CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

125553Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES- BIOSIMILAR PRODUCT

NDA/BLA #:	BLA 125533
Supplement #:	Original.
Drug Name:	EP2006, biosimilar to reference product Neupogen
Indication(s):	Cancer patients receiving myelosuppressive chemotherapy
	Patients with acute myeloid leukemia receiving induction or consolidation chemotherapy
	Cancer patients receiving bone marrow transplant
	Patients undergoing peripheral blood progenitor cell collection and therapy
	Patients with severe chronic neutropenia
Applicant:	Sandoz Biopharmaceuticals
Date(s):	Stamp Date: May 8, 2014
	BsUFA Goal Date: March 8, 2015
Review Priority:	Standard
Biometrics Division:	Division of Biometrics V
Statistical Reviewer:	Kyung Yul Lee, Ph.D.
Concurring Reviewers:	Lei Nie, Ph.D., Lead Mathematical Statistician Thomas Gwise, Ph.D., Deputy Director Rajeshwari Sridhara, Ph.D., Division Director
Medical Division:	Division of Hematology Products
Clinical Team:	Dr. Donna Przepiorka, MD, Ph.D., Clinical Reviewer
	Dr. Albert Deisseroth, MD, Ph.D., Clinical Team Leader
Project Manager:	Jessica Boehmer, RPM, MBA

Keywords:

Link to keywords: biosimilar, non-inferiority, duration of severe neutropenia, febrile neutropenia, ANCOVA, , equivalence, no clinically meaningful difference

Table of Contents

1 EXECUTIVE SUMMARY 6

2 INTRODUCTION 8

2.1 OVERVIEW 8
2.2 DATA SOURCES 8

3 STATISTICAL EVALUATION 9

3.1 DATA AND ANALYSIS QUALITY 9
3.2 EVALUATION OF EFFICACY 9
3.2.1 Study Design and Endpoints 9
3.2.2 Statistical Methodologies 11
3.2.3 Patient Disposition, Demographic and Baseline Characteristics 12
3.2.4 Results and Conclusions 14
3.3 EVALUATION OF SAFETY 22

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 22

4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION 22
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS 23

5 SUMMARY AND CONCLUSIONS 24

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 24
5.2 CONCLUSIONS AND RECOMMENDATIONS 24

LIST OF TABLES

Table 1 : List of all studies included in analysis 8
Table 2 : Planned Treatment Groups 9
Table 3 : Primary Reason for Premature Treatment and Study Discontinuations 12
Table 4 : Analysis Population Sets 13
Table 5 : Demographic Characteristics: Randomized Population 13
Table 6 : Disease Characteristics: Randomized Population 14
Table 7 : DSN in Cycle 1: PP population 15
Table 8 : DSN with ANC <1x10^9/L in Cycle 1 15
Table 9 : DSN in Cycle 1: FAS population 16
Table 10: DSN in Cycle 1: FAS population Excluding Subjects with Exposure of Commercial Drug and Subjects in Site 703 17
Table 11: Results for sensitivity analyses for DSN in Cycle 1 17
Table 12: Sponsor's Depth and Time Occurred ANC nadir in Cycle 1: PP population 20
Table 13: Sponsor's Time to ANC Recovery in Cycle 1: PP population 20
Table 14: Sponsor's Incidence of FN in Cycle 1: PP population 21
Table 15: Incidence of FN in Cycle 1: FAS population 21
Table 16: Sponsor's Number of days in fever in Cycle 1: PP population 21
Table 17: Subgroup Analyses of DSN: Age and Region (90% CI): PP population 23
Table 18: Subgroup Analyses for DSN: Baseline Disease Characteristics (90% CI): PP population 23

LIST OF FIGURES

Figure 1 : Daily Mean ANC in Cycle 1: FAS population 18
Figure 2 : Sponsor's Time course of ANC in Cycle 1 (PP set) 19
Figure 3 : Mean Daily Maximum Temperatures in Cycle 1: FAS Population 22

1 EXECUTIVE SUMMARY

Sandoz submitted a biologics license application BLA125553 under section 351(k) of the Public Health Service Act (PHS Act) to support EP2006 as a biosimilar product to US-licensed Neupogen (filgrastim). Sandoz is seeking licensure of EP2006 for the same indications as currently approved for Neupogen: The indications are as follows:

1) to decrease the incidence of infections, as manifested by febrile neutropenia, in patients with nonmyeloid malignancies receiving myelosuppressive anti-cancer drugs associated with a significant incidence of severe neutropenia with fever;

2) for reducing the time to neutrophil recovery and the duration of fever, following induction or consolidation chemotherapy treatment of adults with AML;

3) to reduce the duration of neutropenia and neutropenia-related clinical sequelae, e.g., febrile neutropenia in patients with nonmyeloid malignancies undergoing myeloablative chemotherapy followed by marrow transplantation;

4) for the mobilization of hematopoietic progenitor cells into the peripheral blood for collection by leukapheresis; and

5) for chronic administration to reduce the incidence and duration of sequelae of neutropenia (e.g., fever, infections, oropharyngeal ulcers) in symptomatic patients with congenital neutropenia, cyclic neutropenia, or idiopathic neutropenia.

To support a demonstration of biosimilarity, a stepwise approach was used following the FDA's scientific recommendation. The stepwise approach starts with structural and functional characterization of both the proposed biosimilar product and the reference product. Results of nonclinical and/or clinical studies follow to assess remaining questions with regards to potential residual uncertainty about biosimilarity.

This review is to evaluate the results of the clinical study, EP06-302 (PIONEER) which was a randomized, double-blind, parallel-group, multi-center study of EP2006 and Neupogen® in histologically proven breast cancer patients. Patients eligible for neoadjuvant or adjuvant treatment were treated with myelosuppressive TAC chemotherapy (Taxotere® [docetaxel 75 mg/m2] in combination with Adriamycin® [doxorubicin 50 mg/m2] and Cytoxan® [cyclophosphamide 500 mg/m2]), all given IV on day 1 of each of six 21-day cycles).

A total of 192 patients were planned to be assigned into four arms (48/group) randomly; Group 1 EP2006 for Cycle 1 through 6; Group 2 EP2006 for Cycles 1, 3, and 5 and Neupogen for Cycles, 2, 4, and 6; Group 3 Neupogen cycles 1, 3, and 5 and EP2006 for Cycles 2, 4, and 6; Group 4 Neupogen for Cycles 1 through 6 (See Table 2).

The pre-specified primary objective of this study was to demonstrate non-inferiority of EP2006 versus Neupogen® (US-licensed) with respect to the mean duration of severe neutropenia

(DSN), which was defined as the number of consecutive days with grade 4 neutropenia (absolute neutrophil count [ANC] less than 0.5×10^9/L), during Cycle 1 of the neoadjuvant or adjuvant TAC regimen in breast cancer patients.

The primary endpoint was the duration of severe neutropenia (DSN) in days in cycle 1 and analysis conducted in the per-protocol population (PP) (101 patients in the EP2006 group and 103 patients in the Neupogen group). The randomization stratification factor was kind of therapy (adjuvant therapy vs. neoadjuvant therapy). The primary analysis was analysis of covariance with covariates treatment status (adjuvant vs neoadjuvant) and baseline absolute neutrophil count, based on the per-protocol population (the subgroup of subjects who received treatment and had no major protocol violations).

No similarity margins for equivalence testing were proposed by the sponsor. The data provided in the submission could be used to evaluate the claim that the products are similar by considering the width of the confidence interval for the difference in mean DSN. If the difference is sufficiently small (±1 day) with a narrow confidence interval, one might conclude that the difference is not clinically meaningful.

We conclude that there was no clinically meaningful difference between the EP2006 group and the Neupogen group with respect to the efficacy endpoint results. The mean DSN in Cycle 1 was 1.17 days and 1.20 days for EP2006 and Neupogen, respectively. The 90% CI of the mean difference is (-0.21, 0.28). The analysis showed that EP2006 is equivalent to Neupogen in terms of efficacy as measured by the mean difference of DSN between EP2006 and Neupogen being less than 1 day for both the upper and lower bounds of the 90% CI

Our conclusion is consistent with the advisory committee's recommendation. The advisory committee meeting for oncology drug products was held on January 7, 2015 for this application. The advisory committee voted unanimously (14-0) that EP2006 should receive licensure as a biosimilar product for each of the five indications for which US-licensed Neupogen is currently approved.

2 INTRODUCTION

2.1 Overview

Granulocyte colony-stimulating factor (G-CSF) is a lineage-specific colony-stimulating factor which is produced by monocytes, fibroblasts, and endothelial cells. G-CSFs restore the number of neutrophils and keep the neutrophil count above the critical level at which febrile neutropenia (FN) can occur. The clinical use of recombinant human G-CSF (rhG-CSF) is to reduce the duration of neutropenia and the incidence of febrile neutropenia (FN) in patients with malignancies treated with myelosuppressive chemotherapy regimens as well as to reduce the duration of neutropenia in patients undergoing myeloablative therapy prior to bone marrow transplantation.

The first approved recombinant human G-CSF is Amgen's filgrastim (Neupogen®). The European Commission granted a marketing authorization valid throughout the EU for Ratiograstim® (a biosimilar filgrastim) to ratiopharm GmbH on September, 2008. The FDA and European Medicines Agency (EMA) approved in 2002 the first second-generation, recombinant methionyl form of human G-CSF (PEG-r-metHuG-CSF) that is pegylated under the INN pegfilgrastim.

In February 2009, EP2006 was approved by the European Medicines Agency (EMA) in the same indications as those of EU-approved Neupogen® and unrestricted renewal of the authorization has been granted by the EMA in the meantime.

Study EP06-302 was designed to demonstrate non-inferiority of EP2006 to US-licensed Neupogen® in the prevention of neutropenic complications in breast cancer patients treated with established myelosuppressive chemotherapy.

