Protocol B1871039

A PHASE 4 SAFETY AND EFFICACY STUDY OF BOSUTINIB (BOSULIF®) IN PATIENTS WITH PHILADELPHIA CHROMOSOME POSITIVE CHRONIC MYELOID LEUKEMIA PREVIOUSLY TREATED WITH ONE OR MORE TYROSINE KINASE INHIBITORS

Statistical Analysis Plan (SAP)

Version: 3.0

Author: PPD, PhD, PPD, Clinical Statistician

Date: 12-NOV-2020

TABLE OF CONTENTS

APPENDICES

1. AMENDMENTS FROM PREVIOUS VERSION(S)

The SAP version 3.0 amended SAP v2.0 dated September 28, 2017. The main changes are COVID-19 related and additional exploratory analyses which included:

- Impact on analyses from COVID-19;
- Additional analyses of deep molecular response.

The SAP version 2.0 amended the original SAP v1.0 dated March 18, 2014. The main changes to align with protocol amendment #1 included:

- Change the primary endpoint analysis to (1) require maintenance of baseline response for at least one year on-treatment for patients with complete cytogenetic response (CCyR) at baseline, (2) require confirmation of initial on-treatment cytogenetic response (in the absence of major molecular response [MMR]) by a second consecutive response at least 28 days apart, and (3) be based on treated patients with a valid baseline efficacy assessment;
- Addition of secondary endpoints duration of CCyR and duration of MMR;
- Include imputation of cytogenetic response from molecular monitoring data if cytogenetics is not performed;
- Update appendices for molecular, cytogenetic and hematologic response criteria;
- Update required number of 4th line patients from at least 75 to 45.

2. INTRODUCTION

On September 4, 2012, bosutinib was approved by the US Food and Drug Administration (FDA) for the treatment of adult patients with chronic (CP), accelerated (AP), or blast phase (BP) Philadelphia chromosome positive (Ph+) Chronic myelogenous leukemia (CML) with resistance or intolerance to prior therapy (United States Prescribing Information BOSULIF®). More recently, on March 27, 2013, the European Medicines Agency (EMA) granted conditional marketing authorization in the European Union (EU), for the treatment of adult patients with CP, AP and BP Ph+ CML previously treated with one or more tyrosine kinase inhibitors (TKIs) and for whom imatinib, nilotinib and dasatinib are not considered appropriate treatment options (EMA Summary of Product Characteristics).

Furthermore, in 5%–10% of patients with of chronic myelogenous leukemia (CML), the Philadelphia chromosome (Ph) is not identified despite the presence of the associated BCR-ABL molecular abnormality (Ph-negative, BCR-ABL-positive CML) because of submicroscopic rearrangements. Regardless, the presence of the BCR-ABL fusion gene will translate into a BCR-ABL fusion protein with dysregulated tyrosine kinase activity. Therefore, a TKI targeting BCR-ABL is also an effective treatment also for Ph negative CML patients. The FDA and EMA approvals were granted based on the results obtained from the single arm, Phase 1/2 study (3160A4-200-WW, "Study-200") in adult patients with Ph+

leukemias who had failed prior TKI therapy, and with the support of the safety results obtained from the Phase 3 study (3160A4-3000-WW) comparing bosutinib with imatinib in newly diagnosed CP Ph+ CML patients. In addition, the EMA was provided with descriptive narrative information from the bosutinib compassionate use program in patients with Ph+CML who had received at least one prior TKI treatment and had progressed or were intolerant, and otherwise not considered suitable for other TKI therapy, as well as efficacy and safety analyses from a subset of Study-200 patients whose disease had failed prior imatinib and/or dasatinib or nilotinib and were contraindicated for treatment with dasatinib or nilotinib, as well as 4th-line CML patients on the study.

The purpose of this phase 4 study is to fulfill the post-authorization commitment made by Pfizer to the EMA in providing additional safety and efficacy data in approximately 150 Ph+ CML patients with high unmet medical need, including at least 75 CP, AP or BP patients in the 4th or later line treatment setting (ie, after treatment with at least 3 other TKIs). The EMA conditional approval requires completion of this post-authorization study in order to convert the conditional approval to a full marketing authorization approval.

This study was initiated in November 2014 and the original target projection was 75 patients in each cohort. However, the 4th line cohort has not met its planned targeted enrollment objective; whereas, the 2nd/3rd-line cohort has overenrolled (n=106). Therefore, Pfizer decided based on CHMP feedback to stop recruitment after at least 45 patients have been enrolled in the 4th line cohort. It was further agreed, there are sufficient numbers of patients in each cohort to perform an informative descriptive analysis of the safety and efficacy and safety of bosutinib.

2.1. Study Design

This is a single-arm, open-label, non-randomized, multi-center Phase 4 study to evaluate bosutinib (Bosulif ®) in patients with CP/AP/BP Ph+ CML whose disease has failed prior treatment with commercially available TKIs due to drug resistance or intolerance, or are otherwise contraindicated for treatment with commercially available TKIs such as imatinib, dasatinib, or nilotinib (ie, presence of a BCR-ABL mutation or medical condition making commercially available TKIs unsuitable for a patient). Patients will receive bosutinib for at least 4 years from the time of first dose, unless disease progression, unacceptable toxicity, patient withdrawal of consent, death or Sponsor discontinuation of study. Patients discontinuing bosutinib prior to completing at least 4 years of therapy will be followed for survival until they complete at least 4 years of follow-up from the time of first dose. Patients completing at least 4 years of bosutinib with continued benefit may be switched to commercially available therapy at that time.

