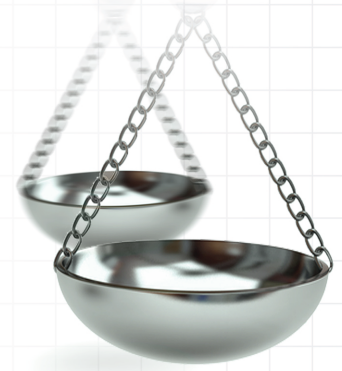




Evaluating the feasibility of a network meta-analysis comparing treatment options in polycythemia vera

Journal of **Comparative Effectiveness Research**

Paul Walden^{‡,1}, Noemi Hummel^{*,2} , Agnieszka Kopiec³, Zuzanna Maliszewska³ & Emi Naslazi⁴

¹PharmaEssentia, Burlington, MA, USA

²Certara GmbH, Lörrach, Germany

³Certara, Krakow, Poland

⁴Certara, Netherlands

*Author for correspondence: noemi.hummel@certara.com

[‡]Affiliation at the time of the study

Aim: Polycythemia vera (PV), a rare, chronic myeloproliferative neoplasm, that negatively impacts patient outcomes, and optimal therapy remains unclear due to a lack of head-to-head trials. A targeted literature review and feasibility assessment for an indirect comparison of ropeginterferon alfa-2b-njft versus peginterferon alfa-2a or ruxolitinib, using standard of care comprising hydroxyurea (HU) as a common comparator was conducted. **Materials & methods:** A targeted literature review evaluated clinical comparative evidence for PV treatments published between January 2014 and May 2024 in PubMed and relevant conference abstracts. End points of interest included complete hematologic response, molecular response, allele burden, event-free survival and safety. The feasibility of a network meta-analysis (NMA) was evaluated based on homogeneity of patient populations, treatment regimens and end point definitions. **Results:** Of 193 PubMed records and 460 conference abstracts screened, 40 records were included, representing evidence from 11 randomized controlled trials and 10 observational studies. Among these, 20 studies formed connected evidence networks for the end points of interest. Substantial heterogeneity across studies precluded a robust NMA: patient populations varied (newly diagnosed, high-risk, low-risk, HU-refractory or -intolerant), complete hematologic response definitions differed (e.g., requirement for absence of disease-related symptoms), molecular response thresholds were inconsistent, follow-up durations varied and definitions of standard of care ranged from almost exclusive use of HU to mixed regimens. **Conclusion:** An NMA for PV treatments was not feasible due to significant clinical and methodological heterogeneity across studies, including differences in patient characteristics, treatments, outcome definitions and follow-up times. These findings highlight the importance of standardized clinical trial designs and outcome definitions to enable robust comparative evidence generation for rare conditions like PV.

Plain language summary

What is this article about? This article explores the possibility of comparing different treatments for polycythemia vera (PV), a rare blood cancer, using a method called network meta-analysis (NMA). NMA allows researchers to compare treatments even when direct clinical trials are missing. The study reviewed existing research to assess if this method can be reliably used, given the differences across PV studies.

What were the results? The review included 40 records that were assessed for feasibility and found significant variability between studies, in how treatments and patient outcomes were defined, which patients were included, what treatments were used as 'best available therapy', and when outcomes were measured. Due to this variability, the requirements for a valid NMA were not met, making it unsuitable for comparing PV treatments based on the current evidence.

What do the results mean? These findings mean that using NMA to compare PV treatments is currently not reliable given the inconsistencies across studies. Alternative methods, like patient level analyses (e.g., matching adjusted indirect comparison or simulated treatment comparison), might be more appropriate but require detailed patient data that are often unavailable. This highlights the challenges of assessing treatments for rare conditions like PV and underscores the importance of more standardized study designs and data to enable better informed treatment decisions.

First draft submitted: 8 September 2025; Accepted for publication: 14 November 2025; Published online: 2 December 2025

Keywords: indirect treatment comparison • network meta-analysis • polycythemia vera • ropeginterferon alfa-2b-njft

Polycythemia vera (PV) is a rare chronic blood cancer within the myeloproliferative neoplasm (MPN) family, defined by an abnormal increase in red blood cell production and sometimes increased production of leukocytes and platelets – driven by clonal stem cell expansion, primarily due to somatic mutations in the *JAK2* gene [1]. Consequently, patients with PV are at high risk for thromboembolic events, cardiovascular complications and increased risk of disease progression to myelofibrosis or acute myeloid leukemia [2–4]. In addition to these clinical sequelae, patients with PV commonly experience fatigue, pruritus, night sweats and cognitive difficulties that persist despite treatment and substantially impair quality of life [5]. Together, these clinical and symptom burdens contribute to a substantial economic impact, with evidence demonstrating increased healthcare resource utilization and costs compared with noncancer populations [6,7].

Management of PV aims to control hematocrit, reduce thrombotic risk and alleviate symptoms through risk-adapted therapy. According to clinical guidelines, low-risk patients (typically younger than 60 years with no prior thrombosis) can be managed with phlebotomy (Phleb), low-dose aspirin and cytoreductive therapy where indicated [8–10], while higher-risk patients are treated with cytoreductive agents such as hydroxyurea (HU), *JAK2* targeted agents or interferon- α -based treatments [8,11,12]. Ropenylated INF α -2b, a newly designed interferon, was approved to treat PV in the USA in October 2021 [13]. Other treatment options include ruxolitinib (RUX), especially for patients intolerant or resistant to HU [14,15], and agents such as pipobroman, anagrelide and immunomodulators [12,16]. Best available therapy (BAT) in clinical trials typically includes a range of treatments such as HU, INF α -s and other agents, or sometimes observation without active pharmacotherapy, reflecting real-world (RW) variability in clinical practice [14].