Table 1 : List of all studies included in analysis

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
EP006-302	*Phase 3*	*18 weeks*	*6 weeks*	*192 (48/arm)*	*Breast cancer*

2.2 Data Sources

The study report and data were provided electronically; the location/names of study report, analysis datasets (ADAM) including STDM datasets and SAS programs are as follows;

Study Reports:
\\cdsesub1\bla\eCTD_Submissions\STN125553\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\neutropenia-scn-pbpc-hiv\5351-stud-rep-conntr-sr-ep06-302
\\cdsesub1\bla\eCTD_Submissions\STN125553\0005\m1\us\

Dataset
\\cdsesub1\bla\eCTD_Submissions\STN125553\0000\m5\datasets\ep06-302\analysis\datasets

\\cdsesub1\bla\eCTD_Submissions\STN125553\0005\m5\datasets\ep06-302\analysis\datasets

Program
\\cdsesub1\bla\eCTD_Submissions\STN125553\0000\m5\datasets\ep06-302\analysis\programs

\\cdsesub1\bla\eCTD_Submissions\STN125553\0005\m5\datasets\ep06-302\analysis\programs

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

Reviewer reviewed the quality and integrity of the submitted data. Examples of relevant issues include the following:

- It is possible to reproduce the primary analysis dataset, and in particular the primary endpoint, from the original data source.
- The sponsor didn't provide subgroup analysis results at the initial BLA submission, so we requested subgroup results through information request.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

This study was a randomized, double-blind, parallel-group, multi-center study comparing the efficacy and safety of EP2006 and Neupogen® in histologically proven breast cancer patients treated with TAC combination chemotherapy (docetaxel 75 mg/m^2, doxorubicin 50 mg/m^2, and cyclophosphamide 500 mg/m^2). A total of 192 patients were randomized to either EP2006 or US-licensed Neupogen® in four groups (48/group) from 25 centers; 10 centers in Russia, 6 centers in Ukraine and 6 centers in Hungary, 1 center in Latvia, 1 center in Slovakia, and 1 center in Czech Republic. The four groups are as follows;

Table 2 : Planned Treatment Groups

Group	n	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
1 EP	48	EP2006	EP2006	EP2006	EP2006	EP2006	EP2006
2 EPNEU	48	EP2006	Neupogen	EP2006	Neupogen	EP2006	Neupogen
3 NEUEP	48	Neupogen	EP2006	Neupogen	EP2006	Neupogen	EP2006
4 NEU	48	Neupogen	Neupogen	Neupogen	Neupogen	Neupogen	Neupogen

The patients underwent TAC combination chemotherapy (docetaxel 75 mg/m^2, doxorubicin 50 mg/m^2, and cyclophosphamide 500 mg/m^2), administered intravenously on Day 1 of each chemotherapy cycle and given for six cycles with 3 weeks /cycle. Study drug (EP2006 or Neupogen®) was administered daily starting on Day 2 of each chemotherapy cycle (at least 24 hours after chemotherapy ended) and continued until the ANC recovered to 10×10^9/L after the nadir or up to a maximum of 14 days (whichever occurred first). EP2006 and Neupogen® were injected subcutaneously with a daily dose of 5 mcg/kg body weight.

The total study duration was up to 24 weeks, including up to three weeks screening, approximately 18 weeks of active treatment (6 TAC chemotherapy cycles), and a follow-up visit about six weeks after the start of the last cycle (approximately four weeks after the last study medication administration).

Patient's ANC, platelet values and hemoglobin values had to be above the defined limits (ANC $\geq$1.5 x 10^9/L, platelets $\geq$100 x 10^9/L, and Hemoglobin $\geq$ 10g/dL) at the Day 1 of Cycle 1.

In Cycle 1, blood samples for the determination of the ANC were taken on Day 1, daily until the ANC recovered to 10×10^9/L after the nadir or until Day 15, whichever occurred first.
In Cycles 2 to 6, blood samples were taken on Day 1 prior to chemotherapy and daily from Day 7 onwards until the ANC recovered to 10×10^9/L after the nadir or until Day 15, whichever occurred first.

Primary Endpoint:

The primary efficacy endpoint was the mean DSN in Cycle 1. The DSN was set to 0 in patients who did not experience severe neutropenia in Cycle 1. In patients who experienced several episodes of severe neutropenia, the number of days for each episode was summed up.

Secondary endpoints:

The key two secondary endpoints were the depth of ANC nadir and time to ANC recovery. The depth of ANC nadir was defined as the patient's lowest ANC in a chemotherapy cycle. Time to ANC recovery was defined as the time from ANC nadir day until the patient's ANC increases to $\geq$ 2 x 10^9/L day after the nadir in cycle 1.

The depth of ANC nadir was analyzed with descriptive statistics for Cycle 1 and for each cycle. A descriptive analysis was performed for the combined treatment groups 1 + 2, and 3 + 4 only. If the nadir was $\geq 2 \times 10^9$/L for all time points after administration of chemotherapy the time was set to 0 day.

The other secondary endpoints were the incidence of FN, the number of days of fever, the frequency of infections and duration of hospitalization due to FN.

The incidence of FN was calculated as the number of patients with at least one episode of FN divided by the number of patients at risk in a given time interval (in each cycle the period

between Day 2 to Day 15 was considered for the analysis). FN was defined as having both an oral temperature ≥ 38.3°C and an ANC < 0.5×10^9/L on the same day. The incidence of FN was analyzed separately for each cycle and over all cycles (overall incidences).

3.2.2 Statistical Methodologies

The following rules were pre-specified to treat missing data in assessing the primary endpoint.

- The ANC before and after the missing day was ≥ 0.5×10^9/L: the day could be ignored as a potential day of severe neutropenia.
- If at both neighboring days the ANCs were < 0.5×10^9/L, then the missing day was to be set to severe neutropenia.
- If the day before was < 0.5×10^9/L and the day after ≥ 0.5×10^9/L, then the missing day was to be set to severe neutropenia.
- If the day before was ≥ 0.5×10^9/L and the day after < 0.5×10^9/L, then the missing day was to be set to severe neutropenia.
- If any of the neighboring days were also missing, severe neutropenia could not be determined and the data remained missing.

The primary efficacy endpoint was analyzed using an analysis of co-variance (ANCOVA) with treatment group, kind of chemotherapy and baseline ANC value as a covariate.

The full analysis set (FAS) included all randomized patients who received at least one dose of study medication. The per protocol (PP) set is a subset of the FAS including those patients who completed the first chemotherapy cycle without major protocol deviations. The primary analysis population was the PP population. The primary endpoint was additionally analyzed based on the FAS as a sensitivity analysis to evaluate the robustness of the results.

A one-sided 97.5% Clopper-Pearson CI for the difference of overall FN incidence between the switched and un-switched (between EP2006 and Neupogen®) patients was calculated. Switching was to be considered non-inferior to not switching if the lower bound of the one-sided 97.5% CI was above the non-inferiority margin of -15%.

No similarity margins for equivalence testing were proposed by the sponsor. The data provided in the submission could be used to evaluate the claim that the products are similar by considering the width of the confidence interval for the difference in mean DSN. If the difference is adequately small with a narrow confidence interval, one might conclude that the difference is immaterial.

The maximum daily temperature was analyzed with descriptive statistics separately for each cycle and over all cycles. Fever was defined as an oral temperature ≥ 38.3°C. The number of patients who had fever at least once was presented with counts and percentages for each cycle and over all cycles.

Sample Size Calculation
The sample size was calculated based on the non-inferiority of EP compared to Neupogen® concerning the DSN defined as days with ANC <0.5 x 10^9/L in cycle 1. The non-inferiority margin was set to 1 day and non-inferiority should be regarded as confirmed if the upper limit of the two-sided 95% CI for the difference of the expected DSN between EP and Neupogen® would be less than 1 day. Assuming the difference between EP and Neupogen® of 0.25 days in favor of Neupogen® and the standard deviation of about 1.5 days, the sample size should be at least 86 patients per treatment group assuring 90% power. Based on primary analysis population with per-protocol population, 10% of the randomized patients were expected to be excluded in the per-protocol population. The sample size of 192 patients (96/treatment group) was planned.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

A total of 258 patients were screened and 218 patients were randomized and 40 patients were excluded; 3 due to not meeting inclusion criteria, 2 due to meeting exclusion criteria, and 34 due to other reasons. The first patient enrolled date was on December 26, 2011 and the last patient entered on June 17, 2013.

Among 218 randomized patients, 54 patients were allocated to EP group, 55 patients were allocated to EPNEU group, 55 patients were allocated to NEUEP group and 54 patients were allocated to NEU group.

A total of 34 patients did not complete the study or discontinued study treatment prematurely; 29 patients did not complete the study drug treatment and 33 patients did not complete the study.

The primary reason for premature treatment and study discontinuations are summarized in Table 3.

Table 3 : Primary Reason for Premature Treatment and Study Discontinuations

Primary reason	EP N=54 n (%)	EPNEU N=55 n (%)	NEUEP N=55 n (%)	NEU N=54 n (%)	Total N=218 n (%)
Treatment Discontinuation					
Withdrawal	5 (9.3)	3 (5.5)	3 (5.5)	2 (3.7)	13 (6.0)
Lost to follow up	0	0	1 (1.8)	1 (1.9)	2 (0.9)
Death	1 (1.9)	0	0	0	1 (0.5)
Physician decision	1 (1.9)	1 (1.8)	3 (5.5)	0	5 (2.3)
Other	2 (3.7)	2 (3.6)	0	4 (7.4)	8 (3.7)
Withdrawal from the study					
Withdrawal	4 (7.4)	3 (5.5)	3 (5.5)	2 (3.7)	12 (5.5)
Lost follow up	1 (1.9)	1 (1.8)	2 (3.6)	1 (12.9)	5 (2.3)
Death	1 (1.9)	0	0	0	1 (0.5)
Other	6 (11.1)	3 (5.5)	3 (5.5)	3 (8.8)	15 (6.9)

A total of 29 patients discontinued the treatment. A total of 33 patients were withdrawn from the study. Among 218 randomized patients, 14 patients had major protocol deviations. The primary reasons for the protocol deviation were due to administration of commercial filgrastim (9 patients); and due to no study drug during the cycles (5 patients).

The sponsor's analysis population sets are summarized in Table 4.

