Supplementary Appendix:

A) Study Design
At the time of planning this study, only few prospective data from clinical trials were available on the incidence of clinical events following cytoreductive treatment with interferon alpha, hydroxyurea or platelet reduction with anagrelide for treatment of PSVG-ET.

- In a prospective trial long-term treatment with interferon alpha abolished or reduced symptoms associated with thrombocytosis in patients with myeloproliferative disorders\textsuperscript{50}. Total symptoms per patient/year were 6.00 before therapy and 1.03 after therapy. Symptoms included bleedings, venous thrombosis, pulmonary embolism and arterial microcirculatory disturbances.

- A prospective randomized trial had demonstrated a significant reduction of platelet counts and a reduction of major thrombotic events (including severe transitory ischemic attacks) for hydroxyurea to 3.6\% over a median period of 27 months compared to 24\% in the comparator arm with no therapy\textsuperscript{6}.

- An open prospective explorative GCP phase II study with anagrelide demonstrated efficacy of platelet reduction and showed an event rate of 2.1\% for major (thrombohemorrhagic) events, and 15.3\% for minor (transitory ischemic events and microcirculatory disturbances) events, respectively, over 6 months\textsuperscript{21}. In addition to diagnostic PSVG parameters histopathology was used for diagnosis of ET in this trial.

Neither of the studies used WHO criteria for diagnoses, so the incidence of clinical events in patients diagnosed according to WHO was unknown at the time of designing the study. Considering an estimated incidence of about 4\%/year for major thrombohemorrhagic events during cytoreductive therapy with hydroxyurea or platelet reductive therapy with anagrelide in patients with WHO-ET, which is a rare disease, a sample size of about 1600 patients (recruitment time of 5 years and a drop out rate of 10\% per year) would be required to demonstrate a $d = 0.5$ in a superiority trial, i.e. a reduction from 4 to 2\% for the test drug\textsuperscript{6,21}. 
• This approach was considered unfeasible in a GCP compliant setting for regulatory purposes in patients diagnosed for WHO-ET for the following reasons: ET is a rare disease, and diagnosis according to WHO-ET further limits patient numbers to 30 to 50% of patients compared to diagnosis of ET according to PSVG criteria\(^5\).

• From the feasibility analysis it was concluded that the duration of a superiority trial including 30 centers would have required >10 years of recruitment taking a recruitment of 3 to 5 previously untreated patients per year per center into account.

As this study was designed as a phase III study for regulatory approval and as the comparator was an active treatment a test for non-inferiority was chosen. This test includes a confirmatory test for superiority as soon as non-inferiority can be demonstrated\(^5\). The non inferiority design was justified by the following arguments: based on the hypothesis that anagrelide and hydroxyurea would not differ substantially as far as their clinical event rate is concerned we thought it would be scientifically more meaningful to prove non-inferiority by statistical significance than to disprove superiority or equivalency due to a lack of statistical significance. From a statistical point of view it is inappropriate to conclude equivalency or non inferiority based on observing a non-significant result of the null hypothesis when there is no difference between the investigational drug and the active comparator\(^5\).

In order to raise the benchmark for non-inferiority in respect to clinical events a sequence of 3 primary criteria had to be passed before non-inferiority for ET events as the 4\(^{th}\) primary criterion could be tested based on the following rational:

a) Platelet reduction is the primary pharmacodynamic activity of anagrelide, thus it was hypothesized that anagrelide had to be non-inferior to hydroxyurea in terms of platelet reduction. It was assumed, in case anagrelide was inferior to hydroxyurea in terms of platelet reduction, that it would be less likely that anagrelide had the same efficacy in reducing the clinical events.

b) Leukocytes seem to contribute to the development of arterial events in patients with ET diagnosed according to PSVG. If leukocytes decreased under hydroxyurea, but remained stable under anagrelide, one would expect fewer arterial events under hydroxyurea.
c) Anagrelide has been previously associated with anemia (24% of patients experienced a hemoglobin decrease of >3 g/dl) in ET patients diagnosed according to PSVG after a mean treatment period of 10.8 years\textsuperscript{54}. If anagrelide lowered hemoglobin levels significantly more than hydroxyurea following long-term observation, the benefit risk ratio would change dramatically against anagrelide.

d) Thus the most important non-inferiority criterion clinical events could be tested only after having tested the hypotheses for criteria a) to c).

