16	111	-4.7 (25.34)	40	7.9 (36.40)
	Week 24	98	-5.4 (24.50)	39	3.0 (28.06)
	Week 48	69	-1.9 (28.80)	28	3.0 (17.60)
Dyspnea	Week 8	114	-7.0 (24.08)	48	4.9 (23.81)
	Week 16	112	-4.5 (28.12)	40	2.5 (29.61)
	Week 24	97	-1.7 (28.61)	38	5.3 (31.51)
	Week 48	69	-6.3 (30.40)	28	4.8 (34.80)
Insomnia	Week 8	114	-14.3 (31.97)	48	-4.2 (22.41)
	Week 16	112	-10.4 (32.89)	40	5.0 (30.71)
	Week 24	97	-9.3 (31.08)	39	6.0 (28.48)
	Week 48	68	-12.3 (35.46)	28	6.0 (24.09)
Appetite Loss	Week 8	114	-10.8 (24.89)	48	3.5 (24.06)
	Week 16	112	-8.0 (28.03)	40	3.3 (31.85)
	Week 24	98	-10.9 (24.77)	39	0.9 (34.61)
	Week 48	69	-8.2 (27.06)	28	9.5 (33.77)
Diarrhea	Week 8	109	-2.1 (27.69)	47	6.4 (25.67)
	Week 16	107	-4.0 (24.96)	38	14.0 (33.45)
	Week 24	95	-5.3 (25.87)	37	2.7 (31.80)
	Week 48	65	-4.1 (23.20)	27	-6.2 (24.52)

* Subscales with no treatment difference (i.e. difference in mean change scores <10 points at all post-baseline visits) are not presented.

FACT-Lym (Studienbericht, S. 134)

"LymS, FACT-Lym TOI, FACT-G total, and FACT-Lym total scores were summarized by treatment at each time point (Table 11-16). The mean change from baseline scores were observed to be higher (indicating better QoL) in the ruxolitinib arm compared to the BAT arm at Week 8 and remained so through Week 48. The difference between treatments ranges from 4.7 (7.8% of the scale range) for the LymS to 13.8 (8.2% of the scale range) for FACT-Lym total scale. The scores for the FACT-Lym total score, FACT-TOI consistently worsened for the BAT group through week 48 while they improved and then stabilized for the ruxolitinib group during the same time period."

Table 11-16 Summary of Change from Baseline Scores for Select FACT-Lym Subscales (FAS)

	Score Range	Ruxolitinib N=146		BAT N=73	
		n	Mean (SD)	n	Mean (SD)
FACT-Lym Total Score	0 - 168				
Week 8		116	12.1 (22.1)	50	-0.1 (19.9)
Week 16		116	10.5 (22)	42	-3.3 (13.4)
Week 24		100	11.9 (20.3)	40	-0.8 (17.1)
Week 48		70	11.3 (25.0)	29	-0.9 (19.6)
FACT-TOI	0 - 116				
Week 8		116	10.1 (17.0)	50	0.0 (14.4)
Week 16		116	8.7 (15.9)	42	-2.7 (10)
Week 24		100	9.0 (14.2)	40	-0.4 (13.8)
Week 48		70	9.1 (18.8)	29	-0.9 (16.1)
FACT-G Total Score	0 - 108				
Week 8		116	6.3 (15.0)	50	-0.1 (12.5)
Week 16		116	5.3 (16.4)	42	-2.9 (10.4)
Week 24		100	6.3 (17.2)	40	-1.1 (11)
Week 48		68	8.9 (15.9)	29	0.1 (13.8)
LymS	0 - 60				
Week 8		112	5.9 (7.3)	49	1.2 (6.3)
Week 16		113	5.7 (7.3)	42	-0.4 (5.99)
Week 24		99	6.1 (6.7)	40	0.3 (8.4)
Week 48		67	6.0 (7.6)	28	0.7 (9.1)

