Janssen Research & Development

Statistical Analysis Plan

A Phase 2 Study of Cusatuzumab Plus Azacitidine in Patients with Newly Diagnosed Acute Myeloid Leukemia who are not Candidates for Intensive Chemotherapy

Protocol 74494550AML2001; Phase 2

JNJ-74494550 (Cusatuzumab)

Status:	Approved
Date:	22 April 2021
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-499105, 1.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze the data and report results for study JNJ-74494550-AML-2001.

JNJ-74494550-AML-2001 is an open-label, 2-part, multicenter, Phase 2 study to evaluate efficacy and safety of 2 dose levels of cusatuzumab (10 mg/kg and 20 mg/kg) in combination with azacitidine in participants with previously untreated acute myeloid leukemia (AML) who are not eligible for intensive chemotherapy.

1.1. Study Objectives

1.1.1. Primary Objective

The primary objective is to determine the complete response (CR) rate per ELN 2017 (Dohner 2017) of 2 dose levels of cusatuzumab in combination with azacitidine.

1.1.2. Secondary and Exploratory Objectives

Secondary objectives are to evaluate the following:

- Rate of CRh
- Rate of CR + CRh
- Rate of CRi
- Overall response rate (ORR) = CR + CRh + CRi
- Rate of CR without MRD
- Time to response, defined as time from randomization/enrollment to achieving the first response of CRi, CRh, or CR
- Duration of response, defined as time from achieving first response of CRi, CRh, or CR to disease relapse, progression or death of any cause
- Safety profile of adverse events (AE) and serious adverse events (SAE)
- Transfusion independence (TI), defined as having a period of at least 56 consecutive days with no red blood cell or platelets transfusion between randomization and the last dose of study drug + 30 days
- Pharmacokinetics (PK) of cusatuzumab alone and in combination with azacitidine
- Immunogenicity/anti-drug antibody of cusatuzumab alone and in combination with azacitidine

Exploratory objectives are to summarize the following:

- Progression-free survival (PFS), defined as the time from randomization (part 1) or enrollment (part 2) to progressive disease, relapse from CR/CRh/CRi, or death from any cause, whichever occurs first
- Overall survival (OS), defined as the time from randomization (part 1) or enrollment (part 2) to death from any cause

- Exploratory biomarkers including baseline level of CD70+ blasts, CD70+ LSC, sCD27, NK cells: to assess the pharmacodynamics and to explore biomarkers predictive of clinical response
- PROs using the FACT-Leu and EQ 5D 5L

1.2. Trial Design

The study consists of 2 sequential parts, Part 1 and Part 2. The purpose of Part 1 is to select a dose level (10 mg/kg or 20 mg/kg) based on the totality of the data, including CR rate and other efficacy endpoints, PK, biomarkers, and safety. The purpose of Part 2 is to further evaluate and confirm the efficacy and safety of the combination therapy at the selected dose of cusatuzumab.

During Part 1, participants are randomly assigned in a 1:1 ratio, with stratification by types of AML (de novo versus secondary AML) and ECOG score (0-1 versus 2), in one of the 2 treatment groups:

- Group A: Azacitidine 75 mg/m^2 SC or IV on Day 1 through Day 7, cusatuzumab 10 mg/kg IV on Day 3 and Day 17 of each 28-day cycle
- Group B: Azacitidine 75 mg/m^2 SC or IV on Day 1 through Day 7, cusatuzumab 20 mg/kg IV on Day 3 and Day 17 of each 28-day cycle

A 3-stage monitoring approach, as the statistical guidance, is utilized in Part 1 to select a dose level (see Section 1.5, Interim Analysis). The cumulative number of participants for each dose group at each of the 3 stages is 15, 30, and 50, respectively. Available data will be evaluated after the last enrolled participant has been treated for 4 cycles with the combination therapy.

Additionally, after 5 participants have enrolled on each arm and been treated for 2 cycles, cumulative safety data will be evaluated by the DRC to ensure continued study treatment is safe.

For Part 2, the treatment group to be studied will be dependent upon the results from Part 1. Approximately 100 participants are included for the selected treatment group from Parts 1 and 2 combined. A formal interim analysis of the primary endpoint, CR rate, is planned during Part 2 to include 80 participants from the selected dose group. The final analysis will be based on 100 participants.