Patients are allocated to the cohort based on the derived assignment from the baseline disease stage and prior TKI therapy. Patients are either:

- CP Ph+ CML second line (CP2L) resistant or intolerant to imatinib, dasatinib, or nilotinib;
- CP Ph+ CML third line (CP3L) resistant or intolerant to imatinib and dasatinib, imatinib and nilotinib, or dasatinib and nilotinib;

- CP Ph+ CML fourth or later line (CP4L) resistant or intolerant to imatinib, dasatinib, and nilotinib;
- AP Ph+ CML second line (AP2L) resistant or intolerant to imatinib, dasatinib, or nilotinib;
- AP Ph+ CML third line (AP3L) resistant or intolerant to imatinib and dasatinib, imatinib and nilotinib, or dasatinib and nilotinib;
- AP Ph+ CML fourth or later line (AP4L) resistant or intolerant to imatinib, dasatinib, and nilotinib;
- BP Ph+ CML second line (BP2L) resistant or intolerant to imatinib, dasatinib, or nilotinib;
- BP Ph+ CML third line (BP3L) resistant or intolerant to imatinib and dasatinib, imatinib and nilotinib, or dasatinib and nilotinib;
- BP Ph+ CML fourth or later line (BP4L) resistant or intolerant to imatinib, dasatinib, and nilotinib, or
- Ph- CML.

Patients will be summarized based on these cohorts or a combination of these cohorts (eg, CP with 1 or 2 prior lines of TKI therapy includes CP2L and CP3L, CP total includes CP2L, CP3L, and CP4L). AP (AP2L, AP3L, AP4L) and BP (BP2L, BP3L, BP4L) patients combined are known as advanced patients. All Ph- CML patients will be summarized separately. Patients who were both resistant and intolerant to the prior TKI or neither will be considered intolerant to that TKI for analysis purposes.

2.2. Study Objectives

Primary Objectives

- *To estimate the 1-year (52-week) probability of cumulative confirmed Major Cytogenetic Response (MCyR) in CP Ph+ CML patients with 1 or 2 prior lines of TKI therapy.*
- *To estimate the 1-year (52-week) probability of cumulative confirmed MCyR in CP Ph+ CML patients with 3 or more prior lines of TKI therapy.*
- *To estimate the 1-year (52-week) probability of cumulative confirmed Overall Hematological Response (OHR) in AP and BP Ph+ CML with any prior TKI therapy.*

Secondary Objectives

- *To estimate the cumulative probability of MCyR in each disease phase (CP, AP and BP) for Ph+ CML patients.*

- *To estimate the cumulative probability of confirmed OHR in each disease phase (AP and BP) for Ph+ CML patients by number of lines of prior therapy.*
- *To characterize the distributions of best response (molecular, cytogenetic, or hematologic) in the CP, AP, and BP Ph+ CML patient populations.*
- *To estimate the probability of MCyR at 3, 6, 12, 18, and 24 months in the CP, AP, and BP Ph+ CML patient populations.*
- *To estimate the probability of confirmed OHR at 3, 6, 9, 12, 18, and 24 months in the AP and BP Ph+ CML patient populations.*
- *To estimate the probability of cumulative confirmed complete hematologic response (CHR) in the CP, AP and BP Ph+ CML patient populations.*
- *To estimate the probability of cumulative MMR in the CP, AP, and BP Ph+ CML patient populations.*
- *To estimate the duration of CCyR.*
- *To estimate the duration of MMR.*
- *To evaluate the overall safety profile of bosutinib in the study population.*

Exploratory Objectives

- *To obtain preliminary estimates of Progression Free Survival (PFS) and Overall Survival (OS) in 4th or later line CP, AP and BP Ph+ CML patients.*
- *To identify BCR-ABL mutations associated with sensitivity or resistance to bosutinib, as well as to catalog possible acquired resistance mutations.*
- *To explore efficacy (hematologic and molecular response) in Ph- CML patients.*
- *To evaluate patient reported outcomes measures.*
- *To collect banked biospecimens for exploratory research, unless prohibited by local regulations or ethics committee decision.*

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

There are no interim analyses planned for this study. The analysis for the primary objective will be done after all patients have completed one year of follow-up and all data through one year have been collected and processed. There will be one follow-up analysis when all study data are complete.

This is a single-arm, open-label study so unblinding is not required.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

There are no statistical hypotheses in this study.

4.2. Statistical Decision Rules

There are no statistical decision rules for this study and no hypothesis testing will be done.

5. ANALYSIS SETS

5.1. Full Analysis Set

The full analysis set is all patients who receive at least one dose of bosutinib. This is the same as the safety analysis set.

5.2. 'Per Protocol' Analysis Set

Per Protocol analyses will not be performed in this study.

5.3. Safety Analysis Set

The safety analysis set is all patients who received at least one dose of bosutinib. This is the same as the full analysis set thus only the "full analysis set" terminology will be used.

5.4. Other Analysis Sets

The evaluable analysis set are those treated patients with a valid baseline efficacy assessment for the respective endpoint (ie, [1] at least 20 metaphases from the baseline bone marrow cytogenetic assessment or a CCyR with at least 200 cells analyzed from the fluorescence in situ hybridization [FISH] baseline assessment or MMR from the baseline molecular assessment for cytogenetic endpoints or [2] a valid baseline hematologic assessment for hematologic endpoints as defined in Appendix 1 or [3] a valid baseline molecular assessment from the central laboratory for molecular endpoints).

5.5. Treatment Misallocations

B1871039 is a single arm study so there were will be no treatment misallocations. Patients who are enrolled but not treated will not be included in the safety or efficacy analyses.

5.6. Protocol Deviations

The full list of protocol deviations for the study report will be compiled prior to database closure. A listing of COVID-19 related protocol deviations will be provided.

6. ENDPOINTS AND COVARIATES

In general, the baseline value will be defined as the last non-missing value prior to first dose, unless specifically stated otherwise. Assessments performed on Week 1 day 1 will be assumed to be pre-dose per the schedule of activities in the Protocol.

An on-treatment evaluation is defined as any evaluation done during bosutinib therapy or within 28 days of last dose.

6.1. Efficacy Endpoint(s)

Primary

- *Cumulative confirmed MCyR defined as Complete Cytogenetic Response (CCyR) or Partial Cytogenetic Response (PCyR) by 1 year (52 weeks) in 2nd and 3rd line CP patients.*
- *Cumulative confirmed MCyR by 1 year (52 weeks) in 4th and later line CP patients.*
- *Cumulative confirmed OHR defined as Complete Hematological Response (CHR) or Return to Chronic Phase (RCP) by 1 year (52 weeks) in AP and BP patients.*

The local cytogenetic assessments will be used for the primary endpoint analysis in CP patients. The local hematologic assessments will be used for the primary endpoint analysis in AP and BP patients.