Table 4 : Analysis Population Sets

	EP N=54 n (%)	EPNEU N=55 n (%)	NEUEP N=55 n (%)	NEU N=54 n (%)	Total N=218 n (%)
ITT	54	55	55	54	218
FAS	53	54	55	52	214
SAF	53	54	55	52	214
PP	50	51	52	51	204
PP-I	40	45	44	46	175

SAF = Safety (set); FAS = Full analysis set; PP = Per protocol (set); SAF-I = Safety interchangeability (set); PP-I = Per protocol interchangeability (set); sponsor's Table 11-1

Among 218 randomized patients, 214 patients were treated with study drug (full analysis set (FAS)) after excluding 4 patients who were not treated or only treated with commercial filgrastim. This is the same with the safety analysis population (SAF). The protocol deviations were 4 patients from EP; 4 patients EPNEU, 3 patients NEUEP and 3 patients from NEU. After excluding 14 protocol deviation patients from 218 randomized patients, the PP included 204 patients. There were 4 patients who were treated in the study, but did not receive the study drug after Cycle 1, 19 patients who did not complete all six cycles, and 16 patients who completed all six cycles, but had major protocol violations. After excluding all 39 patients from 214 PP populations, the PP-I included 175 patients. The analyses population for switched (Group 1 [EP] and 4 [NEU]) vs. un-switched (Group 2 [EPNEU] and 3 [NEUEP]) was PP-I.

The patients' demographics are summarized in Table 5.

Table 5 : Demographic Characteristics: Randomized Population

	EP (N=109) n (%)	NEU (N=109) n (%)
Age		
Mean (SD) (years)	49.4 (11.5)	48.4 (10.9)
<65	100 (91.7)	98 (89.9)
≥65	9 (8.3)	11 (10.1)
Chemotherapy		
Adjuvant	64 (58.7)	62 (56.9)
Neo-adjuvant	45 (41.3)	47 (43.1)

Region		
Russia	81 (74.3)	87 (79.8)
Ukraine	17 (15.6)	16 (14.7)
Other	11 (10.1)	6 (5.5)

The demographic characteristics were similar between the two groups. Mean ages were 49 years in the EP2006 group and 48 years in the Neupogen group. Patients who had adjuvant therapy were 59% in the EP2006 group and 57% in the Neupogen group. The most patients were enrolled from Russia in both groups.

The disease characteristics are summarized in Table 6.

Table 6 : Disease Characteristics: Randomized Population

	EP (N=109) n (%)	NEU (N=109) n (%)
Stage		
I	7 (6.4)	8 (7.3)
II	57 (52.3)	55 (50.5)
III	45 (41.3)	46 (42.2)
Surgery		
Yes	86 (78.9)	83 (76.2)
No	23 (21.1)	26 (23.9)
Radio Therapy		
Yes	9 (8.3)	10 (9.2)
No	100 (91.7)	99 (90.8)
ECOG Status		
0	84 (77.1)	84 (77.1)
1	25 (22.9)	25 (22.9)
Months since first diagnosis (months)		
Mean (SD)#	2.8 (16.3)	1.2 (1.9)

The disease characteristics were similar between the two groups. Majority patients were with breast stage II or III, ECOG score 0, yes surgery, no radio therapy. The mean months since first diagnosis were 2.8 months in the EP006 group and 1.2 months in the Neupogen group. The difference was one outlier (171 months) in the EP2006 group. The median months since first diagnosis were both one month.

3.2.4 Results and Conclusions

The primary analysis is summarized in Table 7, which are the same as the sponsor's.

Table 7 : DSN in Cycle 1: PP population

	EP (N=101)		NEU (N=103)
DSN			
Mean (SD)	1.17 (1.11)		1.20 (1.02)
Difference *		-0.04	
95% CI		(-0.33, 0.26)	
90% CI		(-0.28, 0.21)	
DSN (days), n (%)			
0	37 (36.6)		32 (31.1)
1	23 (22.8)		30 (29.1)
2	32 (31.7)		30 (29.1)
3	5 (4.9)		10 (9.7)
4	4 (4.0)		1 (1.0)

Sponsor's Table 11-4

The mean DSN in Cycle 1 was 1.17 days and 1.20 days for EP2006 and Neupogen, respectively. The estimated mean difference of DSN was -0.04 days and the upper limit of 95% of 0.26 (95% CI: -0.33, 0.26) which is below 1 day of non-inferiority margin. The analysis showed that EP2006 is equivalent to Neupogen in terms of efficacy as measured by the difference of DSN between EP2006 and Neupogen being less than 1 day for both the upper and lower bounds of the 90% CI

Reviewer's comment:
We believe that the equivalence margin of 1 day is appropriate. Please refer to Dr. Gootenberg's clinical review in STN125031, dated Jan 31, 2002, for the basis for use of DSN as a surrogate for FN and the non-inferiority margin of 1 day in DSN was used. Dr. Gootenberg also stated " a 1-day difference in DSN would be anticipated to result in approximately a 10% difference in febrile neutropenia. This was felt to be a meaningful and practical difference to exclude when comparing Pegfilgrastim and Filgrastim".

Reviewer's sensitivity analyses

FDA's sensitivity analysis 1:

In a sensitivity analysis of the primary endpoint, DSN was defined as days of ANC <1× 10^9/L and the results are summarized in Table 8.

Table 8 : DSN with ANC <1x10^9/L in Cycle 1

	EP (N=101)		NEU (N=103)
DSN			
Mean (SD)	1.76 (1.23)		1.84 (1.25)
Difference *		-0.08	
95% CI		(-0.43, 0.26)	
90% CI		(-0.37, 0.21)	

DSN (days), n (%)		
0	22 (21.8)	22 (21.4)
1	17 (16.8)	11 (10.7)
2	33 (32.7)	40 (38.8)
3	21 (20.8)	23 (22.3)
≥ 4	8 (7.9)	7 (6.8)

FDA's sensitivity analysis 2:

This reviewer analyzed the DSN in cycle 1 based on FAS population and the results are summarized in Table 9.

Table 9 : DSN in Cycle 1: FAS population

	EP (N=107)		NEU (N=107)
DSN			
Mean (SD)	1.18 (1.12)		1.20 (1.02)
Difference *		-0.02	
95% CI		(-0.31, 0.27)	
90% CI		(-0.26, 0.22)	
DSN (days), n (%)			
0	37		32
1	23		30
≥ 2	41		41

The mean DSN in cycle 1 was 1.18 and1.20 days for EP and NEU, respectively. The estimated mean DSN difference between EP and NEU was -0.02 days (95% CI:-0.31, 0.27*). The results based on FAS population were also consistent to those of PP population.*

The sample size, based on an equivalence test with margin (-1, 1), was 45 patients with standard deviation of 1 and 90% power at 2-sided α=0.05. The sample size based on equivalence test with margin of (-0.74, 0.74) was 99 patients with standard deviation of 1.11 and 90% power at 2-sided α=0.05.

FDA's sensitivity analysis 3:

In the site 703, 75% patients had commercial filgrastim, the reviewer analyzed the sensitivity analysis for DSN in cycle 1 excluding patients in the site 703 and patients who had commercial filgrastim. The results are summarized in Table 10.

Table 10: DSN in Cycle 1: FAS population Excluding Subjects with Exposure of Commercial Drug and Subjects in Site 703

	EP (N=92)	NEU (N=89)
DSN		
Mean (SD)	1.15 (1.12)	1.13 (1.02)
Difference *	0.01	
95% CI	(-0.30, 0.33)	
90% CI	(-0.25, 0.28)	
DSN (days), n (%)		
0	35	30
1	20	27
≥ 2	37	32

The results were similar to the primary analysis results.

FDA's Sensitivity analysis 4:
There are missing ANC values from Day 10 to Day 15 in the Cycle 1. We did not impute missing DSN days in the control group but imputed 0.1(sensitivity 1), 0.2 (sensitivity 2) days in the missing DSN in the EP2006 group for sensitivity analyses. These sensitivity analyses results are summarized in Table 11.

Table 11: Results for sensitivity analyses for DSN in Cycle 1

	EP	NEU	Differences (95% CI)
Sponsor's results (PP)	N=101	N=103	
Mean (SD)	1.17 (1.11)	1.20 (1.02)	-0.04 (-0.33, 0.26)
Reviewer's results (FAS)	N=107	N=107	
Mean (SD)	1.18 (1.12)	1.20 (1.02)	-0.02 (-0.31, 0.27)
Sensitivity 1 (EP 0.1)	N=107	N=107	
Mean (SD)	1.64 (1.07)	1.20 (1.02)	0.46 (0.18, 0.75)
Sensitivity 2 (EP 0.2)	N=107	N=107	
Mean (SD)	2.10 (1.03)	1.20 (1.02)	0.90 (0.63, 1.18)

The sensitivity analysis results were robust except sensitivity analysis number 2. The missing data mostly occurred after the ANC recovery and we normally assume ANC should be above $.5\times 10^9$ /L.

FDA's sensitivity analysis 5:

The assumption of normality of ANCOVA analysis does not hold, so the reviewer used negative binomial distribution assumption with Genmod based on PP population. The difference (NEU-EP2006) and 95% CI and 90% CI are as follows;

Difference (EP2006-NEU) (95% CI): -0.03 (-0.28, 0.22)

Difference (EP2006-NEU) (90% CI): -0.03 (-0.24, 0.18)