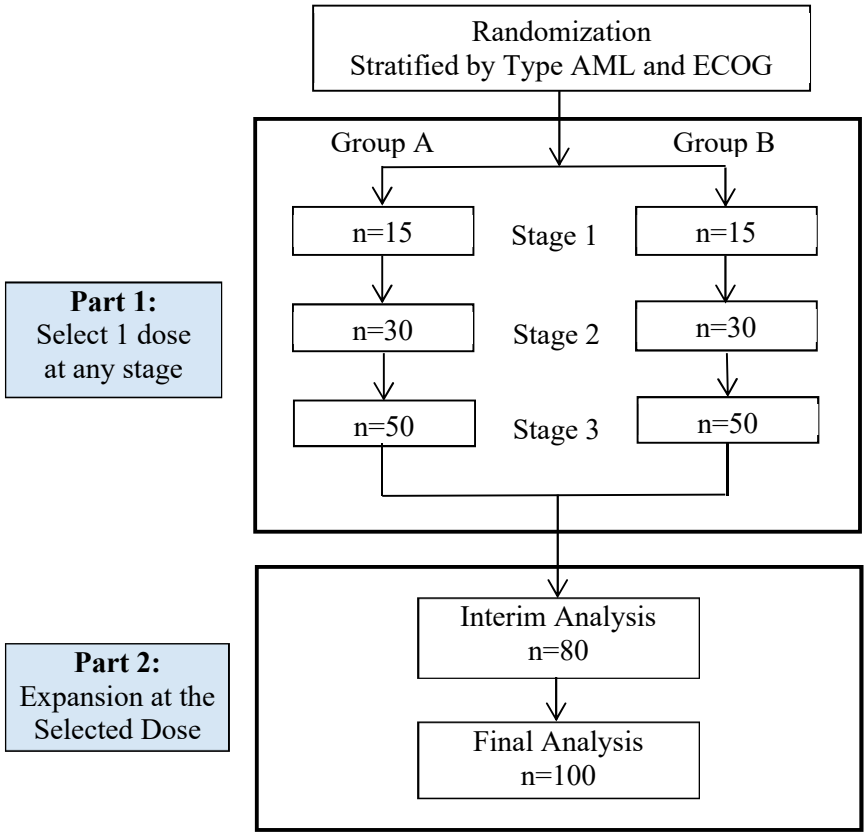
Upon 1 dose level being selected to continue, ongoing participants who are not receiving the selected cusatuzumab dose may switch to that dose. This will be following review and approval of the selected dose by the Data Review Committee (DRC) (see below.

A diagram of the study design is provided in Figure 1.

A Data Review Committee (DRC) will be established to review cumulative safety and efficacy data at planned interim time points, to ensure participants' safety, to evaluate efficacy results, and to make recommendations on dose selection, study continuation, and other study conduct related

matters. There will be a separate DRC charter that describes the composition of the DRC, and the role and responsibilities of the DRC members.

Figure 1.



Notes:
- **Group A: A**zacitidine 75 mg/ m^2 SC or IV on D1 through D7, cusatuzumab 10 mg/kg IV on Day 3 and 17 of each 28 day cycle.
- **Group B**: Azacitidine 75 mg/ m^2 SC or IV on D1 through D7, cusatuzumab 20 mg/kg IV on Day 3 and 17 of each 28 day cycle.
- Number of participants (n) at each Stage of Part 1 and Interim and Final Analyses are cumulative.

1.3. Statistical Hypothesis for Trial Primary Objective

The primary hypothesis is that participants treated with the combination of cusatuzumab (at a dose level of 10 mg/kg or 20 mg/kg) and azacitidine can achieve a CR rate of 35% or greater against the null hypothesis of 20%.

1.4. Sample Size Justification

Up to 50 participants will be randomized into each dose group in Part 1 of the study to select a dose and continue into Part 2. Participants will be enrolled into Part 2 at the selected dose until a cumulative total of 100 participants have been treated with the selected dose from Part 1 and Part 2. The study has a >90% power to reject the null hypothesis of a CR rate of 20% under the alternative hypothesis of 35%. This sample size is based on the single-arm Wilson Score test without continuity correction at an overall 1-sided type 1 error rate <2.5%.