Cumulative response is any response on-treatment up through Week 52 (based on the date of assessment) with a 4 week window (≤Week 56). Response is confirmed for cytogenetic and hematologic with 2 consecutive responses at least 28 days apart. The date of confirmed response is the date of the first response. Cytogenetic and hematologic responses are defined in Appendix 2 and Appendix 1. MCyR is defined as a CCyR or PCyR. OHR is defined as CHR or RCP. To be considered a responder, the patient must have maintenance of baseline response for ≥5 weeks for hematologic response or ≥52 weeks for cytogenetic response or an improvement from baseline. Patients with PCyR at baseline must attain CCyR on-treatment to be counted as a cytogenetic responder. Patients with at least MMR and a deeper molecular response than baseline are counted as confirmed CCyR.

Secondary

- *Cumulative MCyR in CP, AP and BP patients.*
- *Cumulative confirmed OHR in AP and BP patients by line of therapy.*
- *Cumulative best response in CP, AP, and BP patients.*
- *MCyR at 3, 6, 12, 18, and 24 months in CP, AP, and BP patients.*
- *Confirmed OHR at 3, 6, 9, 12, 18, and 24 months in AP and BP patients.*
- *Cumulative confirmed CHR.*
- *Cumulative MMR in CP, AP, and BP patients.*
- *Duration of CCyR.*
- *Duration of MMR.*

The local cytogenetic assessments will be used for the secondary endpoint analyses for all patients. The local hematologic assessments will be used for the secondary endpoint analyses for all patients. *Molecular assessments will be performed at a central laboratory. Endpoint analysis for molecular response will be carried out by independent central analysis.*

Cumulative response is any response on-treatment. Landmark "at" responses need to occur within a 2 week window of the landmark for those before Week 39 and within a 4 week window for Week 39 and after (eg, Month 3 is ≥Week 11 to ≤Week 15, Month 18 is ≥Week 74 to ≤Week 82, etc.). Molecular responses are defined in Appendix 2. Response is unconfirmed for cytogenetic and molecular and confirmed for hematologic with 2 consecutive responses at least 28 days apart. The date of hematologic confirmed response is the date of the first response. To be considered a responder, the patient must have maintenance of baseline response while on-treatment or an improvement from baseline. The hierarchy to assess for best response from best to worst is ≥4.5-log reduction of BCR-ABL transcripts from standardized baseline on the international scale ($MR^{4.5}$), MR^4 (≥4-log reduction), MMR (≥3-log reduction, same as MR^3), CCyR, PCyR, CHR, and OHR.

Duration of response (DOR) for CCyR and MMR, is defined as the time from first response to confirmed loss, progression of disease as assessed by the investigator, or on-treatment death due to any cause ([date of loss – date of first response + 1]/30.4). For CCyR, confirmed loss is defined as 2 consecutive assessments with >0 Ph+ metaphases at least 28 days apart. For MMR, confirmed loss is defined as 2 consecutive assessments at least 28 days apart with a <3-log (>0.1%) reduction in transcripts from standardized baseline one of which corresponds to a ≤2-log reduction (≥1%). The date of confirmed loss is the earlier of first loss, progression, or death date. Patients without confirmed loss, progression, or death are censored at the last valid assessment where response could be assessed.

Exploratory

- *PFS.*
- *OS.*
- *Presence of BCR-ABL mutations.*
- *For Ph- CML patients, cumulative best response (OHR and MMR) anytime on treatment.*
- *Patient reported outcome (PRO) endpoints as measured by Functional Assessment of Cancer Therapy-Leukemia (FACT-Leu).*
- *To collect banked biospecimens for exploratory research, unless prohibited by local regulations or ethics committee decision. Collection of banked biospecimens unless prohibited by local regulations or ethics committee decision. Additional information on collection and potential use is provided in the Banked Biospecimens section.*

- Time to MMR, MR^4, $MR^{4.5}$, CCyR, and MCyR.
- Duration of MR^4 and MCyR.
- Time to confirmed transformation on-treatment.

The local cytogenetic assessments will be used for the exploratory endpoint analyses for all patients. The local hematologic assessments will be used for the exploratory endpoint analyses for all patients. *Molecular assessments will be performed at a central laboratory. Endpoint analysis for molecular response will be carried out by independent central analysis.*

Time to response is the time from first dose to first response, unconfirmed for molecular and cytogenetic ([date of first response – first dose date +1]/30.4). Patients without events are censored at the last molecular or cytogenetic assessment where response could be assessed for the respective endpoint.

DOR for MCyR follows the same rules as the secondary endpoint of DOR for CCyR. DOR for MR follows the same rules as the secondary endpoint of DOR for MMR except confirmed loss is defined as 2 consecutive assessments at least 28 days apart with a <4-log (>0.01%) reduction in transcripts from standardized baseline one of which corresponds to a ≤3-log reduction (≥0.1%).

PFS is defined as the time from first dose to objective progression or relapse as assessed by the investigator and noted as the reason for permanent treatment discontinuation on the case report form (CRF) or on-treatment death due to any cause ([date of progression – first dose date +1]/30.4). Patients without events are censored at the last cytogenetic, hematologic, or laboratory assessment where progression could be assessed. Progression is defined as:

- *For CP, at least one of the following criteria:*
  - *Development of AP or BP CML;*
  - *Loss of CHR (in absence of cytogenetic response) confirmed by 2 complete blood counts (CBC) at least 2 weeks apart; Loss of CHR is defined as:*
    - *WBC count that rises to >20.0 x 10^9/L;*
    - *Platelet count that rises to ≥600 x 10^9/L;*
    - *Appearance of extramedullary disease not present at baseline including hepatosplenomegaly by palpation or other extramedullary involvement proven by biopsy;*
    - *Appearance of ≥5% myelocytes + metamyelocytes in the peripheral blood or appearance of blasts or promyelocytes in the peripheral blood:*
      - *Loss of MCyR (loss of CCyR or loss of PCyR);*

- *Increasing white blood count (WBC) in a patient without CHR, defined as doubling of WBC, to >20 x 10^9/L on 2 occasions at least 2 weeks apart (after the first 4 weeks of treatment).*