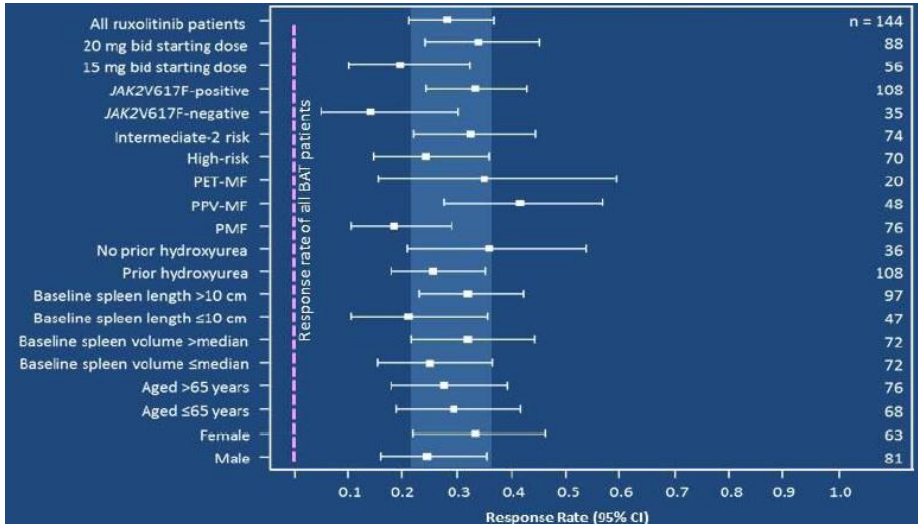
Source: Table 14.2.4.4

ECOG Performance-Status (Studienbericht, S. 135)

“ECOG status was similar between subjects in the ruxolitinib arm and those in the BAT arm at baseline (Table 11-17). More than 35% of subjects in each arm had ECOG status of 0 (best status) and approximately half of the subjects had ECOG status of 1. Compared with the BAT arm, the ruxolitinib arm had a higher percentage of subjects with ECOG status of 0 at Week 4, both among all randomized subjects and among only subjects with the Week 4 assessment. This was also observed at Weeks 8, 16, and 24. At Week 48, the ruxolitinib arm had a higher percentage of subjects with ECOG status of 0 or 1 than the BAT arm; there were 43.2% and 58.9% missing evaluations for the ruxolitinib and BAT arms, respectively.”

Table 11-17 Subjects with ECOG status of 0 or 1 by time and treatment (FAS)

Timepoint	ECOG Status	Ruxolitinib N=146 n (%)	BAT N=73 n (%)
Baseline	0	58 (39.7)	26 (35.6)
	1	77 (52.7)	37 (50.7)
	missing	0 (0)	0 (0)
Week 4	0	71 (48.6)	25 (34.2)
	1	57 (39.0)	29 (39.7)
	missing	11 (7.5)	10 (13.7)
Week 8	0	72 (49.3)	25 (34.2)
	1	51 (34.9)	27 (37.0)
	missing	17 (11.6)	14 (19.2)
Week 16	0	65 (44.5)	16 (21.9)
	1	53 (36.3)	26 (35.6)
	missing	21 (14.40)	25 (34.2)
Week 24	0	61 (41.8)	13 (17.8)
	1	41 (28.1)	22 (30.1)
	missing	42 (28.80)	35 (47.9)
Week 48	0	52 (35.6)	19 (26.0)
	1	27 (18.5)	8 (8.2)
	missing	63 (43.2)	43 (58.9)