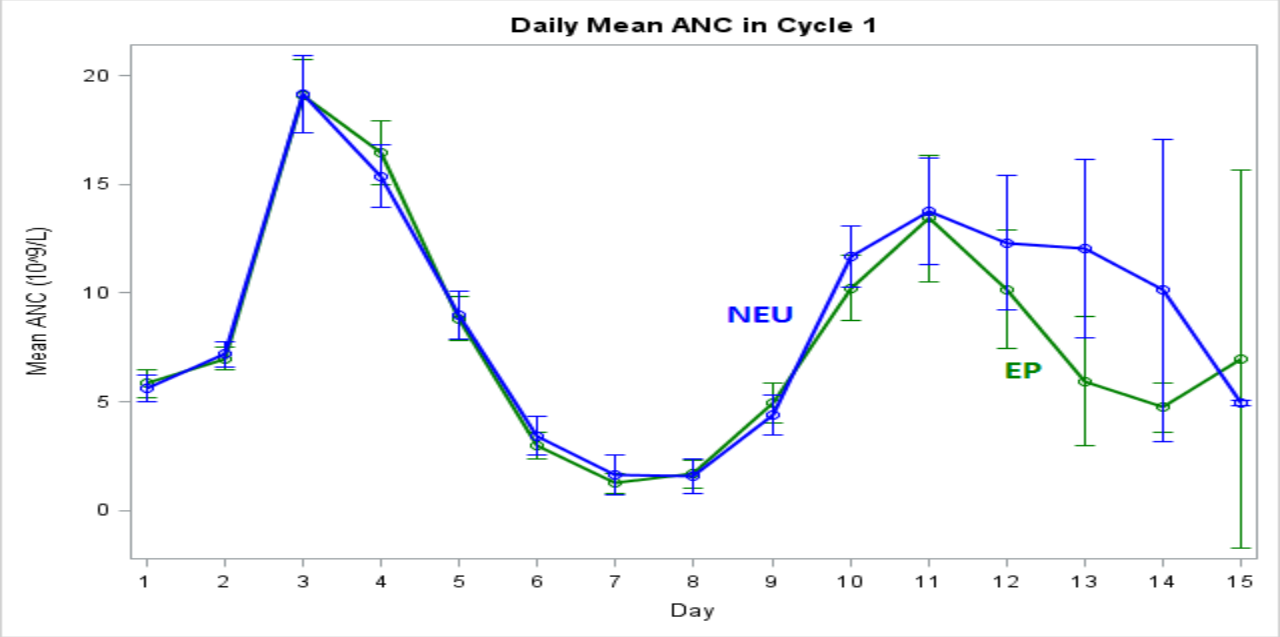
The results were similar to that of ANCOVA. In addition, we analyzed the data using Poisson regression and bootstrap method, the results were similar.

Secondary Endpoints

1. Depth of ANC Nadir

The key secondary endpoint was the depth of ANC nadir and time to recovery of the ANC nadir in cycle 1. The daily mean ANC in Cycle 1 is plotted in Figure 1.

Figure 1 : Daily Mean ANC in Cycle 1: FAS population

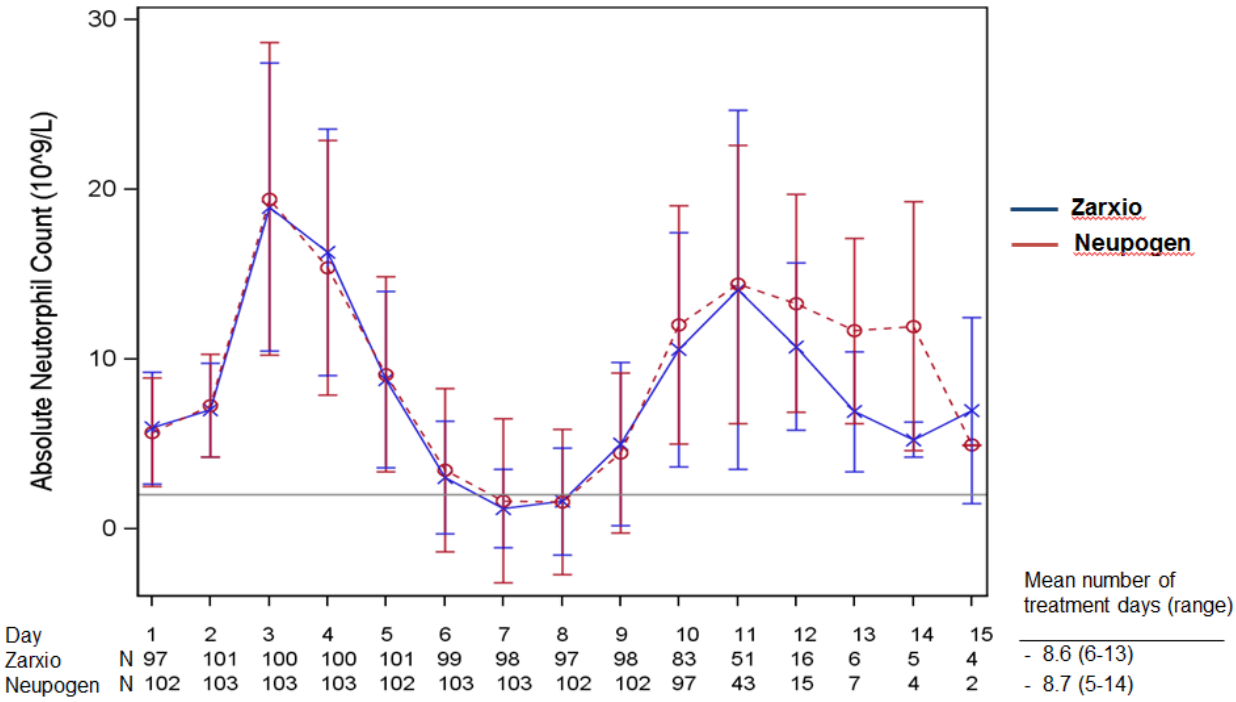


N															
	EP	101	101	101	101	101	101	101	99	98	87	56	25	16	15	14
	NEU	103	103	103	103	103	103	103	102	102	98	45	17	9	6	4

The daily means ANC in Cycle 1 between EP and NEU were similar until Day 10, a time when AND recovery was observed.

During the ODAC meeting of January 7, 2015, the FDA presented a similar graph of ANC profile in the PP set (Figure 1), which was found different from the sponsor's graph (Figure 2) in the same PP set.

Figure 2 : Sponsor's Time course of ANC in Cycle 1 (PP set)



Below, we list the sources of the discrepancies of those two graphs, Figure 1 and Figure 2.

1) If a subject's ANC value is missing (but this subject is still being monitored) at any day, this subject is not counted in the sample size on that day in the sponsor's graph but is counted in FDA's graph.

2) While the sponsor plotted mean+/- standard deviation, the FDA plotted mean with 95% confidence limits for each day.

3) While the sponsor plotted ANC from day 1 to day 21 (with no data from day 16-day 20), the FDA plotted up to day 15.

Both graphs are reasonable, but have different display preferences. Other than these minor differences, the two graphs are the same.

The results for ANC depth and time occurred ANC nadir based on PP population are summarized in Table 12.

Table 12: Sponsor's Depth and Time Occurred ANC nadir in Cycle 1: PP population

	EP (N=101)	NEU (N=103)
Mean (SD)	0.73 (1.14)	0.76 (1.31)
ANC nadir at Day, n (%)		
Mean(SD)	7.27 (1.31)	7.45 (1.45)
1-5	4 (4.0)	2 (1.9)
6	2 (2.0)	2 (1.9)
7	60 (59.4)	58 (56.3)
8	31 (30.7)	37 (35.9)
9	3 (3.0)	1 (1.0)
10-15	1 (1.0)	3 (2.9)

The mean depth of ANC nadir was 0.73 in the EP2006 group and 0.76 in the Neupogen group. The mean time occurred ANC nadir was 7.27 days in the EP2006 group and 7.35 days in the Neupogen group. The mean depth and the mean time of ANC nadir occurred were similar between the two groups in Cycle 1.

2. Time to Recovery of ANC Nadir

The results of time to ANC recovery in Cycle 1 based on PP population are summarized in Table 13.

Table 13: Sponsor's Time to ANC Recovery in Cycle 1: PP population

ANC Recovery Day	EP (N=101)	NEU (N=103)
Mean (SD)	1.79 (0.97)	1.68 (0.81)
Difference*	0.13	
95% CI*	-0.14, 0.36	

*:Difference and 95% CI were estimated using ANCOVA with treatment group and type of chemotherapy and a baseline ANC as a covariate

The mean times to recovery from ANC nadir were 1.79 days and 1.68 days, for EP2006 and Neupogen, respectively. The mean times to ANC nadir recovery were similar. The estimated mean differences in time to recovery ANC nadir was 0.13 days with 95% CI of (-0.14, 0.36). The results were the same with that of sponsor's.

3. Incidence of FN

The FN was defined as oral temperature ≥ 38.3°C while having an ANC < 0.5×10^9/L (both measured on the same day). The sponsor's results for incidence of FN are summarized in Table 14, confirmed by the reviewer.

Table 14: Sponsor's Incidence of FN in Cycle 1: PP population

	EP (N=101) n (%)	NEU (N=103) n (%)
FN		
Number of FN	4 (4.0)	2 (1.9)
Exact 95 % CI	(1.1, 9.8)	(0.0, 5.3)
Days of FN		
1	4 (4.0)	1 (1.0)
2	0	1 (1.0)
Missing	1 (1.0)	

Four patients in the EP group (4%) and 2 patients (1.9%) in the NEU group had FN cases. There were no clinically meaningful differences in the incidence of FN between the two groups.

Reviewer's additional analysis
The results for incidence of FN based on FAS population in Cycle 1 are summarized in Table 15.

Table 15: Incidence of FN in Cycle 1: FAS population

	EP (N=107)	NEU (N=107)
FN	N=106	N=107
Number of FN	5 (4.7)	1 (0.9)
Exact 95% CI	(1.5, 10.6)	(0.0, 5.1)
Days of FN		
1	5	1
2	0	1
Missing	1	

Five patients in the EP group (4.7%) and one patient (0.9%) in the NEU group had FN cases. These results were similar to that of PP population.

4. Number of Days of Fever

Fever was defined as an oral temperature ≥ 38.3°C. One patient's temperature was not available. The sponsor's mean fever days in Cycle 1 are summarized in Table 16, confirmed by the reviewer.