- *For AP, at least one of the following criteria:*
  - *Development of confirmed BP;*
  - *Loss of previous hematologic response over a 2-week period;*
  - *Loss of CHR is defined as:*
    - *WBC count that rises to >20.0 x 10^9/L;*
    - *Platelet count that rises to ≥600 x 10^9/L;*
    - *Appearance of extramedullary disease not present at baseline including hepatosplenomegaly by palpation or other extramedullary involvement proven by biopsy;*
    - *Appearance of ≥5% myelocytes + metamyelocytes in the peripheral blood or appearance of blasts or promyelocytes in the peripheral blood.*
  - *No decrease from baseline levels (if considered clinically relevant) in percentage blasts in peripheral blood or bone marrow on all assessments over a 4-week period.*

- *For BP (of any phenotype):*
  - *Increasing blasts (if considered clinically relevant) peripheral blood or bone marrow over a 4-week period.*

For DOR and PFS, if there is an unacceptable gap between the date of progression or death and the date of the most recent prior disease assessment (>52 weeks), the event will not be used and censorship will be at the most recent prior assessment.

Time to on-treatment transformation, defined by the first progression bullet above, is the time from first dose to confirmed transformation ([date of transformation – first dose date +1]/30.4). Confirmed transformation is defined as 2 consecutive assessments at least 1 week apart or 1 assessment confirmed by progression of disease or death due to any cause. The date of confirmed transformation is the date of the first transformation. Patients without events are censored at the last hematologic or extramedullary assessment where transformation could be assessed.

BCR-ABL mutational status at baseline and treatment discontinuation will be assessed. Emergent mutations are defined as those mutations which were not present at baseline. Patients who were not assessed for baseline mutations are not assessable for emergent mutations.

OS is defined as the time from first dose to the date of death ([date of death – first dose date +1]/30.4). Patients who have not died will be censored at the last known alive date.

For Ph- CML patients, cumulative best response anytime on treatment will be provided. The hierarchy to assess for best response is $MR^{4.5}$, MR^4, MMR, MR^2 (≥2-log reduction), MR^1 (≥1-log reduction), CHR, and OHR.

PRO endpoints are defined in the outcomes research section.

6.2. Safety Endpoints

Treatment-emergent adverse events (TEAE) are defined as any event increasing in severity from baseline or any new event starting during bosutinib therapy or within 28 days of the last dose of study drug (using 28 day lag).

6.2.1. Exposure and Compliance

Exposure and compliance data will be collected during this study. The primary exposure endpoints while the patient is on treatment are duration of treatment, cumulative dose received, dose over time, reasons for dose modifications (reduction, delay, or escalation), and reasons for treatment discontinuation. Duration of treatment is the time from first non-zero dose to last non-zero dose ([date of last non-zero dose – first dose date +1]/30.4). Planned dose is 500 mg/day.

6.2.2. Laboratory Evaluations, Vital Signs, and Cardiac Evaluations

Laboratory values, electrocardiogram (ECG) data, and left ventricular ejection fraction (LVEF) data from echocardiogram (ECHO) or multiple gated acquisition (MUGA) scans will be collected during this study from local laboratories. Collection time points for these measurements are specified in the protocol. Laboratory values and LVEF data will be graded according to NCI CTC version 4.03. ECG assessments (time intervals) will be graded as Normal: <450, Grade 1: ≥450-480, Grade 2: >480-500, and Grade 3: >500 milliseconds.

If multiple post-baseline observations occur at the same visit, the most serious (highest grade) evaluation will be used for all summaries.

6.2.3. Laboratory Evaluations of Special Interest

Laboratory parameters including but not limited to alanine aminotransferase (ALT), aspartate aminotransferase (AST), serum creatinine, estimated glomerular filtration rate (eGFR), platelets, neutrophils, WBC, and hemoglobin will be summarized. Endpoints include distribution of maximum toxicity, time to first event (time from first dose to date of first event including only non-partial dates), duration of any grade event (time from start date to stop date including only non-partial dates), and cumulative duration of grade 3b/4/5 events for eGFR and grade 3/4 events for others (sum of time from start dates to stop dates including only non-partial dates for all events).

eGFR will be derived based on the Modification of Diet in Renal Disease (MDRD) study formula: 175 x SerumCreatinine$^{-1.154}$ * age$^{-0.203}$ * 1.212 (if patient is black) * 0.742 (if female), creatinine measured in mg/dL. eGFR will be graded according to the Kidney Disease Improving Global Outcomes (KDIGO) Kidney function stages (Grade 1: ≥90, Grade 2: 60-<90, Grade 3a: 45-<60, Grade 3b: 30-<45, Grade 4: 15-<30, and Grade 5: <15 mL/min/1.73 m^2).

6.2.4. Non-study Medications

Non-study medications are coded using the World Health Organization (WHO) Drug Dictionary. The Prior, Concomitant, After (PCA) flag for non-study medications will be derived: prior non-study medications are defined as any non-study medications taken before the first dose of test article administration; concomitant non-study medications are defined as any medications taken during the treatment period (includes the 28 days after the last dose of the test article); after non-study medications are defined as any non-study medications taken during the follow-up period (more than 28 days after last dose).

6.2.5. Cross Intolerance

Patients who discontinued prior TKI therapy due to intolerance of imatinib, dasatinib, or nilotinib and who discontinue bosutinib due to the same AE or AE cluster will be classified as having cross intolerance.

6.2.6. Death

Death data will be collected during this study. Mandatory variables captured are date of death and cause of death. Deaths within 28 days of last dose and death due to study treatment toxicity are safety endpoints.

6.2.7. Adverse Events

A 3-tier approach will not be used to classify AEs since this study is a single-arm non-comparative study.

To ensure accuracy and consistency in data analysis and reporting, all adverse events must be classified using the Medical Dictionary for Regulatory Activities (MedDRA) before database lock (DBL). The severity of AEs will be assessed using the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 4.0. The causal relationship between test article and an AE will be assessed by investigator's judgment.