		Subgruppenergebnisse (Studienbericht, S. 137) Subgroup analyses were also performed on the primary endpoint; general conclusions could only be made for the ruxolitinib treatment arm given that no BAT-treated subjects reported a $\geq 35\%$ reduction in spleen volume from baseline at Week 48. Differences in the ruxolitinib treatment arm were observed with regards to - Gender: more female subjects met the primary endpoint (female: 33.3%; male: 24.7%). - MF type: more PPV-MF subjects achieved a $\geq 35\%$ reduction in spleen volume from baseline at Week 48 than PMF or PET MF subjects (41.7%, 18.4%, and 35.0%, respectively). - Baseline prognostic category as defined at randomization: more Intermediate risk-2 subjects met the primary endpoint than High risk subjects (32.4% vs. 24.3%, respectively). - Platelet count: subjects with a platelet count $> 200K$ were more frequently reported to exhibit a $\geq 35\%$ reduction in spleen volume from baseline at Week 48 ($> 200K$: 35.2%; $\leq 200K$: 17.9%). - Baseline spleen volume: more subjects with a baseline spleen volume above the median met the primary endpoint ($> \text{median}$: 31.9%; $\leq \text{median}$: 25.0%). - Previous HU use: subjects not previously treated with HU were more frequently observed to achieve a $\geq 35\%$ reduction in spleen volume from baseline at Week 48 (prior HU use: 25.9%; no prior HU use: 36.1%). There was no difference between age groups in achieving a $\geq 35\%$ reduction in spleen volume from baseline at Week 48 (≤ 65 years: 29.4%; > 65 years: 27.6%). In addition, no conclusions could be made for the influence of race due to the fact that the vast majority of Ruxolitinib-treated subjects were Caucasian."  Quelle: Dossier, Modul 4, S. 89
21	Unerwünschte Therapie-wirkungen	Unerwünschte Ereignisse nach Systemorganklasse (Studienbericht, S. 142) "The overall frequencies of AEs of all grades for any primary system organ class (SOC) were higher in the ruxolitinib treatment arm (99.3%) than in the BAT arm (90.4%); median exposure durations were 51.4 and 45.1 weeks, respectively. Grade 3-4 AEs were more frequently observed in the ruxolitinib

arm compared to the BAT arm (ruxolitinib: 43.2%; BAT: 24.7%).

Table 12-5 Adverse events regardless of study drug relationship by system organ class (Safety set)

Primary system organ class	Ruxolitinib N=146		BAT N=73	
	All grades n (%)	Grades 3-4 n (%)	All grades n (%)	Grades 3-4 n (%)
Any primary system organ class	145 (99.3)	63 (43.2)	66 (90.4)	18 (24.7)
Blood and lymphatic system disorders	103 (70.5)	30 (20.5)	18 (24.7)	7 (9.6)
Infections and infestations	93 (63.7)	19 (13.0)	32 (43.8)	3 (4.1)
Gastrointestinal disorders	84 (57.5)	11 (7.5)	36 (49.3)	7 (9.6)
General disorders and administration site conditions	79 (54.1)	10 (6.8)	36 (49.3)	4 (5.5)
Musculoskeletal and connective tissue disorders	70 (47.9)	11 (7.5)	25 (34.2)	0
Respiratory, thoracic and mediastinal disorders	56 (38.4)	3 (2.1)	32 (43.8)	7 (9.6)
Investigations	49 (33.6)	9 (6.2)	20 (27.4)	1 (1.4)
Skin and subcutaneous tissue disorders	48 (32.9)	1 (0.7)	28 (38.4)	1 (1.4)
Nervous system disorders	42 (28.8)	5 (3.4)	15 (20.5)	0
Vascular disorders	31 (21.2)	2 (1.4)	11 (15.1)	2 (2.7)
Cardiac disorders	29 (19.9)	5 (3.4)	8 (11.0)	2 (2.7)
Injury, poisoning and procedural complications	23 (15.8)	5 (3.4)	6 (8.2)	0
Metabolism and nutrition disorders	22 (15.1)	2 (1.4)	10 (13.7)	1 (1.4)
Psychiatric disorders	20 (13.7)	0	8 (11.0)	0
Renal and urinary disorders	19 (13.0)	3 (2.1)	12 (16.4)	4 (5.5)
Eye disorders	15 (10.3)	1 (0.7)	9 (12.3)	0
Ear and labyrinth disorders	10 (6.8)	0	3 (4.1)	1 (1.4)
Hepatobiliary disorders	8 (5.5)	3 (2.1)	7 (9.6)	3 (4.1)
Neoplasms benign, malignant and unspecified	8 (5.5)	1 (0.7)	6 (8.2)	3 (4.1)
Reproductive system and breast disorders	7 (4.8)	0	1 (1.4)	0
Immune system disorders	0	0	1 (1.4)	0

(Häufigste) unerwünschte Ereignisse (alle Grade und Grad 3 / 4 nach Preferred Terms (Studienbericht, S. 144))

Table 12-6 Frequent adverse events (at least 5% in any group) regardless of study drug relationship by preferred term (Safety set)