1.5. Interim Analyses

For each analysis (3 stages in Part 1 and the formal interim and final analysis in Part 2), the data cutoff will be the time point when the last participant in the respective analysis has been treated for a minimum of 3 cycles (Part 1) or 4 cycles (Part 2) of the combination therapy.

Study Part 1

Part 1 of the study consists of 3 stages of monitoring to select a dose schedule. The cumulative number of participants for each dose group at each of the 3 stages are 15, 30, and 50, respectively. At each stage, each dose group's response rate will be determined independently (without formal statistical comparison to the other group) if that dose group will be stopped or continued into the next stage or into study Part 2. Only 1 dose group will be selected for Part 2. The statistical criteria for such a determination will be based on the number (and percentage) of participants in each dose group achieving CR.

The purpose of the 3-stage monitoring in Part 1 is to select, or to eliminate a dose, which will be based on the Pocock beta-spending function. The overall beta level is chosen as 10%, allocated to the 3 stages with respective and cumulative sample sizes of 15 (30%), 30 (60%), and 50 (100%) participants per group; the corresponding nominal significance level (calculated against the alternative hypothesis of 35% CR rate) are 0.04157, 0.04618, and 0.05632. Per this approach, the statistical decision criteria and corresponding inferences are summarized in Table 1, where 2 possible outcomes resulting in a stop or continuation decision are presented for each stage.

The statistical criteria shown in Table 1 ensure that the false negative error rate (ie, the probability of a stopping decision, when the true CR rate is ≥35%) is ≤6% at any stage; or equivalently, the upper limit of the 1-sided >94% confidence interval (CI) of the observed CR rate is <35% (which means that a CR rate of 35% or greater has been statistically ruled out). It's important to control the false negative error rate at a low level when the number of participants is small to ensure that a stopping decision will not be made unless it is statistically convincing that a CR rate of ≥35% has been ruled out. Since all the analyses in study Part 1 are futility evaluations without the intention to stop for efficacy, the overall type I error of the entire study is not inflated.

If the number of participants at each stage is not exactly as specified in Table 1, the Pocock beta-spending function with an overall beta level of 10% will be adjusted based on the actual number participants.

The above statistical criteria for Part 1 will be served as guidance; the totality of the data including other efficacy outcomes (eg, duration of CR), available safety, and PK/PD data will be considered in making a decision at each stage. If both dose schedules warrant proceeding into Part 2 per the above statistical criteria, only one dose schedule will be extended into Part 2 based on the observed CR rate as well as available data of other efficacy, safety, and PK/PD.

Table 1: Statistical Decision Parameters for Study Part 1: Observed CR, 1-Sided Confidence Interval (CI), and Statistical Decision

	Number (%) of Observed CR	1-Sided CI Upper Bound[1] (%)	Decision	False Negative Error[2]	False Positive Error[3]
Stage 1, n=15	1 (6.7) 2 (13.3)	26.7 34.9	Stop	≤4%	NA
	3 (20.0) 4 (26.7)	42.3 49.3	Continue to Stage 2	NA	≤50%
Stage 2, n=30	5 (16.7) 6 (20.0)	30.5 34.3	Stop	≤5%	NA
	7 (23.3) 8 (26.7)	37.9 41.5	Continue to Stage 3	NA	≤32%
Stage 3, n=50	11 (22.0) 12 (24.0)	32.3 34.4	Stop	≤6%	NA
	13 (26.0) 14 (28.0)	36.6 38.7	Continue to Part 2	NA	≤14%

[1] 96%, 95%, and 94% for Stage 1, 2, and 3, respectively.
[2] False negative error: probability of a "stop" decision, when the true CR rate is ≥35%.
[3] False positive error: probability of a "continue" decision, when the true CR rate is ≤20%.
Note: All statistical inferences are based on Wilson Score statistics without continuity correction.

Study Part 2

For Part 2, 1 interim analysis of CR rate is planned to include 80 participants. The null hypothesis of a CR rate of 20% will be rejected if the 1-sided nominal p-value is ≤0.01221. If the null hypothesis is not rejected at this interim analysis, the final analysis will be performed with 100 participants, and the null hypothesis will be rejected if the 1-sided nominal p-value is ≤0.02144. The O'Brien-Fleming alpha-spending, and Wilson Score test without continuity correction is to be implemented for the interim and final analyses.