While randomized controlled trials (RCTs) remain the gold standard for evaluating treatment efficacy, direct head-to-head comparisons of the full range of PV therapies are lacking, complicating clinical decision-making. Most clinical trials in PV to date have compared individual agents against historical standard of care or placebo, often enrolling heterogeneous patient populations and applying variable outcome measures. Consequently, indirect treatment comparisons (ITCs), including network meta-analyses (NMAs), have emerged as valuable tools to estimate relative treatment effects by synthesizing evidence across trials that share common comparators such as BAT [17]. However, the validity of ITCs, especially more complex approaches such as NMA, depends on critical methodological assumptions, including similarity in patient populations (homogeneity), consistency in outcome definitions and timing and comparability of background treatments or standard of care [18,19]. When these assumptions are violated, indirect estimates may be biased or unreliable.

Given the increasing need to understand the comparative effectiveness of PV treatments such as ropegylated INF α -2b (ROPEG), pegylated INF α -2a (PEG), and RUX, there is an urgent need to systematically assess the comparative effectiveness and safety of these treatments using all available evidence. To address this gap, we conducted a targeted literature review (TLR) of clinical trials and RW studies to comprehensively evaluate the efficacy and safety profiles of current PV therapies. Subsequently, we performed a structured feasibility assessment to determine whether a robust NMA comparing ROPEG to PEG and RUX, using BAT as the common comparator, can be conducted given the existing evidence base. This assessment focused on examining heterogeneity in study design, patient populations, treatment regimens and outcome definitions to evaluate the suitability of existing data for reliable indirect comparisons in this rare and clinically heterogeneous disease.

Materials & methods

Overview

A TLR was conducted to identify and assess comparative clinical evidence on the efficacy and safety of systemic therapies for patients with PV. The primary objective was to assess the feasibility of conducting an NMA by evaluating the consistency and comparability of trial characteristics, patient populations, treatment regimens and reported outcomes across available studies.

Literature search

The literature search was conducted on 9 May 2024, covering a 10-year period from 1 January 2014 to 9 May 2024. The primary electronic database searched was PubMed. Search terms combined disease-specific keywords (e.g., “polycythemia vera”, “PV”) with treatment-related terms (e.g., “ropeginterferon alfa-2b-njft”, “ruxolitinib”, “hydroxyurea”) and study design filters for comparative clinical studies and RW evidence. Detailed full search strategy of this TLR is provided in [Supplementary Table 1](#). Additionally, a manual search was conducted of conference proceedings from key hematology and oncology congresses, including the American Society of Clinical Oncology (ASCO), American Society of Hematology (ASH) and European Hematology Association (EHA), and experts were consulted on the completeness of the search. Only studies published in English were considered eligible. Inclusion criteria were pre-specified following a PICOS framework, as detailed in [Supplementary Table 2](#).

Study selection & data extraction

Eligible studies included RCTs and comparative RW studies enrolling adult patients diagnosed with PV. Studies were included if they reported outcomes for at least one of the following treatments: ROPEG, PEG, RUX, phlebotomy (Phleb) or historical standard-of-care therapies such as HU, Inf α , pipobroman, anagrelide, immunomodulators or no active medication. One reviewer screened titles and abstracts for eligibility, followed by full-text review. Discrepancies were resolved by consensus or a third reviewer. Data were extracted using standardized forms and included study design, patient demographics and baseline characteristics, intervention and comparator arms, follow-up duration and key efficacy and safety outcomes. Extracted data were validated by a second reviewer to ensure completeness and accuracy.

Outcomes of interest

Key end points assessed included complete hematologic response (CHR), molecular response, allele burden reduction, event-free survival and safety outcomes such as any adverse event and thromboembolic or thrombotic events. End point definitions, measurement criteria and time points were systematically collected to assess consistency and comparability across studies.

Feasibility assessment for NMA

The feasibility of performing an NMA was evaluated by examining whether key assumptions, such as homogeneity of trial populations and consistency in end point definitions were met. Clinical and methodological characteristics were reviewed across the included studies, focusing on trial design features (e.g., randomization and blinding), treatment regimens and dosing, eligibility criteria, baseline patient characteristics and reported end points. The assessment determined whether a connected treatment network could be constructed based on shared comparators, particularly BAT, and whether clinical end points were reported in a sufficiently consistent and comparable manner to support quantitative synthesis. Based on this evaluation, we assessed whether key methodological assumptions for conducting an NMA could be reasonably met and considered whether alternative indirect comparison methods might be appropriate.

Results

Targeted literature review

A total of 653 records were initially identified through database and conference abstract screening, including 193 PubMed records and 460 conference abstracts. Following title and abstract screening and full-text review, 36 records from database and 30 records from conference abstract search were included. In addition to the database and conference abstract searches, four relevant comparative studies meeting the inclusion criteria were identified through expert recommendation. These were reviewed and included in alignment with the predefined PICOS framework. After de-duplication, 40 records met inclusion criteria for qualitative synthesis, covering 11 RCTs and 10 observational comparative RW studies. The study selection process is summarized in the PRISMA flowchart shown in [Figure 1](#). These included studies reported on various systemic treatment regimens for PV, providing the clinical evidence base for subsequent feasibility assessment.