Preferred term	Ruxolitinib N=146		BAT N=73	
	All grades n (%)	Grades 3-4 n (%)	All grades n (%)	Grades 3-4 n (%)
Any preferred term	145 (99.3)	61 (41.8)	66 (90.4)	18 (24.7)
Thrombocytopenia	65 (44.5)	11 (7.5)	7 (9.6)	3 (4.1)
Anaemia	60 (41.1)	16 (11.0)	9 (12.3)	3 (4.1)
Diarrhoea	34 (23.3)	2 (1.4)	9 (12.3)	0
Oedema peripheral	32 (21.9)	0	19 (26.0)	0
Asthenia	26 (17.8)	2 (1.4)	7 (9.6)	1 (1.4)
Dyspnoea	23 (15.8)	1 (0.7)	13 (17.8)	3 (4.1)
Nasopharyngitis	23 (15.8)	0	10 (13.7)	0
Pyrexia	20 (13.7)	3 (2.1)	7 (9.6)	0
Cough	20 (13.7)	0	11 (15.1)	1 (1.4)
Nausea	19 (13.0)	1 (0.7)	5 (6.8)	0
Arthralgia	18 (12.3)	1 (0.7)	5 (6.8)	0
Fatigue	18 (12.3)	1 (0.7)	6 (8.2)	0
Pain in extremity	17 (11.6)	1 (0.7)	3 (4.1)	0
Abdominal pain	16 (11.0)	5 (3.4)	10 (13.7)	2 (2.7)
Headache	15 (10.3)	2 (1.4)	3 (4.1)	0
Bronchitis	14 (9.6)	2 (1.4)	5 (6.8)	1 (1.4)
Weight increased	14 (9.6)	3 (2.1)	0	0
Back pain	14 (9.6)	3 (2.1)	8 (11.0)	0
Muscle spasms	13 (8.9)	0	5 (6.8)	0
Vomiting	13 (8.9)	0	1 (1.4)	0
Epistaxis	12 (8.2)	0	5 (6.8)	0
Haematoma	12 (8.2)	0	3 (4.1)	0
Abdominal pain upper	12 (8.2)	0	4 (5.5)	0
Gastroenteritis	11 (7.5)	2 (1.4)	0	0
Night sweats	11 (7.5)	0	6 (8.2)	0
Constipation	11 (7.5)	0	4 (5.5)	0
Dizziness	11 (7.5)	0	4 (5.5)	0
Platelet count decreased	10 (6.8)	1 (0.7)	2 (2.7)	0
Dyspnoea exertional	10 (6.8)	1 (0.7)	2 (2.7)	0
Urinary tract infection	10 (6.8)	0	2 (2.7)	0
Paraesthesia	10 (6.8)	0	4 (5.5)	0
Bone pain	9 (6.2)	2 (1.4)	4 (5.5)	0
Insomnia	9 (6.2)	0	5 (6.8)	0
Cystitis	8 (5.5)	0	3 (4.1)	0
Pruritus	7 (4.8)	0	9 (12.3)	0
Dyspepsia	6 (4.1)	0	4 (5.5)	0

Deaths, other serious adverse events and other significant adverse events (Studienbericht, S. 146 ff.)

Table 12-8 Safety overview (Safety set)

	Ruxolitinib N=146 n (%)	BAT N=73 n (%)
All deaths	6 (4.1)	4 (5.5)
On-treatment deaths	4 (2.7)	3 (4.1)
AEs	145 (99.3)	66 (90.4)
Related AEs with suspected relationship to study treatment	118 (80.8)	14 (19.2)
SAEs regardless of relationship to study treatment	44 (30.1)	21 (28.8)