B) Definition of ET related events

**Major events**

**Arterial thrombosis**

Cerebrovascular events or stroke, myocardial infarction, peripheral arterial disease, other arterial thrombosis (e.g. obstruction of the mesenterial artery), other major arterial events

**Venous thrombosis**

Ileofemoral thrombosis, pulmonary (infarction) embolism, splanchnic vein thrombosis, portal thrombosis, other major venous events

Major events were diagnosed according to local criteria at the investigators’ institution using detailed description of events including parameters measured, e.g. clinical signs and symptoms, plasma enzyme levels, ECG, CT, MRI, angiography, duplex sonography or other objective means

**Minor events**

**Arterial events**

TIA, angina pectoris, unstable angina, generalized convulsions, erythromelalgia, ocular symptoms, other peripheral arterial microcirculatory disturbances (e.g. ulcers), other minor arterial events (e.g. tinnitus, vertigo)

**Venous events**

Superficial thrombophlebitis, other minor venous events (e.g. hemorrhoids)
TIAs were defined as neurological symptoms of abrupt onset with resolved within 24 hours. Erythromelalgia was defined as pain in the extremities and typical clinical symptoms such as redness. Other minor events were diagnosed based on patients’ symptomatology and clinical judgment of the investigator taking patient diary notes into account.

**Bleeding events**

**Major bleeding events**

Hb drop $\geq 1$ g/dl or red blood cell transfusion required

**Minor bleeding events**

No red blood cell transfusion required and Hb drop $< 1$g/dl

**C) Adaptive Design**

The study was planned as a two-stage adaptive design according to Bauer and Köhne\(^3\). This means that a study stage I with a reduced number of patients is performed. The resulting p-value is interpreted in the following way

- **Stage I:** If $p_1 \geq a_0$ the trial stops with the acceptance of $H_0$.
  - If $p_1 \leq a_1$ the trial stops with the rejection of $H_0$.
  - If $a_1 < p_1 < a_0$ the second stage of the procedure is performed.

- **Stage II:** $H_0$ is rejected if $p_1 \times p_2 < c_{\alpha}$

The global multiple level alpha of the study was defined as $\alpha = 0.025$, one-sided, as required by the ICH Guidance E 9; then the value $\alpha_0$, the boundary for stopping because of futility, was stipulated as $\alpha_0 = 0.5^{53}$. The critical value for decision in stage I is $\alpha_1 = 0.0102$ (instead of 0.025 for a fixed sample size study), and the critical value for testing of the product of the p-values of the two stages is $c_{\alpha} = 0.00380$. Regarding the problem of multiple testing it was decided that only the first of the four primary criteria would be of highest relevance for the sample size estimation, i.e. the adaptive two-stage procedure decision would only be applied on this first criterion. With $\alpha = 0.025$, one-sided, and $\beta = 0.1$ (power = 0.9) and stipulating a non-inferiority margin of the Mann-Whitney measure $= 0.36$ (medium sized according to Colditz), a total of 184
patients resulted\textsuperscript{55}. With regard to usual ambiguities of a study performance the calculated number should be increased to a sufficiently larger number, say \( n = 220 \) patients. Thus the first stage of the study should be performed with a planned sample size of \( n = 140 \) patients. Actually the p-value of the test for platelets reduction just missed the critical value for stage I (\( p = 0.0107 > \alpha_1 = 0.0102 \)) so that the stage II of the study was performed as planned. It turned out that for stage II a \( p = 0.3551 \) would be sufficient to make the combined result a positive result: proof of non-inferiority for platelet reduction. A necessary number of \( n = 94 \) patients was calculated, so that an enhanced number of \( n = 110 \) patients seemed reasonable for stage II of the study, so the total sample size for stage I and II was \( n = 250 \).

References:


Appendix: *Members of the ANAHYDRET Study Group*

In addition to the authors, the following investigators (listed in alphabetical order) participated in the study:

**Austria:** G. Gastl (Innsbruck), B. Gisslinger (Vienna), L. Muellauer (Vienna), E. Schloegl (Vienna); **Czech Republic:** K. Indrak (Olomouc), R. Pytlik (Prague); **France:** J. Briere (Paris), J.J. Kiladjian (Paris); **Germany:** M. Beykirch (Munich), B. Gaede (Hannover), M. Griesshammer (Minden), M. Herbrik-Zipp (Weingarten), H.J. Hurtz (Halle), G. Jacobs (Saarbrücken), S.H. Jäcki (Tübingen), U. Keilholz (Berlin), J. Mezger (Karlsruhe), P. Schlag (Würzburg), F. Schriever (Dorfen); **Hungary:** T. Masszi (Budapest); **Italy:** S. Sacchi (Modena); **Lithuania:** R. Grinipti (Kaunas), I. Tavoriene (Vilnius); **Poland:** A. Dmoszynska (Lublin), A. Hellman (Gdansk), W.W. Jedrzejczak (Warszawa), T. Robak (Lodz), A. Skotnicki (Kraków); **Singapore:** Y.T. Goh (Singapore); **Slovenia:** P. Cernelc (Ljubljana);