Feasibility assessment

Among the 21 studies deemed suitable for feasibility evaluation based on availability of comparative data and relevant outcomes, 20 formed a connected treatment network, as illustrated in [Figure 2](#). One study (Podoltsev,

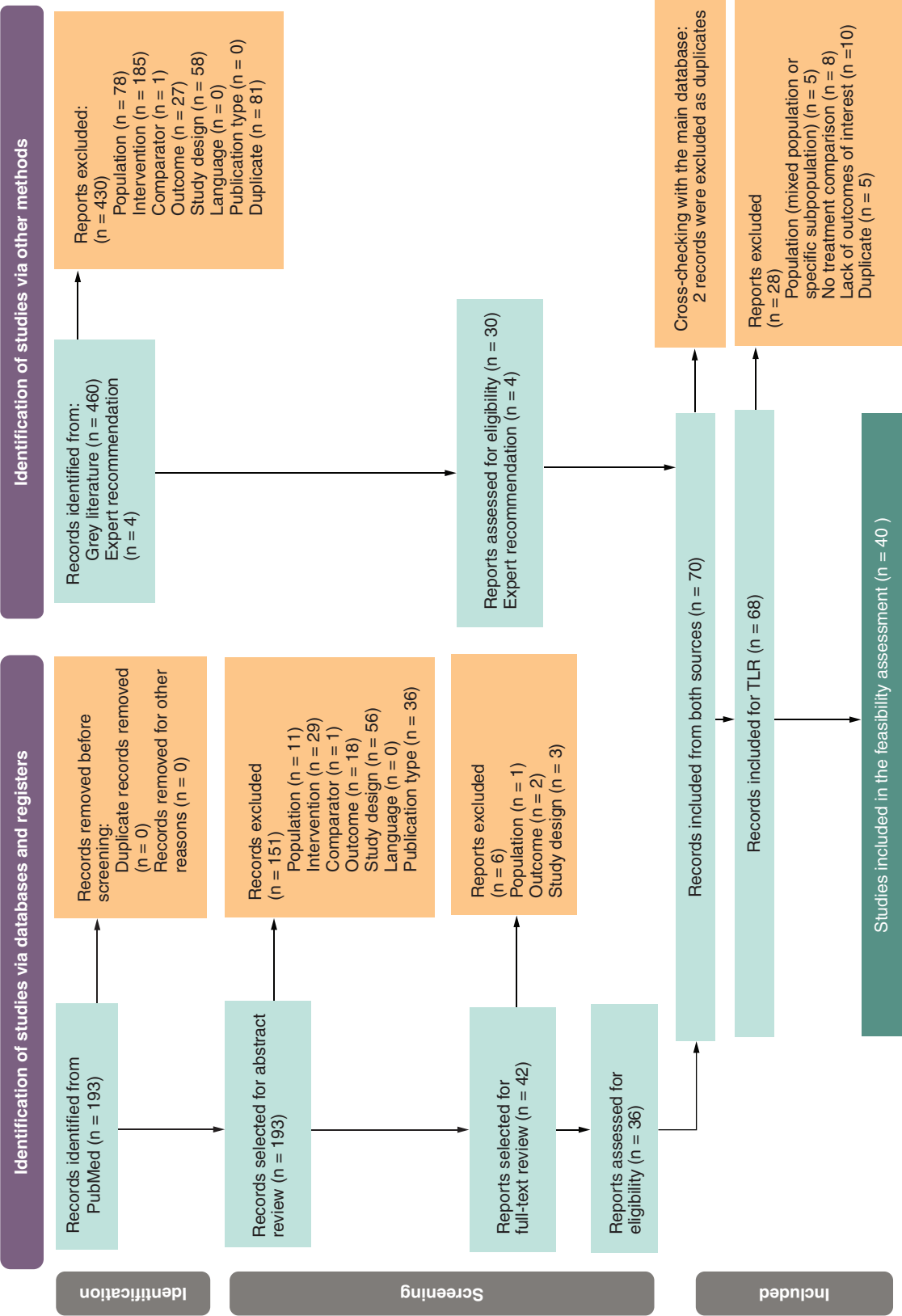


Figure 1. PRISMA flowchart.