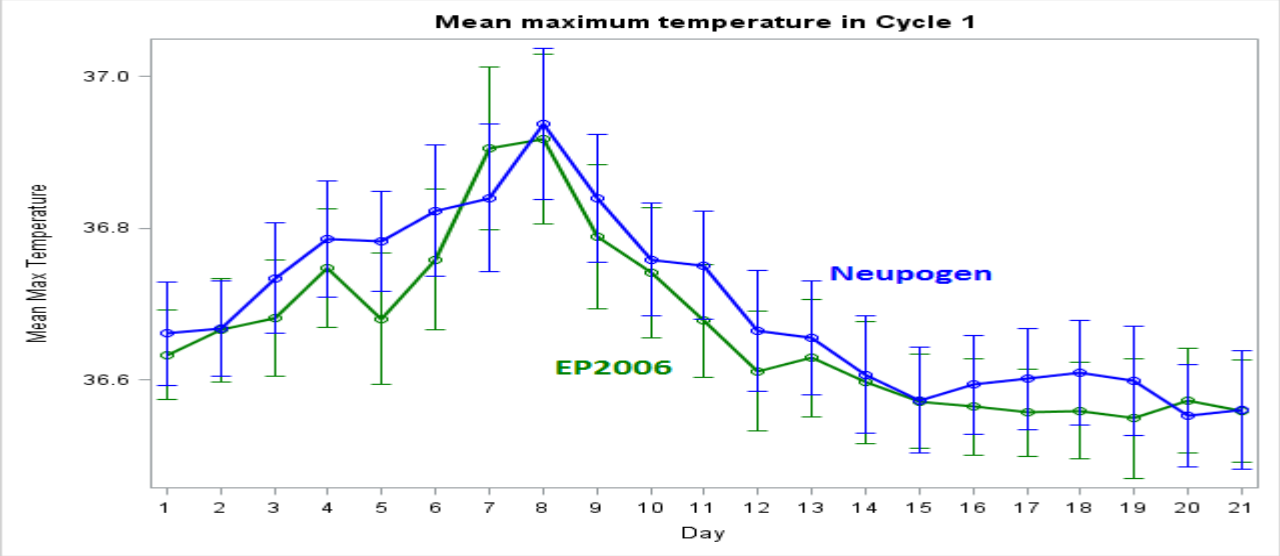
Table 16: Sponsor's Number of days in fever in Cycle 1: PP population

	EP (N=101) n (%)	NEU (N=103) n (%)
Number of days in fever		
Mean (SD)	0.07 (0.29)	0.04 (0.24)

The mean number of days of fever in Cycle 1 was 0.07 days in the EP group and 0.04 days in the NEU group. There were no differences between the two groups in number of fever days.

The mean daily maximum temperatures in Cycle 1 are plotted based on FAS population in Figure 3, confirmed by the reviewer.

Figure 3 : Mean Daily Maximum Temperatures in Cycle 1: FAS Population



N EP	106	106	106	106	105	105	105	105	105	105	101	100	100	99	99	99	99	99	99	96	94
NEU	106	107	107	107	107	107	107	107	107	107	104	100	99	99	98	98	98	98	97	97	93

The daily mean maximum temperatures seem little higher in Neupogen group compared to EP2006 group, but the daily mean maximum temperatures were between 36.5-36.9 ° C.

3.3 Evaluation of Safety

For a detailed summary of the evaluation of safety refer to the review by Dr. Donna Przepiorka.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The reviewer conducted subgroup analyses for age groups (<65 years versus ≥ 65 years) and geographic region (Russia vs. Ukraine vs. Other) of the primary endpoint of DSN using difference and 90% CI are summarized in Table 17.

Table 17: Subgroup Analyses of DSN: Age and Region (90% CI): PP population

	EP (N=101)		NEU (N=103)		
	N	Mean (SD)	N	Mean (SD)	Difference (90% CI)
Age					
< 65	92	1.17 (1.12)	92	1.16 (1.02)	0.01 (-0.25, 0.28)
≥ 65	9	1.11 (0.93)	11	1.55 (1.04)	-0.44 (-1.24, 0.36)
Geographic Region					
Russia	74	1.18 (1.05)	82	1.26 (1.03)	-0.08 (-0.36, 0.19)
Ukraine	17	1.24 (1.43)	16	0.75 (0.86)	0.44 (-0.27, 1.15)
Other	10	1.00 (0.94)	5	1.80 (1.10)	-1.02 (-1.86, -0.18)

The mean DSN was 0.01 days shorter in the NEU group than the EP group in age < 65 group. The mean DSN were 1.11 days in the EP group and 1.55 days in the NEU group in age ≥ 65 group. However, there were only 20 patients in age ≥ 65 group. The results are generally consistent with the whole population except results obtained in subgroups with small sample size.

4.2 Other Special/Subgroup Populations

The reviewer also performed subgroup analyses of the primary endpoint by disease characteristics and summarized results in Table 18.

Table 18: Subgroup Analyses for DSN: Baseline Disease Characteristics (90% CI): PP population

	EP (N=101)		NEU (N=103)		
	N	Mean (SD)	N	Mean (SD)	Difference (90% CI)
Therapy					
Adjuvant	58	1.24 (1.20)	58	1.17 (1.11)	0.08 (-0.28, 0.43)
Neo-adjuvant	43	1.07 (0.96)	45	1.24 (0.91)	-0.17 (-0.50, 0.16)
Stage					
I	6	1.67 (1.63)	8	1.13 (1.13)	0.92 (-0.66, 2.50)
II	56	1.11 (1.11)	50	1.14 (1.05)	-0.02 (-0.37, 0.33)
III	39	1.18 (1.02)	45	1.29 (0.99)	-0.09 (-0.46, 0.28)
Surgery					
Yes	79	1.23 (1.15)	77	1.19 (1.04)	0.04 (-0.26, 0.33)
No	22	0.95 (0.90)	26	1.23 (0.99)	-0.26 (-0.71, 0.19)
Radio Therapy					
Yes	9	1.44 (1.24)	8	1.13 (0.64)	0.17 (-0.76, 1.09)
No	92	1.14 (1.10)	95	1.21 (1.05)	-0.07 (-0.33, 0.19)

ECOG					
0	79	1.18 (1.16)	81	1.20 (1.02)	-0.01 (-0.29, 0.28)
1	22	1.09 (0.92)	22	1.23 (1.07)	-0.08 (-0.61, 0.44)

The mean DSN difference between the two groups and the 90% CI has the upper and lower bound less than 1 day except subgroups with small sample size:

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and collective Evidence

The primary endpoint of mean DSN in Cycle 1 was 1.17 days and 1.20 days for EP and NEU, respectively. The 95% CI (-0.33, 0.26) and 90% CI (-0.28, 0.21) were within (-1, 1).

For secondary endpoints, the mean depth of ANC nadir was 0.73 in the EP2006 group and 0.76 in the Neupogen group. The mean time to ANC nadir recovery was 1.79 days in the EP2006 group and 1.68 days in the Neupogen group. The difference of mean time to recovery of ANC nadir was 0.13 days with 95% CI of (-0.14, 0.36). The mean depth and the mean time to ANC nadir recovery were similar between the two groups in Cycle 1.

For the incidence of FN, 4 patients in the EP group (4%) and 2 patients (1.9%) in the NEU group had FN cases.
The mean number of days of fever in Cycle 1 was 0.07 days in the EP group and 0.04 days in the NEU group. There were no clinically meaningful differences in the incidence of FN between the two groups.

5.2 Conclusions and Recommendations

The analyses of both the primary endpoint (DSN) as well as secondary endpoints in Cycle 1 of study EP06-302 support the conclusion that there was no clinically meaningful difference with respect to efficacy between EP2006 group and US-licensed Neupogen group in Cycle 1.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

KYUNG Y LEE
01/30/2015

THOMAS E GWISE
01/30/2015

LEI NIE
01/30/2015

RAJESHWARI SRIDHARA
01/30/2015

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	125553
REFERENCE PRODUCT	NEUPOGEN (filgrastim)
ESTABLISHED NAME	EP2006
STRENGTH	300 mcg/0.5mL and 480 mcg/0.8 mL
DOSAGE FORM:	Injection or infusion
INDICATION:	Patients With Severe Chronic Neutropenia; Cancer Patients Receiving Myelosuppressive Chemotherapy; Cancer Patients Receiving Bone Marrow Transplant; Patients With Severe Chronic Neutropenia; Patients undergoing progenitor cell therapy
SPONSOR:	Sandoz Inc.
REVIEW FINISHED:	January 30, 2015
NAME OF STATISTICAL REVIEWER:	Xiaoyu (Cassie) Dong, Ph.D.
OBP REVIEWER	Maria Gutierrez Lugo

_____

Reviewer: Xiaoyu Dong, Mathematical Statistician, CDER/OTS/OB/DB VI

Concur: _____

Meiyu Shen, Ph.D., Acting TL, CDER/OTS/OB/DB VI
Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DB VI

Distribution: BLA 125553

CDER/OTS/OB/DB VI/ Yi Tsong
CDER/OTS/OB/DB VI/ Meiyu Shen
CDER/OTS/OB/ Lillian Patrician
CDER/OBP/ Maria Gutierrez Lugo
CDER/TBBT

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY AND RECOMMENDATION 5
III. SPONSOR'S STATISTICAL EQUIVALENCE TESTING 8
III.1. SPONSOR'S STATISTICAL EQUIVALENCE TESTING ON BIOACTIVITY 8
III.2. SPONSOR'S STATISTICAL EQUIVALENCE TESTING ON CONTENT 10
IV. THE AGENCYU'S STATISTICALEQUIVALENCE TESTING 13
IV.1. OVERVIEW 13
IV.2. THE AGENCY'S STATISTICAL EQUIVALENCE TESTING ON BIOACTIVITY 14
IV.3. THE AGENCY'S STATISTICAL EQUIVALENCE TESTING ON CONTENT 16
V. CONCLUSION 18

List of Tables

Table 1 – Summarized Results of Statistical Equivalence Testing for Bioactivity (%) and Content (%) based on the Agency's Analyses..... 5
Table 2: Summary of Sandoz's Responses to the Agency's Information Request Letters on Bioactivity and Content..... 8
Table 3 – Sandoz's Equivalence Testing Results on Bioactivity..... 9
Table 4 – Sandoz's Equivalence Testing Results on Content..... 12
Table 5 - Descriptive Statistics for Bioactivity (%) based on the Agency's Analyses..... 15
Table 6– The Agency's Equivalence Testing Results for Bioactivity (%)..... 15
Table 7 - Descriptive Statistics for Content (%) Data from the Agency's Analysis;..... 17
Table 8– Equivalence Testing Results for Content (%) from the Agency's Analyses..... 17

List of Figures

Figure 1- Sandoz's Plot of the Content Data 11
Figure 2– The Agency's Plot on Bioactivity (%) Data 14
Figure 3– The Agency's Plot on Content (%) Data 16

I. EXECUTIVE SUMMARY AND RECOMMENDATION

For Tier 1 critical quality attributes such as Content (%) and Bioactivity (%), the Agency conducted statistical equivalence testing to assess their similarity to support a demonstration that EP2006 is highly similar to US-licensed Neupogen, and to support the 3-way bridge to justify the relevance of the data generated with EU-approved Neupogen to a demonstration of biosimilarity. Based on the data provided by Sandoz, the testing results are summarized below and also in Table 1:

- For Bioactivity (%), statistical equivalence in mean values is established among EP2006 drug product, US-licensed Neupogen, and EU-approved Neupogen. A total of 15 lots of EP2006 drug product (9 clinical lots and 6 commercial lots), 15 lots of US-licensed Neupogen (10 lots in Pre-filled Syringe (PFS) presentation and 5 lots in Vial presentation), and 34 lots of EU-approved Neupogen were analyzed.
- For Content (%), statistical equivalence in mean values is established among EP2006 drug product, US-licensed Neupogen, and EU-approved Neupogen. A total of 20 lots of EP2006 drug product (13 Clinical lots and 7 commercial lots), 12 lots of US-licensed Neupogen, and 49 lots of EU-approved Neupogen were analyzed.