1.6. Changes of Study Conduct and Planned Analyses

The study did not proceed into Part 2 per decision by study Sponsor and the DRC upon completion of the analyses in all 3 stages in study Part 1. Data from study Part 1 will be summarized with descriptive statistics without formal comparison or hypothesis testing.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Subject Populations

Intent-to-treat (ITT): all participants randomized in Part 1, grouped according to assigned treatment arm. ITT will be used for all summaries of disposition, demographic, baseline disease characteristics, and efficacy analyses.

Modified Intent-to-treat (MITT): all participants randomized in Part 1 who have received at least 1 dose of both study drugs, grouped according to assigned treatment arm. MITT will be used for selected efficacy analyses.

Safety Analysis Set: participants in the ITT population who have received at least 1 dose of study drug. This analysis set will be used for all safety analyses and analyses of exposure.

Pharmacokinetics Analysis Set: participants who have received at least 1 dose of cusatuzamab in combination with azacitidine and whose PK profiles allow for accurate calculation of at least 1 of the PK parameters.

2.2. Definition of Subgroups

Overall response rate will be summarized for the following subgroups:

- Age (<75, ≥75 years)
- Gender (male, female)
- Type of AML (de novo, secondary)
- Cytogenetic risk (adverse, intermediate, favorable)
- Baseline ECOG performance status (0-1, 2)
- Baseline bone marrow blasts (≤50%, >50%)
- ANC (<1.0, ≥1.0 10^9/L)
- Platelet (<100, ≥100 10^9/L)

2.3. Study Day and Relative Day

Study Day 1 is the day of randomization for participants randomized in Part 1. By convention, 1 month equals to 30.4375 days, and 1 year equals to 365.25 days. Duration in days between 2 time points will be calculated as (end date – start date + 1).

2.4. Baseline

Baseline value is defined as the last non-missing value collected on or before study Day 1.

3. SUBJECT INFORMATION

3.1. Demographics and Baseline Characteristics

All demographic and baseline characteristics will be summarized for the ITT population by treatment group and all participants combined.

3.2. Disposition Information

Participant disposition will be summarized for the ITT population by treatment group and total combined with the number and percentages of the followings: study treatment ongoing, discontinued treatment and reason, completed/discontinued the study and reasons.

3.3. Protocol Deviations

Participants with major protocol deviations will be summarized by each category of deviations.

3.4. Prior and Concomitant Medications

Prior and concomitant medications will be coded to generic terms based on the World Health Organization (WHO) dictionary and will be summarized by Anatomical Therapeutic Chemical (ATC) class and drug generic term. Prior anticancer therapies include radiotherapy, surgery, chemotherapy, and systemic therapy.

Premedication for cusatuzumab will be summarized separately.

3.5. Extent of Exposure

The following will be summarized by treatment group, and within each treatment group, by azacitidine and cusatuzumab: total dose prescribed, total dose administered, dose intensity (mg/m^2/cycle azacitidine and mg/kg/cycle for cusatuzumab), and relative dose intensity (actual dose intensity divided by prescribed dose intensity).

Total treatment duration in months, total number of cycles, and cycle delay will be summarized by treatment group.

Number and percent of participants who switched the cusatuzumab dose level will also be provided.

4. EFFICACY

4.1. Primary Efficacy Endpoint, CR Rate

Number and percent of participants achieving CR will be provided by treatment arm and the 2 arms combined along with 95% CI. Analyses will be performed for ITT and MITT populations.

4.2. Secondary and Exploratory Efficacy Endpoints

4.2.1.1. Overall Response and MRD Negativity

Number and percent of participants achieving the following best response outcomes, along with 95% CI, will be provided:

- CRh
- CRi
- MLFS
- CR+CRh
- CR+CRh+CRi (overall response rate, ORR)
- PR
- SD
- PD
- CR without MRD

Subgroup analysis will be provided for ORR as forest plots.

4.2.1.2. Time to Response

Time to response is defined as the time from randomizationto achieving the first response of CR, CRh, or CRi for participants who achieved such a response. Descriptive statistics will be summarized by treatment arm using the Kaplan-Meier method (with no censored observations).

4.2.1.3. Duration of Response

Duration of response is defined for participants who achieved the first response of CR, CRh, or CRi as the time from achieving such a response to relapse or deaths due to any cause. Duration will be censored at the time of last tumor assessment for participants who did not relapse or die. Duration of response will be summarized descriptively using the Kaplan-Meier method.