6.2.8. Adverse Events of Special Interest

Adverse events including ALT, AST, thrombocytopenia (including MedDRA preferred term [PT] Platelet Count Decreased), neutropenia (including PT Neutrophil Count Decreased), leukopenia (including PT White Blood Cell Count Decreased), anemia (including PT Hemoglobin Decreased), diarrhea, nausea, vomiting, cardiac toxicity (definition to be determined before DBL), vascular toxicity (definition to be determined before DBL), and renal toxicity (definition to be determined before DBL) will be summarized. Additional AEs or AE clusters may be summarized if deemed appropriate. AE characteristics to be

summarized include maximum toxicity, time to first event, duration of any grade event, cumulative duration of grade 3/4 events, and successful rechallenge (those patients who were re-dosed and did not discontinue bosutinib due to this AE after having a dose delay).

6.3. Other Endpoints

6.3.1. Demographic, Medical History, and Baseline Characteristics

Demographic and baseline variables will include the following:

- Age ([the date of the signing of the informed consent form – the birth date]/365.25 in integer years), sex, race, height, and weight;
- Medical history;
- Cancer history (prior cancer history including CML and initial and current CML disease stage);
- Prior cancer therapies (interferon, imatinib, dasatinib, and nilotinib) and resistance/intolerance status for prior TKIs;
- Eastern Cooperative Oncology Group (ECOG) performance status.

6.3.2. PK Endpoints

There are no PK data being collected in this study.

6.3.3. PD Endpoints

There are no PD data being collected in this study.

6.3.4. Biomarker Endpoints

Blood and plasma samples will be collected for biobanking if specifically allowed by local EC/IRB. These samples will be retained for potential pharmacogenomic/biomarker analyses related to drug response, unless prohibited by local regulations or ethics committee decision. Formal analysis plans will be developed at a later date and will not be included in this SAP.

6.3.5. Outcomes Research Endpoints

Health Outcomes are measured using the FACT-Leu. The FACT-Leu will be administered at screening, weeks 13, 26, 39, 52, 78, 104, 130, 156, 182 and 208 (End of treatment) where the appropriate translations are available. The health outcomes assessment surveys are for the purpose of exploring the patient's own perceptions about his or her quality of life, and should be completed by the patient only prior to interaction with health care professionals at assigned study visits.

The FACT-Leu is a 44-item, self-reported, reliable and valid assessment of quality of life in patients with leukemia. The FACT-Leu consists of measures health related quality of life across 4 domains: Physical Well-being (7 items; score range, 0–28), Social/Family Well-being (7 items; score range, 0–28), Emotional Well-being (6 items; score range,

0–24), Functional Well-being (7 items; score range, 0–28 and a leukemia-specific subscale, which consists of 17 items (score range, 0–68) that assess patient concerns relating to leukemia. The total score (range, 0-176) can also be calculated. Patients are asked to rate on a scale from '0' ("not at all") to '4' ("very much"), regarding how much each item (ie, pain, fatigue, etc.) was present in the last 7 days; higher scores reflect better quality of life. The minimally important difference (MID), defined as the smallest change in a PRO measure that is perceived by patients as beneficial, or that would result in a clinician considering a change in treatment for this instrument are defined as: 2-3 (PWB), N/A (SWB), 2 (EWB), 2-3 (FWB), 3-7 (FACT-G), 4-7 (LeuS-specific subscale) and 6-12 (FACT-Leu Total).

Each item in FACT-Leu is scored on a scale of 0, 1, 2, 3, and 4 for "not at all" to "very much".

The score for each subscale is the sums of the scores of the items in that subscale. Because some of the items are positively stated and some negatively stated, the scores of the negatively stated items must be reversed so that the larger total scores will always indicate better health state. The FACT-Leu scoring guide identifies those items which must be reversed before being added to obtain subscale totals. The score for FACT-G is the sum of the subscale scores for PWB, SWB, EWB, and FWB. The score for FACT-Leu is the sum of the score for FACT-G plus the score for the subscale leukemia. FACT-Leukemia Trial Outcome Index (FACT-Leu TOI) will be calculated as sum of the PWB, FWB and LeuS sub-scale scores.

6.4. Covariates

Subgroup analyses by gender, age (<65, ≥65), and race (Asian, Black, Other, White) will be explored as deemed appropriate. Additional supportive analyses may be explored if deemed appropriate.

7. HANDLING OF MISSING VALUES

In general, no imputation such as last observation carried forward (LOCF) will be used. For the time-to-event endpoints, such as time to response, DOR, time to transformation, PFS, and OS, the primary missing data handling method will be censoring as described in Section 6.1.

In compliance with Pfizer standards for handling partial dates, if the month and year are present but day is missing, start date is set to 1st of month, and stop date is set to last day of month. If year is present but day and month are missing, start date is set to January 1st, and stop date is set to December 31st.

If there are missing items for the PRO endpoints, subscale scores can be prorated. This is done by multiplying the sum of subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done on the scoring guide or by using the formula below:

Prorated subscale score = [Sum of item scores] x [N of items in subscale] ÷ [N of items answered]

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items were answered (eg, a minimum of 4 of 7 items, 4 of 6 items, etc). The total score is then calculated as the sum of the unweighted subscale scores. The FACT scale is considered to be an acceptable indicator of patient quality of life as long as overall item response rate is greater than 80% (eg, at least 22 of 27 FACT-G items completed). This is not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if greater than 50% of items are answered.

Handling of missing/incomplete patient central molecular data will be done by the following procedure: for post-baseline visits, assessments at a specific visit without a minimum number of ABL transcripts specified by the central laboratory will be not used for analysis.

Handling of missing/incomplete patient local cytogenetic data will be done by the following procedure: for baseline visits, only bone marrow information with ≥20 metaphases will be used, otherwise the baseline response will be set to missing. For post-baseline visits, bone marrow information with 20 to 99 metaphases will be used over FISH information with at least 200 nuclei at a specific visit. FISH will only be used to assess CCyR. No conventional bone marrow cytogenetic assessment will be derived when ≥100 cells are reported as analyzed for Philadelphia chromosome as this will be considered an error in data entry. If on-treatment molecular monitoring is performed in lieu of conventional cytogenetics or FISH, CCyR may be imputed on a specific date if a MMR is achieved on that date. The results of conventional cytogenetic testing, FISH with ≥ 200 nuclei or banding with ≥20 metaphases, will be used in preference to molecular testing, provided they are available and performed on the same date.