The statistical equivalency testing results support the demonstration that that EP2006 drug product is highly similar to US-licensed Neupogen. The results of similarity between US-licensed and EU-approved Neupogen provide relevant information for bridging.

Table 1 – Summarized Results of Statistical Equivalence Testing for Bioactivity (%) and Content (%) based on the Agency's Analyses

Bioactivity (%)	**# of Lots**	**Mean Difference**	**90% Confidence Interval**	**Equivalence Margin**	**Statistical Equivalence?[a]**
EP2006 vs. US	15 vs.15	-5.47	(-8.67, -2.27)	(-9.32, 9.32)	Yes
EP2006 vs. EU	15 vs. 34	-2.47	(-5.47, 0.54)	(-10.07, 10.07)	Yes
EU vs. US	34 vs. 15	-3.12	(-6.34, 0.10)	(-9.32, 9.32)	Yes
Content (%)	**# of Lots**	**Mean Difference**	**90% Confidence Interval**	**Equivalence Margin**	**Statistical Equivalence?[a]**
EP2006vs. US	20 vs.12	-0.86	(-1.87, 0.15)	(-2.26, 2.26)	Yes
EP2006 vs. EU	20 vs. 49	-1.91	(-2.98, -0.85)	(-3.23, 3.23)	Yes
EU vs. US	49 vs. 12	1.18	(0.27, 2.09)	(-2.26, 2.26)	Yes

[a] Statistical equivalence in mean values is established if the obtained confidence interval of the mean difference is completely within the equivalence margin.

Sandoz' analytical similarity assessment of Bioactivity and Content is briefly described in Section II, while a description of Sandoz' statistical analyses for these two attributes is summarized in Section III. Because the Agency did not fully agree with Sandoz's statistical approaches and data selection for the statistical analyses (see Section II for the Agency's comments), the Agency conducted its own statistical

equivalence testing to assess the similarity for Bioactivity and Content. The summary of the Agency's results is shown in Section IV.

II. INTRODUCTION

On May 08, 2014, Sandoz submitted BLA 125553 seeking for licensure under section 351(k) of the Public Health Service Act. The analytical data of Bioactivity and Content were reported in 2.3.P entitled "Quality Overall Summary – Drug Product" and 3.2.R entitled "Overview biosimilarity data". However, due to insufficient number of lots, Bioactivity data didn't establish statistical equivalence for the comparison to US-licensed Neupogen. The content data of EP2006 commercial drug product appeared to be lower than the reference product of US-licensed Neupogen. In addition, Sandoz did not perform the tiered approach for analytical similarity data. Thus, the Agency sent out several Information Request (IR) Letters to Sandoz to address those CMC issues.

FDA IR letter dated 03-Oct-2014:

You provided data to support analytical similarity between EP2006, US-licensed Neupogen® and an EU-approved filgrastim product. The data are derived from two evaluations. Evaluation 1 compared 6 batches of EP2006 drug product (DP), 4 batches of US-licensed Neupogen® and 2 batches of the EU-approved filgrastim product. Evaluation 2 compared 6 batches of EP2006 drug substance (DS) and 5 batches of EP2006 DP with 4 batches of the EU-approved filgrastim.

We are reviewing your analytical similarity data (i.e., evaluation 1 and 2) to evaluate whether you have demonstrated that EP2006 is "highly similar" to US-licensed Neupogen® and whether you have provided adequate analytical data to scientifically justify the relevance of other comparative data obtained using EU-approved filgrastim to support a demonstration that EP2006 is biosimilar to US-licensed Neupogen®.

In your critical quality attribute (CQA) assessment, you identified potency (specific activity in U/mg) and protein concentration (protein content in mcg/ml), both with a criticality score of 140, as two of the most critical quality attributes. However, based on the data you submitted, the min-max ranges for potency and protein content of EP2006 appear to be lower than those of US-licensed Neupogen®. One possible explanation for these observations may be the limited number of batches of US-licensed Neupogen® (4 batches) included in your similarity exercise.

As you have additional US-licensed Neupogen® reference lots that were identified during inspection, you should include these lots of US-licensed Neupogen® in your similarity exercise. We further recommend that you conduct a statistical analysis of the analytical similarity data, including data from these additional lots, to provide more robust support for your efforts to demonstrate that EP2006 is "highly similar" to the reference product with respect to quality attributes, including but not limited to potency and protein content.

Please note, the agency also sent the recommendations on the tiered approach to Sandoz in this IR.

FDA IR letter dated 31-Oct-2014:

Under section 351(k)(2)(A)(i)(IV) of the PHS Act, an applicant must demonstrate that the "strength" of the proposed biosimilar product is the same as that of the reference product. Accordingly, we expect your proposed biosimilar product to have both the same total content of GCSF (in mass or units of activity in a container closure) and the same concentration of GCSF (in mass or units of activity per unit volume) as US-licensed Neupogen (see Q+A I.12 in draft guidance on Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009).You stated that your equivalence testing results for "content" (i.e., concentration as expressed in milligrams per milliliter) of EP2006 in pre-filled syringes (PFS) against US-licensed Neupogen in PFS, and between EP2006 in PFS against the US-licensed Neupogen in both PFS and vials are "inconclusive". In addition, FDA analysis of content of drug product batches manufactured at Lek Pharmaceuticals d.d., Slovenia (LEK), IDT Biologika GmbH,Germany (IDT) and GP Grenzach Produktions GmbH, Germany (GPG) (section 3.2.P.5.4) indicates that the EP2006 drug product validation batches manufactured at GPG (95%-98% from declared content of 0.6 mg/ml) have lower content compared to EP2006 drug product batches manufactured at IDT (97%-100% from declared content of 0.6 mg/ml) and LEK (102%-103% from declared content of 0.6 mg/ml).The lower content of EP2006 drug product manufactured at GPG appears to be a manufacturing issue. Address the lower "content" (i.e., concentration as expressed in milligrams per milliliter) of EP2006 drug product manufactured at GPG and submit data to demonstrate that EP2006 drug product manufactured at GPG, the proposed site for your intended commercial product, has the same "strength" as US-licensed Neupogen.

FDA IR letter dated 07-Nov-2014:

1. In your response, dated October 16, 2014, to our request for information, dated October 02, 2014, you provided additional data to support analytical similarity of EP2006 and the reference product, US-licensed Neupogen® and to establish an analytical bridge between EP2006, the reference product and EU-approved Neupogen®.Provide the following additional information for the bioactivity potency data present in Table 2-8.

a. *Clarify how many replicates were obtained to calculate the reportable result for each lot.*
b. *For bioactivity data in the table, the five data points of US-licensed Neupogen® of Vial are consistently lower than those data from US-licensed Neupogen® of PFS. Provide an explanation as to why such difference is observed between the vial and PFS presentations of US-licensed Neupogen®. In addition, please submit all available potency data for US-licensed Neupogen® for both Vial and PFS presentations.*

2. Specify the expiry date for the tested US-licensed Neupogen® and EU-approved Neupogen® as well as the manufacturing date for the EP2006 in your Table 2-2 for Content and Table 2-8 for Bioactivity. Also specify the testing date for each lot value listed in Table 2-2 and Table 2-8.

Sandoz's responses to each of the above IR letters are summarized in Table 2 below. The Agency carefully evaluated similarity studies for bioactivity and content provided in the original BLA submission along with all updated information found in Sandoz's responses to IR letters. Our comments regarding Sandoz's statistical equivalence testing (Tier 1 approach) is provided in Section III, and our independent statistical equivalence testing analysis are present in Section IV.

Table 2: Summary of Sandoz's Responses to the Agency's Information Request Letters on Bioactivity and Content

Date	Document Title	Summary of Updated Information
16-Oct-2014	Response to request for information #1, dated 03-Oct-2014	Sandoz performed FDA's tiered approach for their quality attributes, including statistical equivalence testing for Content, Bioactivity, Target Binding, Higher Order Structures, Subvisiable particles, and Product Related Variants. Additional data on US-licensed and EU-approved Neupogen were also provided.
14-Nov-2014	Response to request for information, dated 07-Nov-2014	Sandoz provided the number of replicates for Bioactivity and Content. In addition, Sandoz submitted manufacturing data, testing date and expiry for each batch of Bioactivity and Content. A few typos in Content data were corrected.
17-Nov-2014	Response to request for information, dated 07-Nov-2014	Sandoz performed statistical equivalence testing between PFS and Vial lots of US-Neupogen for Bioactivity. Sandoz also provide additional information to evaluate similarity between Ep2006 and US-Neupogen for Bioactivity.
25-Nov-2014	Response to request for information, dated 31-Oct-2014	Sandoz submitted additional data for content including seven independent lots of EP2006 commercial drug product and two additional lots of EP2006 clinical drug product. Sandoz also conducted statistical equivalence testing on the content data.

III. SPONSOR'S STATISTICAL EQUIVALENCE TESTING

Sandoz conducted statistical equivalence testing for both Bioactivity and Content based on the Agency's recommendation sent out on October 03, 2014. Their analyses results on those two critical quality attributes followed by the Agency's comments are provided in Section III.1 and Section III.2.