		<table> <tr> <th></th><th>Ruxolitinib N=146 n (%)</th><th>BAT N=73 n (%)</th></tr> <tr> <td>SAEs with suspected relationship to study treatment</td><td>14 (9.6)</td><td>1 (1.4)</td></tr> <tr> <td>CTC grade 3 or 4 AEs</td><td>63 (43.2)</td><td>18 (24.7)</td></tr> <tr> <td>AEs leading to discontinuation from study drug</td><td>12 (8.2)</td><td>1 (1.4)</td></tr> <tr> <td>AEs requiring dose reduction or interruption of study treatment</td><td>92 (63.0)</td><td>11 (15.1)</td></tr> <tr> <td>AEs requiring concomitant medication</td><td>122 (83.6)</td><td>51 (69.9)</td></tr> <tr> <td>Clinically notable AEs</td><td>105 (71.9)</td><td>16 (21.9)</td></tr> </table>		Ruxolitinib N=146 n (%)	BAT N=73 n (%)	SAEs with suspected relationship to study treatment	14 (9.6)	1 (1.4)	CTC grade 3 or 4 AEs	63 (43.2)	18 (24.7)	AEs leading to discontinuation from study drug	12 (8.2)	1 (1.4)	AEs requiring dose reduction or interruption of study treatment	92 (63.0)	11 (15.1)	AEs requiring concomitant medication	122 (83.6)	51 (69.9)	Clinically notable AEs	105 (71.9)	16 (21.9)
	Ruxolitinib N=146 n (%)	BAT N=73 n (%)																					
SAEs with suspected relationship to study treatment	14 (9.6)	1 (1.4)																					
CTC grade 3 or 4 AEs	63 (43.2)	18 (24.7)																					
AEs leading to discontinuation from study drug	12 (8.2)	1 (1.4)																					
AEs requiring dose reduction or interruption of study treatment	92 (63.0)	11 (15.1)																					
AEs requiring concomitant medication	122 (83.6)	51 (69.9)																					
Clinically notable AEs	105 (71.9)	16 (21.9)																					
		Categories are not mutually exclusive; AEs that occurred outside of Randomized Treatment Phase are not summarized. Source: Table 14.3.1-1.15																					
		<u>Zusammenfassung der Sicherheitsergebnisse (Studienbericht, S. 166)</u> <i>“Despite the anticipated AEs of anemia and thrombocytopenia based upon the known mechanism of action of ruxolitinib as a JAK pathway inhibitor, there were few discontinuations related to cytopenias. Grade 3-4 laboratory findings of anemia and thrombocytopenia were reported in subjects treated with ruxolitinib at rates of 38.4% and 8.2%, respectively, compared to 20.5% and 6.8% on BAT. In the Extension Phase, Grade 3-4 thrombocytopenia was observed in 11% of BAT crossover subjects, which was higher than what was observed in the Randomized Phase. One potential explanation could be that BAT-treated subjects entering the Extension Phase had more advanced disease than those randomized to ruxolitinib at the time of randomization; subjects randomized to BAT could only enter the Extension Phase after some measure of progression. Discontinuation as a result of myelosuppression in the ruxolitinib treatment arm was rare; only one ruxolitinib-treated subject discontinued as a result of thrombocytopenia. In general, these side effects were tolerable, manageable and treated by dose reduction, brief periods of interruption, or blood product transfusions. Dose reductions on account of thrombocytopenia as an AE were required in 41.1% of ruxolitinib-treated subjects. The higher percentage of dose reduction on account of thrombocytopenia is not surprising given that the study protocol mandated dose reductions in such cases. There was no signal of increased deaths in the ruxolitinib arm of the trial. The study was not powered to detect any difference in OS. Causality was not suspected by the Investigator in most of the deaths on study, indicating these were representative events for subjects with MF. With respect to body weight, a severe loss of weight is a known characteristic of MF due to the catabolic nature of the disease; the weight gains observed (sometimes up to 20%) may be due to some subjects having experienced undesired prior weight loss on account of their underlying disease.”</i>																					
22	Fazit der Autoren	-																					
#: Die Verblindung des Auswerters wird zumindest für den primären Endpunkt beschrieben																							

Verzerrungspotenzial auf Studien- Endpunktebene (COMFORT II)

Verzerrungspotenzial auf Studienebene

RCT mit dem zu bewertenden Arzneimittel (Analyse nach 24 bzw. 48 Wochen [Primäranalyse])

Studie	Adäquate Erzeugung der Randomisierungs- sequenz	Verdeckung der Gruppeneinteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver- zerrungspotenzial beeinflussende Punkte	Verzerrungs-potenzial auf Studienebene
			Patient	Behandler			
COMFORT II	Ja	Ja	Nein	Nein	Nein	Nein	Niedrig