Handling of missing/mismatched patient hematologic efficacy data will be done using the following procedure: for each patient, the observations will be sorted by visit and the relative day to first dose.

First, for each observation, peripheral blood and/or bone marrow differentials should add up to 100%. Some sites may not record 0% to reflect uncounted cell types as instructed. Instead, the corresponding field may appear as missing or “ND” even though data collection procedures have been vigorously followed. Also, the differentials may not exactly add up to 100%. To correct this problem, each domain of the differentials will be summed together. If the total is between 98.5% and 101%, the missing values are assigned 0%. Otherwise, the observation is left as is and the record is also to be reported as a data issue.

Second, for the set of variables collected to derive peripheral blood or bone marrow differentials at each post-baseline visit up to week 13, if a given set are still incomplete, a non-missing set of lab variables obtained ±3 days from the scheduled visit will be utilized to complete the derivation. For visits on or after week 13, a non-missing set of lab variables obtained ±7 days of the visit will be utilized to complete that post-baseline assessment. The last non-missing peripheral blood or bone marrow differentials collected at week 1 day 1 or up to 30 days prior to week 1 day 1 will be used for baseline assessment. For example, if the bone marrow blast count was missing, the entire non-missing bone marrow assessment, not just the blast field, from another observation will be used. Some data handling rules for data combination are specified below.

- Data combination is needed only when the full dataset is not available on a single day at a given visit. The full dataset includes peripheral blood differential for chronic phase, and both bone marrow (when available) and peripheral blood differentials for advanced phase. The extramedullary assessment when absent should not preclude the derivation, and when present should be taken into account per response criteria rules.

- During the post-baseline data combination, the latest date will be used as the date of visit and the more conservative value will be taken if there is more than one measurement for a variable. For platelet counts, in case of multiple values, the lowest value will be taken. If any value falls above or below normal range, the out-of range value will be taken.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Analyses of Continuous Data

For continuous variables, the descriptive statistics include n, mean, median, standard deviation (SD), and range (minimum, maximum).

8.1.2. Analyses of Categorical Data

For categorical variables, the descriptive statistics include the count in each category, the total n, and percentage.

8.1.3. Analyses of Binary Endpoints

Exact 2-sided 95% CIs will be provided for molecular, cytogenetic, and hematologic response rates.

8.1.4. Analyses of Time-To-Event Endpoints

For time-to-event endpoints such as time to response, DOR, time to transformation, PFS, and OS, the method of analysis will either be Kaplan-Meier (K-M) or cumulative incidence of the event adjusting for the competing risk of treatment discontinuation due to any reason without the event (Gray).[1]

When applying the K-M method, medians and quartiles with the associated 2-sided 95% CI will be provided. The CIs based on the Brookmeyer-Crowley log(-log) transformation method will be provided. The K-M yearly rates up through year 4 with the associated 2-sided 95% CI based on Greenwood's formula using a log(-log) transformation will also be provided.

When applying the cumulative incidence method, yearly rates up through year 4 with the associated 2-sided 95% CI based on the delta method using a log(-log) transformation will be provided.

8.2. Statistical Analyses

The efficacy endpoints will be analyzed according to the cohorts or a combination of these cohorts specified in Section 2. Ph- CML patients will be summarized separately.

8.2.1. Analysis of Primary Efficacy Endpoints

The response rate analyses are based on the evaluable analysis set for Ph+ CML patients only. The primary efficacy analysis is the percentage of responders by Week 52, done separately for:

- Confirmed MCyR for combined CP2L and CP3L;
- Confirmed MCyR for CP4L;
- Confirmed OHR separately for AP and BP.

Any patient not demonstrating a response by Week 52 is deemed a non-responder.

8.2.2. Analysis of Secondary Efficacy Endpoints

The response rate and DOR analyses are based on the evaluable analysis set for Ph+ CML patients only. The efficacy analysis is the percentage of responders cumulative anytime on-treatment or at a specific landmark, for:

- Cumulative MCyR separately for CP, AP, and BP;
- Cumulative confirmed OHR separately for AP2L, AP3L, AP4L, BP2L, BP3L, BP4L;
- Best response separately for CP, AP, and BP;
- Landmark MCyR at 3, 6, 12, 18, and 24 months separately for CP, AP, and BP;
- Landmark confirmed OHR at 3, 6, 9, 12, 18, and 24 months separately for AP and BP;
- Cumulative confirmed CHR separately for CP, AP, and BP;
- Cumulative MMR separately for CP, AP, and BP;
- Duration of CCyR separately for CP, AP, and BP;
- Duration of MMR separately for CP, AP, and BP.

Any patient not demonstrating a response on-treatment for cumulative and at the specific visit for landmark is deemed a non-responder for the respective analysis. For landmark analyses, if the patient is missing the landmark assessment however the patient is responding at the scheduled before and after visit (eg, missing Month 6/Week 26 but responding at Week 13 and Week 39, missing Month 18/Week 78 but responding at Week 52 and

Week 104, etc.), then the patient will be deemed a responder for that landmark. For cumulative MMR (including MR^4 and $MR^{4.5}$) and MCyR (including CCyR), supportive analyses will be provided excluding patients with baseline response and also for patients who responded before and after initial dose reduction to 400 mg, 300 mg, and 200 mg due to an AE

The duration of cytogenetic and molecular response analyses are based on the evaluable analysis set for the subset of patients who respond for the respective endpoint. The K-M method, as described in Section 8.1, will be used for estimation of the distributions of duration provided most responding patients remain on treatment until progression or loss of response. The percentages of CCyR and MMR for the respective endpoint, along with percent of censored patients still on-treatment and percent of reasons for failure will be summarized.

No COVID-19 related efficacy analyses will be performed as the primary endpoint only summarizes data up through Week 52 and all patients reached that milestone prior to the March 11, 2020 COVID-19 anchor date and for secondary endpoints, >80% of patients were already off treatment as of the anchor date. A summary of missing data before and after the anchor date will be summarized.

A listing of COVID-19 related study discontinuations will be provided.