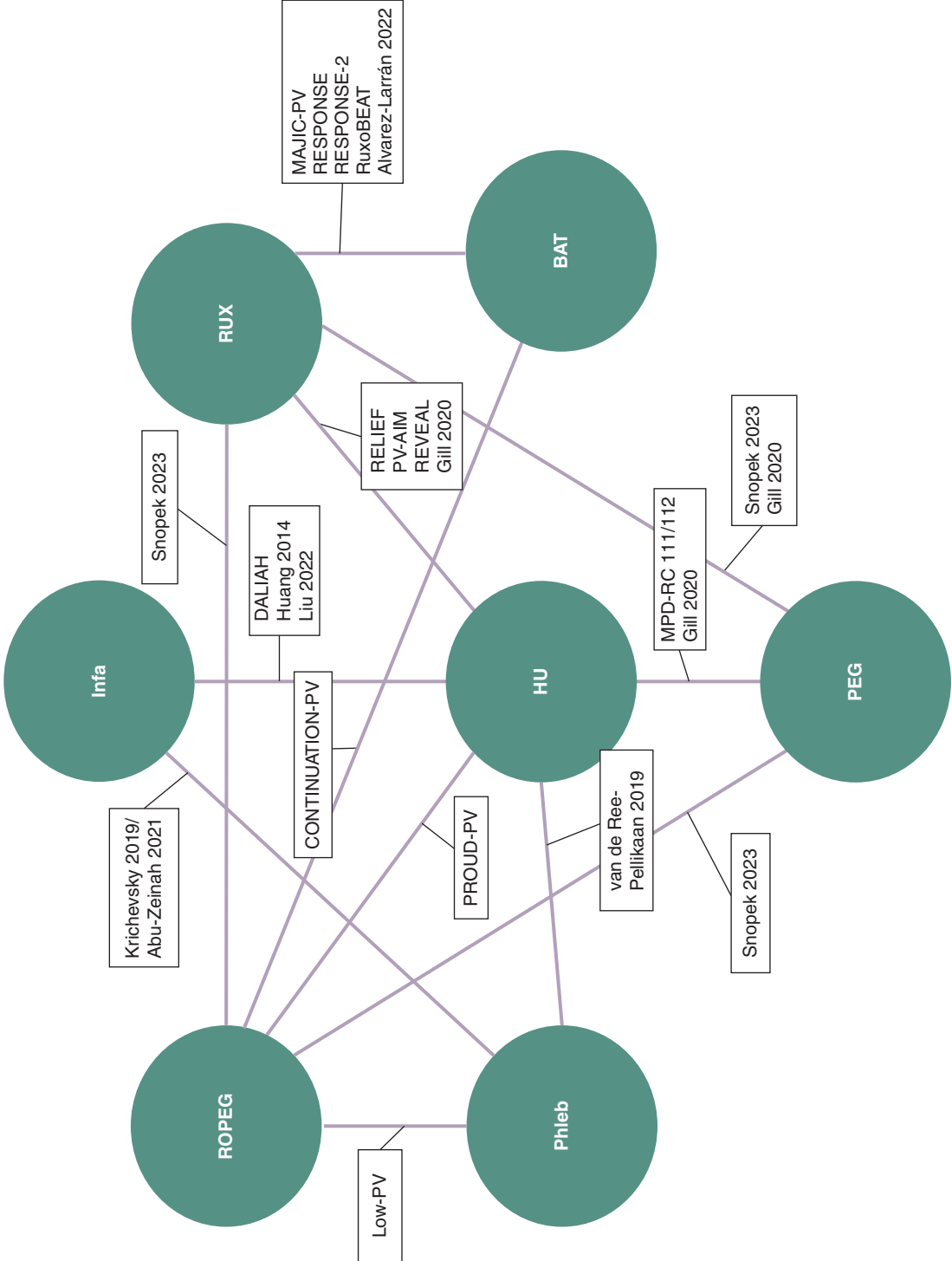


Figure 2. Best case evidence network for comparison of polycythemia vera treatments. BAT: Best available therapy; HU: Hydroxyurea; INF α : Interferon- α ; PEG: Pegylated INF α ; Phleb: Phlebotomy; PV: Polycythemia vera; ROPEG: Ropegylated INF α ; RUX: Ruxolitinib.

Table 1. Availability of outcomes across included studies.

Study	Comparison	Follow-up (months)	CHR [†]	Molecular response ^{††}	Allele burden ^{§§§}	Event-free survival
PROUD-/CONTINUATION-PV	ROPEG vs HU	Up to 72	ü [‡]	ü ^{‡‡}	ü	ü ^{###}
Low-PV	ROPEG vs Phleb	Up to 24		ü ^{§§}	ü	
MPD-111/112	PEG vs HU	Up to 36	ü		ü ^{¶¶¶¶}	ü ^{††††}
DALIAH	INFα vs HU	36	ü	ü ^{¶¶}	ü	
Huang 2014	INFα vs HU	Up to 60	ü [§]	ü ^{##}		ü ^{††††}
Liu 2022	INFα vs HU	Up to 60	ü [¶]	ü ^{§§}	ü	ü ^{§§§§}
Krichevsky 2019, Abu-Zeinah 2021	INFα vs HU vs Phleb	Up to 336				ü ^{¶¶¶¶}
van de Ree-Pellikaan 2019	Phleb vs HU vs Phleb + HU	Up to 12	ü [¶]			
Snopek 2023	ROPEG vs PEG vs RUX	15				
RELIEF	RUX vs HU	4				
REVEAL	RUX vs HU	Up to 12				
Gill 2020	RUX vs PEG vs HU	6	ü			
RESPONSE/RESPONSE-2	RUX vs BAT	Up to 8	ü [¶]	ü ^{†††}	ü	
MAJIC-PV	RUX vs BAT	12	ü [#]	ü ^{‡‡‡}	ü	ü ^{¶¶¶¶}
PV-AIM	RUX vs HU	Up to 156				
RuxoBEAT	RUX vs BAT	6	ü			
Alvarez-Larrán 2022	RUX vs BAT	96				ü ^{####}