4.2.1.4. Transfusion Independent

Number and percent of participants achieving transfusion independence (TI), along with 95% CI, will be provided for:

- Red blood cell TI
- Platelets TI
- Red blood cell and platelets TI

Time to and duration of transfusion independence for participants achieving TI will be summarized similarly to time to and duration of response. Time to TI is calculated from randomization to 1 day after the last transfusion episode prior to achieving TI; and duration of TI is calculated from the day of achieving TI to the day of the first transfusion episode after the TI period.

Summaries will be provided for red blood cell and platelet transfusions separately. The number and percent of participants who received transfusions from the date of randomization to 30 days after the last dose will be summarized by treatment arm.

4.2.1.5. Progression-Free Survival

Progression-free survival (PFS) is defined as the time from randomizationto disease progression, relapse from CR, CRh, or CRi, or death from any cause, whichever occurs first. Participants who did not reach any of these events will be censored at the last disease assessment.

Number and percent of participants reaching PFS events (overall, disease progression, relapse, or death) will be provided. Kaplan-Meier estimates of the survival functions will be presented graphically. Median and 25^{th} and 75^{th} percentiles along with associated 95% CI will be provided. Analyses will be performed for ITT and MITT populations.

4.2.1.6. Overall Survival

Overall survival (in months) is defined as the time from randomization to death from any cause. For participants who are not known to have died, survival time will be censored at the date last known to be alive.

Number and percent of deaths will be provided. Kaplan-Meier estimates of the survival functions will be presented graphically. Median and 25th and 75th percentiles along with associated 95% CI will be provided. Analyses will be performed for ITT and MITT populations.

5. SAFETY

All safety analyses will be based on the safety analysis set based on actual treatment received, unless otherwise specified. Selected safety parameters may also be summarized separately for participants who have cusatuzumab dose crossover if number of such participants permits.

For all continuous safety variables, descriptive statistics will include the N, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized using frequency counts and percentages.

Safety assessment will be evaluated through AEs, clinical hematology and chemistry laboratory tests, death, vital signs, electrocardiogram (ECG) and ECOG performance scores. Safety analyses will be based on the safety analysis set and presented by the actual treatment group.

5.1. Adverse Events

The verbatim terms used in the eCRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Toxicities will be graded for severity according to NCI-CTCAE, version 5.0. Intervention-emergent adverse events are adverse events with onset during the intervention phase or that are a consequence of a pre-existing condition that has worsened since baseline. All reported adverse events will be included in the analysis. For each adverse event, the percentage of participants who experience at least 1 occurrence of the given event will be summarized by intervention group. In addition, comparisons between intervention groups will be provided if appropriate.

Specifically, the following will be summarized:

- All adverse events
- Grade 3 or higher adverse events
- Serious adverse events
- Adverse events leading to discontinuation of treatment
- Adverse events leading to death
- Adverse events of special interest

Summaries, listings, datasets, or participant narratives may be provided, as appropriate, for those participants who die, who discontinue intervention due to an adverse event, or who experience a severe or a serious adverse event.

Parameters with predefined National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) toxicity grades will be summarized. Change from baseline to the worst adverse event grade experienced by the participant during the study will be provided as shift tables.

5.2. Clinical Laboratory Tests

Laboratory data will be summarized by type of laboratory test. Reference ranges and markedly abnormal results (specified in the Statistical Analysis Plan) will be used in the summary of laboratory data. Descriptive statistics will be calculated for each laboratory analyte at baseline. Parameters with predefined NCI-CTCAE toxicity grades will be summarized. Change from baseline to the worst adverse event grade experienced by the participants during the study will be provided as shift tables.

6. PATIENT-REPORTED OUTCOMES (PRO)

Patient-reported outcome instruments included in this study are the Functional Assessment of Cancer Therapy – Leukemia (FACT-Leu) and EuroQoL 5 Dimension 5 Level questionnaire (EQ-5D-5L).

The FACT-Leu is a 44-item instrument measuring health-related quality of life concepts. It is comprised of 4 subscales that make up the FACT-G: physical well-being (PWB, 7 items), social/family well-being (SWB, 7 items), emotional well-being (EWB, 6 items), and functional well-being (FWB, 7 items) and a fifth subscale for leukemia-specific concerns (LeuS, 17 items). Items are rated on a response scale from 0 (not at all) to 4 (very much). A Trial Outcome Index (FACT-Leu TOI) score can also be calculated by combining the FACT-PWB, FACT-FWB, and FACT-LeuS subscales. Subscale and total scores will be calculated according to instrument administration and scoring guidelines; higher scores indicate better health-related quality of life.