8.2.3. Analysis of Exploratory Efficacy Endpoints

The time to molecular and cytogenetic response analyses are based on the evaluable analysis set. Cumulative incidence of response adjusting for the competing risk of treatment discontinuation due to any reason without response, as described in Section 8.1, will be used. The percentage of responders without treatment discontinuation and the percentage of treatment discontinuation without response will be summarized.

The duration of cytogenetic and molecular response exploratory analyses follow the same rules as the secondary DOR analyses except the percentages of MCyR, CCyR, and PCyR and MR^4 will be summarized.

The time to transformation analysis is based on the full analysis set. Cumulative incidence of transformation adjusting for the competing risk of treatment discontinuation due to any reason without transformation, as described in Section 8.1, will be used. The percentage of transformation without treatment discontinuation and the percentage of treatment discontinuation without transformation will be summarized. The specific disease stage transformation (accelerated or blast phases) will also be listed.

The PFS analysis is based on the full analysis set. The K-M method, with or without insufficient clinical response as assessed by the investigator as an event or cumulative incidence of on-treatment progression or death adjusting for the competing risk of treatment discontinuation without the event, as described in Section 8.1, will be used.

The OS analysis is based on the full analysis set. The K-M method, as described in Section 8.1, will be used. The number of deaths and causes of death will be summarized. Time on study from date of first dose to date of last contact and reasons for discontinuing participation will also be summarized.

The percentage of patients with baseline and emergent mutations, along with the specific types of mutations, will be summarized. The percentage of patients with cumulative MCyR, CHR, MMR, and OHR (AP and BP patients only) by baseline mutation status will be provided. The distributions of PFS and OS will also be displayed for patients with a baseline mutation versus those assessed but without a baseline mutation.

The percentage of Ph- CML patients in each best response category will be summarized.

Missing values and incomplete data will be handled as described in Section 7 of this analysis plan.

Listings for all efficacy data will also be provided.

8.2.4. Analysis of Safety Endpoints

The full analysis set will be used for all safety analyses unless specifically stated otherwise. The safety endpoints will be analyzed according to the cohorts or a combination of these cohorts as specified in Section 2.

8.2.4.1. Baseline Characteristics

Listings of demography and baseline characteristics data will be provided. The baseline variables listed below will be summarized using descriptive statistics.

- Age in years;
- Race;
- Sex;
- Weight (kg);
- Height (cm);
- ECOG performance status;
- Cancer history (prior cancer history including CML, initial and current CML disease stage, duration since CML diagnosis);
- Prior cancer therapies (number of prior regimens for interferon, imatinib, dasatinib, and nilotinib), duration of prior therapy, and reasons for resistance/intolerance status of prior TKIs;
- Medical history by MedDRA System Organ Class (SOC) and PT.

8.2.4.2. Treatment and Compliance

Descriptive statistics for continuous exposure variable treatment duration and cumulative dose received and the received dose over time (% receiving 600, 500, 400, 300, 200, or 0 mg at specific time points) will be calculated.

The number and percentage of patients with at least one dose reduction due to AE, dose delay due to AE, and escalation to 600mg will be summarized. The number and percentage of patients who experienced more than 1 event will also be summarized.

The number and percentage for reasons for treatment discontinuation will be summarized and listed. A listing of COVID-19 related treatment discontinuations will be provided.

8.2.4.3. Laboratory Evaluations, Vital Signs, and Cardiac Evaluations

The incidence of clinical laboratory, QTcF/QTcB, and LVEF abnormalities by maximum toxicity grade will be presented at baseline, during the treatment period, and post treatment period. Shifts from baseline to maximum on-treatment grade for QTcF/QTcB and LVEF will also be presented.

Laboratory values and vital signs will be provided in the data listings.

8.2.4.4. Laboratory Evaluations of Special Interest

The number and percentage of maximum on-treatment grade 0 through 4 for serum creatinine, ALT, AST, platelets, neutrophils, WBC, and hemoglobin and grade 1 through 5 for eGFR will be summarized. Descriptive statistics for time to first event, duration of any grade event, and cumulative duration of grade 3b/4/5 for eGFR and grade 3/4 events for others will be calculated. Shifts from baseline to maximum on-treatment grade will also be presented.

A listing will be provided for patients who potentially meet the Hy's law criteria. Potential is defined as meeting these 3 criteria any time during the study (non-concurrent): AST or ALT level of ≥3 x upper limit of normal (ULN), total bilirubin level of >2 x ULN, and alkaline phosphatase level <2 x ULN. Total Bilirubin, AST, ALT, and alkaline phosphatase will be listed.

8.2.4.5. Non-study Medications

The percentage of patients using non-study medications prior to, during, and after the treatment period will be summarized.

A listing of non-study medications data will be reported.

8.2.4.6. Cross Intolerance

The percentage of patients experiencing cross intolerance to imatinib, dasatinib. or nilotinib will be summarized overall and by AE. The recurrence percentage of those grade 3/4 TEAEs during treatment with bosutinib will also be estimated.

8.2.4.7. Death

The number of deaths within 28 days of last dose and cause of these deaths will be summarized.

8.2.4.8. Adverse Events

The number and percentage of AEs listed below will be summarized overall and by MedDRA SOC and PT. Sorting will be done alphabetically by SOC and then by decreasing frequency of PT. AEs may also be summarized by other levels if deemed appropriate (eg, Higher Level Group Term [HLGT]).

- TEAEs;
- AEs leading to treatment discontinuations;
- TEAEs leading to dose reductions;
- TEAEs leading to dose delays;
- TEAEs by maximum toxicity grade (grades 1 to 5);
- TEAEs of higher toxicity grades (grades 3 or 4);
- TEAEs related to test article;
- Serious adverse events (SAEs);
- COVID-19 related TEAEs;
- TEAEs before and after initial dose reduction to 400 mg, 300 mg, and 200 mg due to an AE.