[†] For CHR, each of the following must hold: HC <45% w/o phlebotomy, platelet count $400 \times 10^9/l$, WBC count $\leq 10 \times 10^9/l$, normal spleen size on imaging, no disease-related symptoms (microvascular disturbances, pruritus, headache).
[‡] CHR without the spleen-size and/or symptoms requirements are also available.
[§] CHR definition without phlebotomy requirement.
[¶] CHR definition without phlebotomy, spleen-size and symptoms requirements.
[#] CHR definition without symptoms requirements.
^{††} Molecular response is defined as complete response (i.e., reduction of any molecular abnormality to undetectable levels) or partial response, applying only to patients with a baseline value of mutant allele burden $\geq 10\%$ (i.e., reduction of $\geq 50\%$ from baseline value in patients with <50% mutant allele burden at baseline OR reduction of $\geq 25\%$ from baseline value in patients with >50% mutant allele burden at baseline).
^{‡‡} Complete or partial response.
^{§§} Partial response only, applying only to patients with a baseline value of mutant allele burden $\geq 20\%$.
^{¶¶} Partial response only.
^{##} Complete or partial response without restriction on baseline allele burden.
^{†††} Complete response only, and partial response only, applying only to patients with a baseline value of mutant allele burden $\geq 20\%$.
^{‡‡‡} Partial response only without restriction on baseline allele burden.
^{§§§} Allele burden is provided as change from baseline.
^{¶¶¶} Provided as medians over time.
^{###} Events included thromboembolic event, myelofibrosis, acute leukemia, death or disease progression, death and thromboembolic events.
^{††††} Events included major thrombotic event, major hemorrhagic complications, myelofibrosis, acute leukemia or death.
^{‡‡‡‡} Events included thrombosis, bleeding, spleen enlargement, severe myelofibrosis or death.
^{§§§§} Myelofibrosis-free survival and thrombosis-free survival were reported.
^{¶¶¶¶} Myelofibrosis-free survival was reported.
^{####} Events included major thrombosis, major hemorrhage, transformation or death.
 BAT: Best available therapy; CHR: Complete hematologic response; HU: Hydroxyurea; INFα: Interferon-α; PEG: Pegylated INFα; Phleb: phlebotomy; PV: Polycythemia vera; ROPEG: Ropegylated INFα; RUX: Ruxolitinib.

2018) [20] was excluded from the network due to the absence of a shared comparator; it compared HU + Phleb versus no treatment, preventing linkage to other studies. The final list of 39 records of 20 studies forming a connected network is provided in [Supplementary Table 3](#).

Availability and definitions of key clinical end points, such as CHR, molecular response, allele burden and event-free survival across studies varied substantially across studies, as summarized in [Table 1](#). Definitions of CHR frequently omitted spleen size and symptom criteria, and achievement of hematocrit <45% without Phleb was inconsistently required across trials. Molecular response definitions varied with respect to completeness and allele burden thresholds; some studies reported only partial molecular response, while others reported both partial and complete molecular responses, with or without baseline allele burden prerequisites. Assessment timepoints and duration of follow-up also showed wide variation. Definitions of BAT differed widely, ranging from almost exclusive use of HU (97% HU) to heterogeneous mixed regimens with 25–60% of patients treated with HU ([Table 2](#)).

Baseline patient characteristics across the connected network revealed considerable heterogeneity in terms of median age, risk category, spleen size and baseline allele burden ([Table 3](#)). Included patient populations ranged

Table 2. Summary of best available therapy regimens across studies.

Study	BAT sample size (n)	Receiving HU, %	Other BAT components
RESPONSE	112	58.9%	Interferon, anagrelide, immunomodulators, pipobroman and no medication
RESPONSE-2	75	50.7%	Interferon or pegylated interferon, pipobroman, lenalidomide or no treatment
CONTINUATION-PV	66	97%	Hydroxyurea or conventional Interferon
MAJIC-PV	87	32%	The most frequent BATs: Interferon, combination of hydroxycarbomide and interferon, combination of anagrelide and hydroxycarbomide
RuxoBEAT	28	25%	Anagrelide, Interferon alpha or others
Alvarez-Larrán 2022	272	60%	Interferon, anagrelide, busulfan, melphalan, radioactive phosphorus, other treatments and no medication

BAT: Best available therapy; HU: Hydroxyurea.

Table 3. Patient baseline characteristics of studies forming networks.

Study	Study population	Arm	Age, years, median (IQR)	Sex, M (%)	Spleen size median, cm (IQR)	Allele burden at baseline, mean, % (SD)
PROUD-PV	Adults with PV	ROPEG	60 (52, 66)	46	13.1 (11, 15)	41.9 (24)
		HU	60 (48, 67)	47	13 (11.5, 15.2)	42.8 (24)
CONTINUATION-PV	Adults with PV	ROPEG	58 (50, 64)	49	13.5 (11.5, 15)	42.8 (23)
		HU/BAT	59 (49, 65.5)	47	12.8 (11.3, 15.5)	42.9 (23)
Low-PV	Low-risk patients with PV	ROPEG	51.7 (45.5, 55.3)	73.4	2.0 (2.0, 3.0) ^{††}	34 (18, 57) ^{#,§§}
		Phleb	48.2 (43.7, 57.4)	61.9	2.5 (2.0, 5.0) ^{††}	27 (19, 66) ^{#,§§}
MAJIC-PV	Patients resistant/intolerant to HU	RUX	67 (34, 88) [§]	60	14 (9, 26) [§]	64 [¶]
		HU/BAT	66 (28, 85) [§]	56	14 (9, 30) [§]	58 [¶]
Liu 2022	PV patients	IFN α -2b	51 (44, 57)	39	NR	56 (35, 73) [#]
		HU	61 (52, 67)	49	NR	59 (33, 73) [#]
MPD-RC 112 [†]	High-risk ET/PV patients	PEG	60 (19, 79) [§]	60	12.5 (6.5, 22)	34.5 (22.0) [‡]
		HU	63 (18, 87) [§]	56	12.5 (2.1, 20)	36.0 (18.3) [‡]
DALIAH [†]	Newly diagnosed or untreated MPN patients	IFN α	59 (20, 88) [§]	54	NR	33 (19, 51)
		HU	68 (60, 80) [§]	63	NR	37 (17, 52)
RESPONSE	Patients with PV, phlebotomy-dependent patients with splenomegaly	RUX	62 (34, 90)	60	7 (0, 24) [§]	76.2 (17.8)
		BAT	60 (33, 84)	71.4	7 (0, 25) [§]	75 (22.6)
RESPONSE-2	Adults with PV, no palpable splenomegaly, and HU resistance or intolerance	RUX	63 (54, 61)	53	NR	53 (9, 95) [§]
		BAT	67 (61, 74)	63	NR	74 (13, 95) [§]
van de Ree-Pellikaan 2019	Patients with low- and high-profile risk of PV	Phleb	58.7 (13.1) ^{††}	37	NR	NR
		HU	69.1 (9.2) ^{††}	12	NR	NR

Allele burden at baseline refers to JAK2V617F.