Verzerrungspotenzial auf Endpunktebene

Bewertung des Verzerrungspotenzials für den Endpunkt Gesamtüberleben

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver- zerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Ja	Nein	Nein	Niedrig

Bewertung des Verzerrungspotenzials für die Endpunkte Überlebenszeit Leukämiefreies Überleben und Progressionsfreies Überleben

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver- zerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Ja	Nein	Nein	Niedrig

Bewertung des Verzerrungspotenzials für den Endpunkt Rückgang des Milzvolumens um $\geq 35\%$ nach 48 Wochen in MRT oder CT

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Ja ¹	Ja	Nein	Nein	Niedrig
1: Um eine Objektivierung und verblindete Bewertung des üblicherweise herangezogenen klinischen Parameters „tastbare Milzlänge“ zu ermöglichen, wurden MRT-Aufnahmen der Milz gemacht und zentral nur unter der Patientennummer anhand des MRT das Milzvolumen berechnet.						

Bewertung des Verzerrungspotenzials für den Endpunkt Milzvolumenreduktion (prozentuale Veränderung des Milzvolumens gegenüber Studienbeginn nach 24 bzw. 48 Wochen)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Ja ¹	Nein ²	Nein	Nein	Hoch
1: Um eine Objektivierung und verblindete Bewertung des üblicherweise herangezogenen klinischen Parameters „tastbare Milzlänge“ zu ermöglichen, wurden MRT-Aufnahmen der Milz gemacht und zentral nur unter der Patientennummer anhand des MRT das Milzvolumen berechnet. 2: Für diesen Endpunkt wurde eine Bewertung der Observed Cases (ohne Crossover-Patienten) vorgenommen, sodass ein Teil der Patienten – insbesondere in der Placebo-Gruppe – nicht in die Auswertung eingegangen ist.						

Bewertung des Verzerrungspotenzials für den Endpunkt Milzvolumenreduktion (Beginn bzw. Dauer des Ansprechens)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Ja ¹	Nein ¹	Nein	Nein	Hoch
1: Um eine Objektivierung und verblindete Bewertung des üblicherweise herangezogenen klinischen Parameters „tastbare Milzlänge“ zu ermöglichen, wurden MRT-Aufnahmen der Milz gemacht und zentral nur unter der Patientennummer anhand des MRT das Milzvolumen berechnet. 2: In die Auswertung gingen nur die Responder (Milzvolumenreduktion $\geq 35\%$) ein						

Bewertung des Verzerrungspotenzials für den Endpunkt ECOG Performance-Status

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Nein	Nein	Nein	Hoch ¹
1: Da es sich um einen subjektiven Endpunkt bei offenem Studiendesign handelt und die Auswertung der Daten zudem nur die Patienten einschließt, für die ein Wert zu Studienbeginn und zum jeweiligen Auswertungszeitpunkt vorliegt, wird der Endpunkt als hoch verzerrt eingestuft.						

Bewertung des Verzerrungspotenzials für den Endpunkt EORTC QLQ-C30

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Nein	Nein	Nein	Hoch ¹
1: Da es sich um einen subjektiven Endpunkt bei offenem Studiendesign handelt und die Auswertung der Daten zudem nur die Patienten einschließt, für die ein Wert zu Studienbeginn und zum jeweiligen Bewertungszeitpunkt vorliegt, wird der Endpunkt als hoch verzerrt eingestuft.						

Bewertung des Verzerrungspotenzials für den Endpunkt FACT-Lym

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Nein	Nein	Nein	Hoch ¹
1: Da es sich um einen subjektiven Endpunkt bei offenem Studiendesign handelt und die Auswertung der Daten zudem nur die Patienten einschließt, für die ein Wert zu Studienbeginn und zum jeweiligen Auswertungszeitpunkt vorliegt, wird der Endpunkt als hoch verzerrt eingestuft.						

Bewertung des Verzerrungspotenzials für den Endpunkt Unerwünschte Ereignisse

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ¹	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver- zerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
COMFORT II	Niedrig	Nein	Ja	Nein	Nein	Niedrig
1: Die Auswertungen basieren auf der Safety-Population, die alle Patienten einschloss, die mindestens 1 Dosis der Studien- medikation erhalten hatten.						