8.2.4.9. Adverse Events of Special Interest

The number and percentage of maximum treatment-emergent CTCAE grade events and successful rechallenge for ALT, AST, thrombocytopenia (including MedDRA PT Platelet Count Decreased), neutropenia (including PT Neutrophil Count Decreased), leukopenia (including PT White Blood Cell Count Decreased), anemia (including PT Hemoglobin Decreased), diarrhea, nausea, vomiting, cardiac toxicity, vascular toxicity, and renal toxicity will be summarized. Descriptive statistics for time to first event, duration of any grade event, and cumulative duration of grade 3/4 events will be calculated for selected parameters.

8.2.4.10. Additional Safety

Listings of physical examination data, chest x-ray, and ECOG performance status will be provided.

8.2.5. Analyses of Patient Disposition Endpoints

Listings of violations of inclusion/exclusion criteria will be reported. This study will include the summaries of the numbers of patients who: sign the informed consent form (ICF), are screening failures, are enrolled, and are part of the full/evaluable analysis sets. The number of subjects discontinuing study participation, along with the reason, and descriptive statistics for the duration on-study (time from first dose to last contact) will also be calculated.

8.2.6. Analyses of Outcome Research Endpoints

All health outcomes analysis will be based on patients in the Full Analysis Set with available data.

The Outcomes Research endpoints are FACT-Leu overall, FACT-G overall, FACT-G subscales, Leukemia subscale (LeuS), and FACT-Leu Trial Outcome index. Univariate descriptive statistics will be used in the analysis of all the above endpoints.

For each of the endpoints listed above, summary statistics including Mean, Median, Standard Deviation, Minimum, Maximum, 95% CI will be reported at each time point. This will be done based on observed values as well as based on changes from baseline. A graphical display of the observed means over time for each of the endpoints will also be provided.

The above analyses will be done separately for each cohort and combined.

9. REFERENCES

1. A class of k-sample tests for comparing the cumulative incidence of a competing risk. Gray, RJ. 3, 1988, Annals of Statistics, Vol. 16, pp. 1141-1154.

APPENDICES

Appendix 1. CRITERIA TO IDENTIFY PHASE OF CML

Phase [a]	***Findings***
Blast Phase	*≥30% Blasts in blood or bone marrow* *Extramedullary Blasts proliferation, apart from spleen* *If either of these 2 criteria is present, the patient is classified as Blast Phase regardless of other criteria.*
Accelerated Phase	*15-29% Blasts in blood or bone marrow* *or* *>30% Blasts + Promyelocytes (with Blasts <30%) in blood or bone marrow* *OR* *≥20% Basophils in blood* *OR* *Persistent platelets <100 x 10^9/L (not related to therapy) OR* *Clonal chromosomal abnormalities in Ph+ cells (CCA/Ph +), major route, on treatment*
Chronic Phase	*<15% Blasts* *<20% Basophils* *≤30% Blasts + Promyelocytes (with Blasts <15%) in blood or bone marrow* *AND* *Platelets ≥100 x 10^9/L (unless related to therapy)* *AND* *No Extramedullary Blasts (except spleen)*

[a] *Observance of any one criterion in the worse disease phase achieves that phase as the diagnosis (eg, 20% basophils alone elevates the diagnosis from CP to AP).*

Appendix 2. DEFINITIONS OF HEMATOLOGIC, CYTOGENETIC, AND MOLECULAR RESPONSE

Response by Type	*Definitions*
Hematologic ***Complete (CHR)***	 ***WBC <10 x 10^9/L*** ***Peripheral Blood Basophils <5%*** ***No Peripheral Blood myelocytes, promyelocytes, myeloblasts in the differential*** ***Platelet count <450 x 10^9/L*** ***Spleen nonpalpable***
Cytogenetic * ***Complete (CCyR)*** ***Partial (PCyR)*** ***Minor (mCyR)*** ***Minimal (minCyR)*** ***None (noCyR)***	 ***No Ph+ metaphases*** ***1% to 35% Ph+ metaphases*** ***36% to 65% Ph+ metaphases*** ***66% to 95% Ph+ metaphases*** ***>95% Ph+ metaphases***
Molecular ***Deep Molecular Response (MR4, MR4.5)***	***MR4= either (i) detectable disease with <0.01% BCR-ABL1 ^{IS} or (ii) undetectable disease in cDNA with >10 000 ABL1 transcripts*** ***MR4.5 =either (i) detectable disease with < 0.0032% BCR-ABL1 ^{IS} or (ii) undetectable disease in cDNA with >32,000 ABL1 transcripts in the same volume of cDNA used to test for BCR-ABL1.*** ***Ratio of BCR-ABL^{IS} ≤0.1%**.***
Major (MMR)	

Abbreviations: CHR, complete hematologic response; CCyR, complete cytogenetic response, PCyR, partial cytogenetic response; Ph+, Philadelphia Chromosome positive; MCyR, major cytogenetic response; mCyR, minor cytogenetic response; minCyR, minimal cytogenetic response; noCyR, no cytogenetic response; MR4, MR4.5, deep molecular response; MMR, major molecular response.
**If bone marrow metaphases cannot be obtained or evaluated by chromosome banding analysis, the definition of CCyR (defined as <1% for FISH) may be based on interphase FISH of blood cells, provided that it is performed with Dual-Color Dual-Fusion or Dual-Fusion FISH probes and at least 200 nuclei are scored. PCyR and CCyR are counted together and reported as MCyR.*
***With ≥5,000 ABL1 transcripts.*

Definitions of Hematologic Response for AP and BP CML*

*CHR***	*WBC ≤ institutional upper limit of normal (ULN)* *ANC ≥1.0 x 109/L* *Platelet count ≥100 x 109/L* *BM Blasts ≤ 5%**** *No peripheral blood blasts or promyelocytes;* *Peripheral blood myelocytes + metamyelocytes <5%;* *Basophils in peripheral blood <5%** *No evidence of extramedullary involvement*

**The maximum % of basophils in peripheral blood was modified from <2 to <5% for a less restrictive constraint and alignment with the CHR criteria for CP CML patients according to ELN guidelines.*
***If an on- treatment CBC is scheduled within 1 week from the last platelet transfusion or last dose of G-CSF, it should be considered "Not Evaluable" for hematological response assessment, and a CBC should be repeated at least 1 week from the last platelet transfusion or last dose of G-CSF received.*
****CHR for AP/BP that the 5% BM blasts is only required when a BM differential is available.*