[†] Baseline characteristics reported involve both PV and ET, as presented in the original publication, if not stated differently.[‡] For PV patients.[§] Median (range).[¶] Median.[#] Median (IQR).^{††} measured below the costal margin.^{††} Mean (SD).^{§§} Patients responding at Month 12.ET: Essential thrombocythemia; HU: Hydroxyurea; IFN α : Interferon- α ; IQR: Interquartile range; M: Male; MPN: Myeloproliferative neoplasm; PEG: Pegylated IFN α ; Phleb: Phlebotomy; PMF: Primary myelofibrosis; PreMF: Prefibrotic myelofibrosis; PV: Polycythemia vera; ROPEG: Ropegylated IFN α ; RUX: Ruxolitinib; SD: Standard deviation.

from low-risk, treatment-naïve individuals to those with HU-resistant or -intolerant and advanced PV cases. These variations indicate the presence of potential treatment effect modifiers that violate the homogeneity requirement of NMA.

Despite the ability to construct a connected treatment network, the heterogeneity in study populations, end point definitions and treatment regimens prevented fulfillment of key methodological assumptions required for NMA. Consequently, a quantitative synthesis via NMA was considered infeasible in PV.

Discussion

Summary of findings

This feasibility assessment demonstrated that conducting an NMA for systemic therapies in PV is not feasible due to substantial heterogeneity in the available evidence base. Outcome definitions were inconsistent, with CHR varying in inclusion of spleen size, symptom control, and hematocrit ($<45\%$) without phlebotomy. Molecular response criteria also lacked standardization, making cross-trial comparisons challenging. Patient populations ranged widely, from low-risk, treatment-naïve individuals to patients refractory or intolerant to HU, while definitions of BAT varied considerably by the proportion of patients using HU (25–97%) and the mix of treatments including, e.g., IFN α and no active therapy. In addition, assessment timepoints and follow-up durations diverged across studies.

Methodological implications

Collectively, these inconsistencies violated the key methodological assumptions of NMA, making a robust quantitative synthesis inappropriate. Specifically, heterogeneity in end point definitions directly undermines the assumption of homogeneity required for valid indirect comparisons. For example, trials defining CHR solely as achievement of hematocrit $<45\%$ differ fundamentally from those also requiring normalization of spleen size and control of disease-related symptoms. Linking such studies would introduce bias, as the less stringent definition may overestimate the treatment effect relative to stricter criteria, yielding inflated or misleading comparative estimates. Likewise, variability in BAT composition compromises the transitivity assumption by altering the nature of the common comparator across studies. BAT ranged from almost exclusive use of HU to diverse regimens combining HU, IFN α and other agents, reflecting temporal or geographic shifts in clinical practice. These discrepancies mean that the “shared” BAT comparator does not represent a consistent treatment context, invalidating the indirect connections between studies and rendering the resulting network unreliable.

Importantly, the inconsistencies observed in CHR definitions arise even with established European LeukemiaNet (ELN) and International Working Group for Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) consensus criteria in place [21], which provide detailed guidance on hematologic, molecular and clinical end points. This indicates that the challenge is not the absence of standardized definitions, but rather a lack of adherence to these guidelines in trial design and reporting. Addressing this gap through collaboration among trial sponsors, regulatory agencies and research groups is critical to improve the consistency and comparability of PV clinical trial data.

Implication of findings for stakeholders

The lack of comparability across PV trials has practical implications for researchers, clinicians, payers and policymakers. In the absence of a robust NMA, treatment decisions must rely on qualitative synthesis and clinical judgment. For a rare disease like PV, where head-to-head trials are scarce, there is an urgent need for standardized data collection and consistent end point definitions, especially for key measures such as hematocrit control, symptom burden, splenomegaly and JAK2 V617F allele burden, to drive evidence-informed decision-making [22]. Standardization is supported by consensus guidance, such as the ELN recommendations, aimed at harmonizing response criteria in PV and essential thrombocythemia to improve comparability across clinical trials [21].

Comparison with previous research

Our findings mirror broader challenges in the hematologic oncology literature, particularly with rare diseases and MPNs, where inconsistent trial designs and varied patient cohorts impede the validity of NMAs [23]. ITCs can still offer valuable insights when direct evidence is unavailable, but their reliability is critically dependent upon consistency across the underlying studies. In this context, population-adjusted indirect comparison methods, such as matching-adjusted indirect comparison (MAIC) or simulated treatment comparison (STC), have emerged as alternatives to address heterogeneity, although their application is contingent upon access to patient-level data, which is an ongoing challenge in PV research [24]. A recent meta-analysis provides further evidence of variability in treatment outcomes and safety profiles across PV populations, specifically examining cytoreductive therapy in younger adults [25]. Their findings reinforce the need for more age-stratified data and standardized reporting practices to support reliable comparative analyses and inform treatment decisions in subgroups of interest.