III.1. SPONSOR'S STATISTICAL EQUIVALENCE TESTING ON BIOACTIVITY

Bioactivity is measured as the percentage (%) of the measured potency relative to the potency of Sandoz's in-house reference standard. Sandoz submitted the equivalence testing results for Bioactivity (%) on their Responses to IR dated October 16, 2014. In their responses, the following comparisons were conducted by the equivalence testing for Bioactivity"

- EP2006 DP FPS (15 lots) vs. US-Neupogen PFS (10 lots);
- EP2006 DP FPS (15 lots) vs. US-Neupogen PFS + Vial (15 lots in total);

- EP2006 DP FPS (15 lots) vs. EU-Neupogen PFS (34 lots);
- US-Neupogen PFS (10 lots) vs. EU-Neupogen PFS (34 lots);

Their testing results for above comparisons are listed in Table 3 below.

Table 3 – Sandoz's Equivalence Testing Results on Bioactivity

Table 2-9 Equivalence testing results for bioactivity EP2006 DP PFS against bioactivity Neupogen® US PFS

Obs	Variable	Class	Method	Variances	Mean	LowerBound	LowerCLMean	UpperCLMean	UpperBound	Assessment
1	meas	Diff (1-2)	Pooled	Equal	-8.1667	-8.836	-11.4153	-4.9180	8.8364	Inconclusive
2	meas	Diff (1-2)	Satterthwaite	Unequal	-8.1667	-8.836	-11.7780	-4.5553	8.8364	Inconclusive

δ= 1.55* σR, having a power of ~ 80% to state equivalence, if a true difference of σR/2 is present, with 10 batches of Neupogen® US and 15 batches of EP2006 DP, more details see [Module 1.11.1 Quality Information Amendment 5].

Table 2-10 Equivalence testing results for bioactivity EP2006 DP PFS against bioactivity Neupogen® US PFS and vials

Obs	Variable	Class	Method	Variances	Mean	LowerBound	LowerCLMean	UpperCLMean	UpperBound	Assessment
1	meas	Diff (1-2)	Pooled	Equal	-5.4667	-8.836	-8.6680	-2.2653	8.8364	Equivalent
2	meas	Diff (1-2)	Satterthwaite	Unequal	-5.4667	-8.836	-8.6907	-2.2426	8.8364	Equivalent

δ= 1.45* σR, having a power of ~ 80% to state equivalence, if a true difference of σR/2 is present, with 15 batches of Neupogen® US (PFS and vials) and 15 batches of EP2006 DP, more details see [Module 1.11.1 Quality Information Amendment 6].

Table 2-11 Equivalence testing results for bioactivity EP2006 DP PFS against bioactivity Neupogen® EU PFS

Obs	Variable	Class	Method	Variances	Mean	LowerBound	LowerCLMean	UpperCLMean	UpperBound	Assessment
1	potency	Diff (1-2)	Pooled	Equal	2.3431	-8.037	-0.5604	5.2466	8.0375	Equivalent
2	potency	Diff (1-2)	Satterthwaite	Unequal	2.3431	-8.037	-0.0904	4.7767	8.0375	Equivalent

δ= 1.3* σR, having a power of ~ 80% to state equivalence, if a true difference of σR/2 is present, with 34 batches of Neupogen® EU (PFS) and 15 batches of EP2006 DP, more details see [Module 1.11.1 Quality Information Amendment 7].

Table 2-12 Equivalence testing results for bioactivity Neupogen® US PFS against bioactivity Neupogen® EU PFS

Obs	Variable	Class	Method	Variances	Mean	LowerBound	LowerCLMean	UpperCLMean	UpperBound	Assessment
1	potency	Diff (1-2)	Pooled	Equal	-5.8235	-8.656	-9.5039	-2.1431	8.6557	Inconclusive
2	potency	Diff (1-2)	Satterthwaite	Unequal	-5.8235	-8.656	-9.4780	-2.1691	8.6557	Inconclusive

δ= 1.4* σR, having a power of ~ 80% to state equivalence, if a true difference of σR/2 is present, with 34 batches of Neupogen® EU (PFS = reference) and 10 batches of Neupogen® US PFS, more details see [Module 1.11.1 Quality Information Amendment 8].

The Agency's Comments:

The Agency does not fully agree with Sandoz's statistical analyses with respect to Bioactivity for the following reasons.

- *Firstly, a true mean difference of σ/2 used by Sandoz in their equivalence margin determination is too large. Instead, the Agency proposed to use σ/8. Sandoz's calculation to support such a large mean difference is incorrect. In their calculation, low probability of observing the sample mean difference within (-σ/8, σ/8) was used to indicate that σ/8is too stringent. However, this probability is not meaningful to use because it is different from the passing rate of the equivalence testing. Instead, we should compute the probability of the observed mean difference covered the equivalence margin of (-1.5σ, 1.5σ) with an assumed true mean difference. As stated in the Agency's recommendation, this probability is about 85% with 10 biosimilar lots and 10 reference lots when the true mean difference is σ/8. As we can see that, 85% is much higher than Sandoz's claimed value of 22%.*
- *Secondly, Sandoz used different equivalence margins for different sample sizes. The Agency recommended equivalence margins of (-1.5σ, 1.5σ) for all sample sizes. When the sample size is small, the confidence level can be lower than 90% but agreement on this should be reached in advance with FDA scientists.*
- *Thirdly, for the comparison between US-licensed and EU-approved Neupogen, Sandoz used EU-approved Neupogen as the comparator. In contrast, US-Neupogen should be treated as the comparator. Thus, the equivalence margin should be determined by the variability of US-licensed Neupogen, not from EU-approved Neupogen.*

III.2. SPONSOR'S STATISTICAL EQUIVALENCE TESTING ON CONTENT

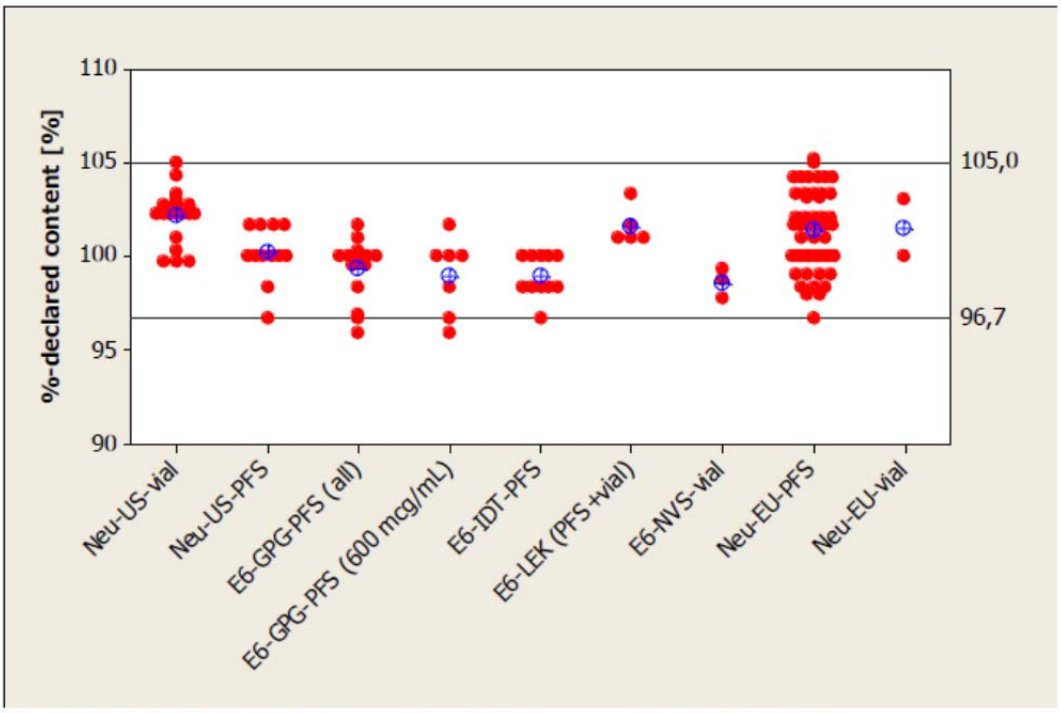
Content is measured as the percentage (%) of the actual protein concentration (mcg/mL) relative to the target value of 600 mcg/mL. Sandoz submitted their ultimate equivalence testing results for Content (%) on their Responses to IR dated November 25, 2014. In their responses, the following comparisons were conducted by the equivalence testing for Content:

- US-Neupogen FPS (12 lots) vs. US-Neupogen Vial (19 lots);
- US-Neupogen FPS (12 lots) vs. vs. EP2006 GPG 600 (7 lots);
- US-Neupogen FPS (12 lots) vs. vs. EP2006 GPG (15 lots);
- US-Neupogen FPS (12 lots) vs. vs. EP2006 All (34 lots);
- US-Neupogen FPF + Vial (31 lots) vs. EU-Neupogen (51 lots);
- US-Neupogen FPF + Vial (31 lots) vs. vs. EP2006 GPG 600 (7 lots);
- US-Neupogen FPF + Vial (31 lots) vs. vs. EP2006 GPG (15 lots);
- US-Neupogen FPF + Vial (31 lots) vs. vs. EP2006 All (34 lots);

Sandoz's graphic display of the data and testing results for above comparisons are shown in Figure 1 and Table 4 below.

Figure 1- Sandoz's Plot of the Content Data

Figure 2-1 **Overview of the content values for Neupogen and EP2006 showing a maximum value of 105.0 % and a minimum value of 96.7 % for Neupogen US**



Min(96.7) and Max (105.0) values are based on Neupogen US (PFS and vials)

Table 4 – Sandoz's Equivalence Testing Results on Content

Table 2-5 Equivalency testing results for %-declared content

Evaluation	Reference	# lots ref.	Comparator	# lots comp.	Diff. of mean	EAC lower limit	EAC upper limit	CI lower limit	CI upper limit	CI lower limit	CI upper limit	Result
					%-declared content					Relative to EAC = -1 to +1		
1	Neupogen US PFS	12	Neupogen US vial	19	1,97	96,0	104,4	101,2	103,0	-0,694	-0,249	Equivalent
2			EP2006 GPG 600	7	1,21	95,4	105,0	97,5	100,3	-0,046	0,548	Equivalent
3			EP2006 GPG	15	0,85	95,9	104,5	98,3	100,3	-0,046	0,443	Equivalent
4			EP2006	34	0,70	96,2	104,2	98,6	100,4	-0,049	0,401	Equivalent
5	Neupogen US	31	Neupogen EU	51	0,02	98,0	104,8	100,6	102,2	-0,215	0,225	Equivalent
6			EP2006 GPG 600	7	2,41	96,9	105,9	97,6	100,2	0,253	0,821	Equivalent
7			EP2006 GPG	15	2,05	97,6	105,2	98,4	100,2	0,298	0,769	Equivalent
8			EP2006	34	1,90	97,9	104,9	98,8	100,2	0,342	0,735	Equivalent

The Agency's Comments:

The Agency does not fully agree with Sandoz's statistical analyses with respect to Content for the following reasons.