The EQ-5D consists of 5 items describing health state and the EuroQoL visual analog scale (VAS) of self-rated overall health. Responses for the 5 descriptive items are converted into a single index value or utility score according to EQ-5D scoring algorithm ranging from -1 to 1 where lower scores indicate worse health status. Utility scores will be calculated using the UK weighting algorithm. EQ-5D-VAS responses range from 0 (the worst health you can imagine) to 100 (the best health you can imagine).

For purposes of determining whether patients have experienced a clinically meaningful change in a given score, the following thresholds for minimum important difference (MID) will be used:

Scale/Sub-Scale	Threshold for definition of worsening or improvement
FACT-Leu Total Score	6
FACT-G (General) Scale	4
FACT-Leu TOI	4
Physical Well-Being (PWB)	2
Social/Family Well-Being (SWB)	2
Emotional Well-Being (EWB)	2
Functional Well-Being (FWB)	2
Leukemia Symptoms/concerns (LeuS)	3
EQ-5D-5L Utility (UK weight)	0.08
EQ-5D-5L VAS	7

Change thresholds for the FACT-Leu scales selected from ranges provided by (Whiteley, 2016).
Change thresholds for EQ-5D-5L scores from (Pickard, 2007; McClure 2017).

PRO Completion

Patient completion of each PRO instrument will be summarized as a percentage (number of completed out of number expected) at each collection time point by dose group.

Patient completion will be summarized using the following categories:

- All questions completed
- At least 50% completed
- At least one question completed

(Note: Items noted as 'optional' in the instrument, specifically item GS7 "I am satisfied with my sex life" will not be considered missing for purposes of tabulating PRO completion.)

PRO Overall Instrument Scores

PRO total scores will include the FACT-Leu Total, FACT-Leu TOI, FACT-G, EQ-5D-5L Utility Score, and EQ-5D-5L VAS. (The EQ-5D-5L VAS is a single item but will be categorized with the overall instrument scores as it represents the concept of overall health status.)

For each overall instrument score, a table of descriptive statistics will be presented including mean, standard deviation, median, minimum and maximum at each time point and change from baseline for each post-baseline assessment, by dose group. Each overall instrument score will also be shown in line graphs of mean score and mean change from baseline by dose groups, with 95% CI.

PRO Domain Scores

PRO domain scores will include the FACT-PWB, FACT-FWB, FACT-EWB, FACT-SWB, and FACT-LeuS.

For each instrument domain score, a table of descriptive statistics will be presented including mean, standard deviation, median, minimum, and maximum at each time point and change from baseline for each post-baseline assessment, by dose group. Each overall instrument score will be shown in line graphs of mean score and mean change from baseline by dose groups, with 95% CI.

PRO Item Analyses

Selected individual items from the FACT-Leu addressing concepts of particular interest in the current study context of use will also be described in terms of their category responses. Specifically:

Item number	Item text
GP1	"I have lack of energy"
GP2	"I have nausea"
GP4	"I have pain"
GP5	"I am bothered by the side effects of treatment"
GP6	"I feel ill"
GF1	"I am able to work (including work at home)"

BRM3	"I am bothered by fevers (episodes of high body temperature)"
P2	"I have certain parts of my body where I experience pain"
An7	"I am able to do my usual activities"

For each item, a bar chart will show the distribution of responses (including missing responses), and a second bar chart will show the distribution of change in response categories from baseline (including missing responses).

7. REFERENCES

Jennifer Whiteley, Arlene Reisman, Mark Shapiro, JorgeE. Cortes & David Cella (2016) Health-related quality of life during bosutinib (SKI-606) therapy in patients with advanced chronic myeloid leukemia after imatinib failure, Current Medical Research and Opinion, 32:8, 1325-1334, DOI: 10.1185/03007995.2016.1174108.

Pickard A, Neary M, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007; 5(70).

McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-defined estimates of the minimally important difference for EQ-5D-5L Index Scores. Value Health. 2017 Apr; 20(4):644-650. doi: 10.1016/j.jval.2016.11.015.