Strengths & limitations

The strength of this work lies in its structured approach, which systematically assessed the feasibility of an NMA by critically evaluating available evidence across both RCTs and comparative RW studies. This approach adds transparency and methodological rigor, allowing for clear understanding of the data limitations. However, the assessment was constrained by the variable quality and reporting practices of the included studies, which often lacked standardized definitions, patient characteristics or follow-up measures. The exclusion of single-arm and noncomparative studies further narrowed the evidence base, although their inclusion would not have resolved the core methodological limitations identified. Crucially, access to patient-level data would allow for more complex indirect comparisons like MAIC or STC to overcome heterogeneity issues.

Implications for future research

Future PV studies should adopt standardized definitions for key clinical outcomes such as CHR, molecular response and other key end points and align patient inclusion criteria, such as risk stratification. Researchers and trial sponsors should also consider aligning follow-up times and clearly specifying BAT regimens to enable more robust evidence synthesis across trials. Collaboration between trial groups, regulatory agencies and patient organizations is essential to develop core outcome sets and consensus definitions, improving evidence synthesis quality.

Conclusion

In conclusion, this feasibility assessment highlights significant limitations within the current PV evidence base for conducting a robust NMA. The substantial clinical and methodological heterogeneity across trials precludes quantitative synthesis and necessitates alternative approaches such as MAIC or STC when patient-level data are available. More broadly, these findings underscore an urgent need for standardization in PV clinical trials, from patient characteristics and outcome definitions to comparator treatments and assessment schedules, to enable valid indirect comparisons in the future. Notably, these inconsistencies occur despite the existence of the ELN and IWG-MRT consensus criteria, highlighting the importance of adherence to established guidelines and collaborative standardization. As the PV treatment landscape continues to evolve with novel interferon formulations, JAK inhibitors and emerging disease-modifying agents, treatment goals are shifting beyond hematocrit control toward broader disease modification and long-term remission. This dynamic environment further reinforces the need for standardized outcome measures and rigorous study designs to ensure comparability across therapies and to generate reliable evidence for clinical and policy decision-making. In rare and complex conditions such as PV, addressing these evidence gaps will be critical for guiding clinical practice, supporting health technology assessments and ultimately optimizing patient outcomes.

Summary points

- Network meta-analysis for systemic treatments in polycythemia vera (PV) was not feasible due to substantial heterogeneity across studies, violating key assumptions.
- Definitions of complete hematologic response varied significantly, including inconsistencies in spleen size, symptom control and hematocrit thresholds (with or without phlebotomy).
- Patient populations ranged widely, from low-risk, treatment-naïve patients to those intolerant or refractory to hydroxyurea, as well as high-risk PV patients.
- Definitions of BAT were inconsistent, spanning almost exclusive use of hydroxyurea to mixed regimens or no active treatment.
- Assessment of time points and follow-up durations varied considerably across studies, complicating cross-trial comparisons.
- These inconsistencies collectively precluded valid quantitative synthesis via network meta-analysis.
- Population-adjusted indirect comparisons, such as matching-adjusted indirect comparison and simulated treatment comparison, may be more appropriate but require access to patient-level data.
- These findings highlight key challenges in evidence synthesis for rare diseases like PV where head-to-head trials are limited, underscoring the need for standardized data and outcome definitions to improve future comparative analyses.
- The rapidly evolving PV treatment landscape, with new interferon formulations, JAK inhibitors and emerging therapeutic strategies, highlights the urgency of adopting standardized end points and rigorous trial designs to ensure outcomes remain comparable across therapies.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://becarispublishing.com/doi/epdf/10.57264/cer-2025-0142>

Author contributions

N Hummel, A. Kopiec, Z Maliszewska and E Naslazi were responsible for acquisition of data and data analysis; P Walden was responsible for study conception and design; all authors were responsible for drafting and revision of the manuscript.

Acknowledgments

The authors thank A Howe, C Castro and H-L Chien for their comments and feedback on earlier versions of this manuscript.

Financial disclosure

This study was funded by PharmaEssentia (MA, USA).

Competing interests disclosure

P Walden was an employee of PharmaEssentia at the time of the study. N Hummel, A Kopiec and E Naslazi are employees of Certara which is a paid consultant to PharmaEssentia. The authors have no other competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing disclosure

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

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

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

- Barbui T, Thiele J, Gisslinger H *et al.* The 2016 who classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J.* 8(2), 15 (2018).
- **Lays the diagnostic foundation (WHO 2016) used by many included trials, important background that impacts trial inclusion criteria and baseline population consistency.**
- Spivak JL. Polycythemia vera: myths, mechanisms, and management. *Blood* 100(13), 4272–4290 (2002).
- Tefferi A, Barbui T. Polycythemia vera and essential thrombocythemia: 2021 update on diagnosis, risk-stratification and management. *Am. J. Hematol.* 95(12), 1599–1613 (2020).
- Silver RT, Abu-Zeinah G. Polycythemia vera: aspects of its current diagnosis and initial treatment. *Expert Rev. Hematol.* 16(4), 253–266 (2023).
- Mesa RA, Niblack J, Wadleigh M *et al.* The burden of fatigue and quality of life in myeloproliferative disorders (mpds): an international internet-based survey of 1179 mpd patients. *Cancer* 109(1), 68–76 (2007).
- Yu J, Gayle J, Rosenthal N *et al.* Resource utilization and inpatient hospitalization costs associated with thromboembolic events among patients with polycythemia vera. *Oncologist* 30(2), oyaf001 (2025).
- **This study quantifies inpatient healthcare resource use and costs associated with thromboembolic events in patients with polycythemia vera (PV) using recent real-world data. Adds real-world burden context; helpful for explaining why consistency in thrombotic endpoints is critical but often lacking.**
- Dores GM, Curtis RE, Linet MS, Morton LM. Cause-specific mortality following polycythemia vera, essential thrombocythemia, and primary myelofibrosis in the US population, 2001–2017. *Am. J. Hematol.* 96(12), E451–E454 (2021).
- Barbui T, Barosi G, Birgegard G *et al.* Philadelphia-negative classical myeloproliferative neoplasms: critical concepts and management recommendations from European Leukemianet. *J. Clin. Oncol.* 29(6), 761–770 (2011).
- Boulnois L, Robles M, Maaziz N *et al.* Benefit of phlebotomy and low-dose aspirin in the prevention of vascular events in patients with EPOR primary familial polycythemia on the island of New Caledonia. *Haematologica* 109(8), 2688–2692 (2024).
- NCCN. NCCN Clinical Practice Guidelines in Oncology: myeloproliferative neoplasms. Accessed: 14 October 2025 Version 2.2025. Available from: <https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1477>
- Gisslinger H, Klade C, Georgiev P *et al.* Ropgeinterferon alfa-2b versus standard therapy for polycythaemia vera (PROUD-PV and CONTINUATION-PV): a randomised, non-inferiority, phase III trial and its extension study. *Lancet Haematol.* 7(3), e196–e208 (2020).

- **Important phase III evidence directly comparing ropeginterferon alfa-2b-njft to hydroxyurea, reporting 3-year and longer-term hematologic and molecular response data; core to assessing endpoint variability and clinical context.**
- 12. Tefferi A, Vannucchi AM, Barbui T. Polycythemia vera treatment algorithm 2018. *Blood Cancer J.* 8(1), 3 (2018).
- 13. FDA. Besremi (ropeginterferon alfa-2b-njft) injection, for subcutaneous use. (Accessed: 14 October 2025). Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/761166s007lbl.pdf
- 14. Harrison C, Kiladjian JJ, Al-Ali HK *et al.* Jak inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N. Engl. J. Med.* 366(9), 787–798 (2012).
- 15. Verstovsek S, Mesa RA, Gotlib J *et al.* A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. *N. Engl. J. Med.* 366(9), 799–807 (2012).
- 16. Bewersdorf JP, Giri S, Wang R *et al.* Interferon alpha therapy in essential thrombocythemia and polycythemia vera – a systematic review and meta-analysis. *Leukemia* 35(6), 1643–1660 (2021).
- 17. Kim H, Gurrin L, Ademi Z, Liew D. Overview of methods for comparing the efficacies of drugs in the absence of head-to-head clinical trial data. *Br. J. Clin. Pharmacol.* 77(1), 116–121 (2014).
- 18. Efthimiou O, Mavridis D, Debray TP *et al.* Combining randomized and non-randomized evidence in network meta-analysis. *Stat. Med.* 36(8), 1210–1226 (2017).
- 19. Phillippo DM, Ades AE, Dias S *et al.* Methods for population-adjusted indirect comparisons in health technology appraisal. *Med. Decis. Making* 38(2), 200–211 (2018).
- 20. Podoltsev NA, Zhu M, Zeidan AM *et al.* The impact of phlebotomy and hydroxyurea on survival and risk of thrombosis among older patients with polycythemia vera. *Blood Adv.* 2(20), 2681–2690 (2018).
- 21. Barosi G, Mesa R, Finazzi G *et al.* Revised response criteria for polycythemia vera and essential thrombocythemia: an ELN and iwg-mrt consensus project. *Blood* 121(23), 4778–4781 (2013).
- **This consensus document from ELN and IWG-MRT updates the 2009 criteria for hematologic, molecular and histologic responses in PV and essential thrombocythemia, incorporating spleen size, symptom assessment, vascular events and bone marrow histology. It aims to standardize endpoints across clinical trials to improve comparability, supporting our call for harmonization.**
- 22. Gotlib J. Treatment and clinical endpoints in polycythemia vera: seeking the best obtainable version of the truth. *Blood* 139(19), 2871–2881 (2022).
- **This expert perspective reviews PV treatment goals, directly aligned with our conclusion that trials focus on short-term outcomes, underpowering long-term insight.**
- 23. Titmarsh GJ, Duncombe AS, McMullin MF *et al.* How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am. J. Hematol.* 89(6), 581–587 (2014).
- 24. Macabeo B, Quenéchdu A, Aballéa S *et al.* Methods for indirect treatment comparison: results from a systematic literature review. *J. Mark. Access Health Policy* 12(2), 58–80 (2024).
- 25. Chamseddine RS, Savenkov O, Rana S *et al.* Cytoreductive therapy in younger adults with polycythemia vera: a meta-analysis of safety and outcomes. *Blood Adv.* 8(10), 2520–2526 (2024).
- **This meta-analysis is one of the most relevant and recent syntheses of evidence on cytoreductive therapy in PV, focusing specifically on younger adult patients, a key subpopulation within the broader PV landscape. It highlights ongoing challenges related to treatment heterogeneity, differences in outcome reporting, and safety profiles across studies. These findings directly support the need for standardization in future PV trials and align with the conclusions of this feasibility assessment regarding the limitations of conducting robust indirect comparisons across diverse PV populations.**