- *First, Sandoz used the standard deviation of EU-Neupogen data for the equivalence margin calculation because EU-Neupogen has the largest number of lots. However, equivalence margin should be established by the standard deviation of the comparator depending on the specific comparison being conduct. That is, for the comparison between the proposed biosimilar and US-Neupogen, the equivalence margin should be established by the variability of US-Neupogen, not the EU-Neupogen.*
- *Secondly, Sandoz determined the multiplier of the equivalence margin by bootstrapping approach to ensure a high power of demonstrating the equivalence between US-Neupogen and EU-Neupogen. In general, equivalence margin determined by bootstrapping would heavily depend on the observed data. In other words, the multiplier cannot be pre-specified. Thus, the type I error rate or other statistical properties using Sandoz's approach are difficult to evaluate. In addition, the equivalence margin determined by comparing the US- and EU-Neupogen is likely to be too tight. The Agency recommends a fixed multiplier of 1.5.*
- *In addition, the Agency does not fully agree with the data selection for the equivalence testing. In the Agency's analysis for Content, only PFS presentation with the same strength of US-licensed Neupogen were included in the equivalence testing. For EP2006 drug product, only those lots manufactured at the current commercial site were analyzed.*

Thus, the Agency conducted independent statistical equivalence testing with regard to Bioactivity. The Agency's testing results are provided in Section IV.

IV. THE AGENCYU'S STATISTICALEQUIVALENCE TESTING

IV.1. OVERVIEW

Statistical equivalence testing in terms of the mean difference is formulated as the following hypothesis.

$$H_0: \mu_T - \mu_C \leq -1.5\sigma_c \text{ or } \mu_T - \mu_C \geq 1.5\sigma_c$$

versus

$$H_a: -1.5\sigma_c < \mu_T - \mu_C \leq 1.5\sigma_c$$

where μ_T and μ_C are the mean responses of the test and the comparator products and σ_c is the standard deviation of the comparator which can be either US-licensed or EU-approved Neupogen, depending on the specific comparison being conducted. The specific equivalence margin was set as constant times the standard deviation of the reference product attributes to ensure an adequate power, when a small but sufficient number of lots are available for testing. After examining a range of possible values for the constant, FDA scientists and statisticians agreed on a value of 1.5. Defining the margin as $\pm 1.5\sigma_c$ assures 85% power of accepting the equivalence hypothesis, if the true mean difference is 1/8 times the standard deviation (σ_C), with 10 biosimilar lots and 10 comparator lots used for testing and assuming a Type I error rate of 5% (i.e., use of a 90% confidence interval for the equivalence testing procedure). When the number of lots is smaller than 10, the test size may be relaxed somewhat, but agreement on this should be reached in advance with FDA scientists. Statistical equivalence in mean values is established if the obtained 90% confidence interval of the mean difference falls inside the equivalence margin of $(-1.5\sigma_c, 1.5\sigma_c)$.

When the sample variances of the test and comparator are similar, we can use the pooled sample variance to compute the confidence interval as described in formula (1)

$$\text{Confidence Interval} = \bar{X}_T - \bar{X}_C \pm t_{1-\alpha,\, NT+NC-2} \times S_{pool}\sqrt{\frac{1}{N_B} + \frac{1}{N_R}} \qquad (1)$$

where $S_{pool}^2 = \frac{(N_T-1)S_T^2+(N_C-1)S_C^2}{N_T+N_C-2}$ is the pooled sample variance with equal variance assumption.

When the sample variances of the test and comparator are different, we can use Satterthwaite's approximation to obtain the confidence interval as described in formula (2)

$$\text{Confidence Interval} = \bar{X}_T - \bar{X}_C \pm t_{1-\alpha,\,\nu} \times \sqrt{\frac{S_T^2}{N_T} + \frac{S_C^2}{N_C}} \quad (2)$$

where the approximate degree of freedom ν is computed by the Satterthwaite approximation. The Agency's equivalence testing results based on the above approach with regard to Bioactivity and Content are summarized below.

IV.2. THE AGENCY'S STATISTICAL EQUIVALENCE TESTING ON BIOACTIVITY

Bioactivity is measured as the percentage (%) of the measured potency relative to the potency of Sandoz's in-house reference standard. The data were provided in Sandoz's response to IR submitted on November 07, 2014, including a total of 15 lots of EP2006 drug product (nine Clinical lots and six commercial lots), 15 lots of US-licensed Neupogen (10 lots in Pre-filled Syringe (PFS) presentation and 5 lots in Vial presentation), and 34 lots of EU-approved Neupogen. In the data set, US-licensed Neupogen lots in PFS presentation have expiry date ranging from 2011 to 2015; US-licensed Neupogen lots in Vial presentation have expiry date ranging from 2011 to 2014; EU-approved Neupogen lots in PFS have expiry date ranging from 2004 to 2014.

The plot and descriptive statistics from the Agency's analysis were provided in Figure 2 and Table 5, respectively.

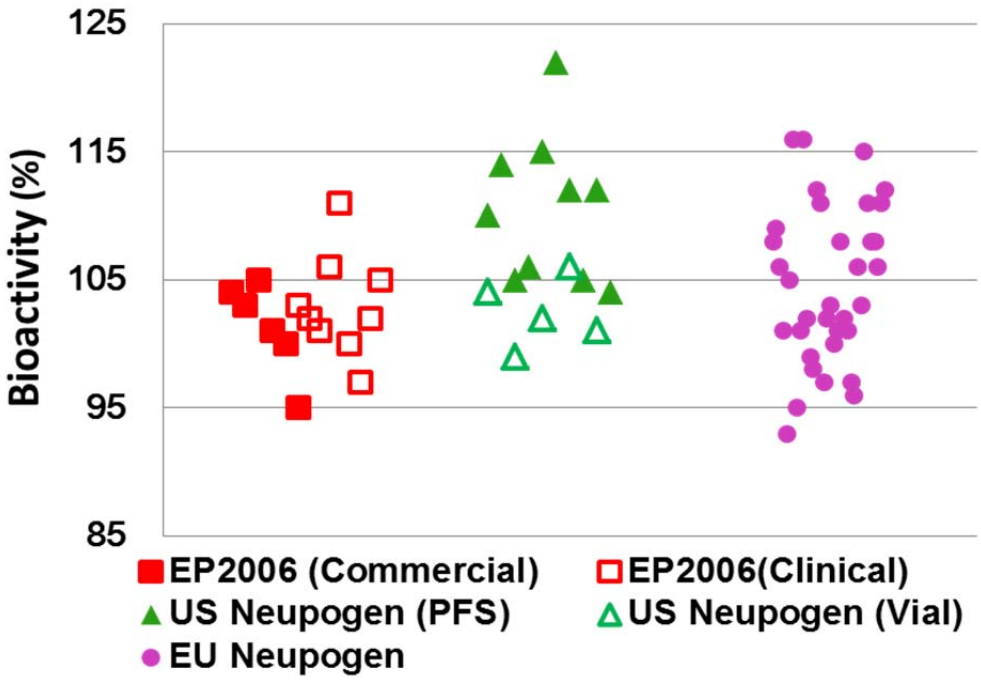


Figure 2– The Agency's Plot on Bioactivity (%) Data

In Figure 2, each data point is the reported value from each lot. The solid squares in red are the bioactivity (%) data from six lots of EP2006 commercial drug product; the open squares in red are the bioactivity (%) data from nine lots of EP2006 clinical drug product; the green solid and open triangles are content data from US-Neupogen lots in PFS and Vial presentations, respectively. The purple solid dots are the bioactivity (%) data from 34 lots of EU-approved Neupogen.

Table 5 - Descriptive Statistics for Bioactivity (%) based on the Agency's Analyses

	# of Batches	Min	Max	Mean	Standard Deviation	CV(%)[a]
US-Neupogen (PFS + Vial)	15	99	122	107.8	6.21	5.76%
EU-Neupogen (PFS)	34	93	116	104.7	6.18	5.91%
EP2006 Clinical (PFS)	9	97	111	103.0	4.00	3.88%
EP2006 Commercial (PFS)	6	95	105	101.3	3.61	3.57%
EP2006 Commercial + Clinical (PFS)	15	95	111	102.3	3.81	3.72%

[a] CV(%) is coefficient of variability and is computed as the percentage of the sample standard deviation relative to the sample mean value;

For Bioactivity (%), the Agency conducted statistical equivalence testing among EP2006 including both clinical and commercial drug products, US-licensed Neupogen including both PFS and Vial presentations, and EU-approved PFS Neupogen. Data of EP2006 commercial and clinical drug products were combined for the evaluation due to two reasons. First, Sandoz demonstrated that the clinical and commercial manufacturing processes are comparable. Secondly, the number of lots for each product is very limited. Suggested by the Agency's CMC reviewer, data of US-licensed Neupogen in both PFS and Vial presentations were also combined for statistical equivalence testing. Those two presentations have the same indications and routes of administrations. Please note, no formal statistical testing was conducted to assess the pooliability of PFS and Vial data for US-licensed Neupogen due to insufficient number of lots.

The summarized statistical equivalence testing results are presented in Table 6 below. As can been seen, the 90% confidence interval of each comparison is entirely within the equivalence margin. Thus, statistical equivalence in mean values is established among EP2006 drug product (Clinical + Commercial), US-licensed Neupogen (PFS + Vial) and EU-approved Neupogen.

Table 6– The Agency's Equivalence Testing Results for Bioactivity (%)

	# of Lots	Mean Difference	90% Confidence Interval	Equivalence Margin	Statistical Equivalence?
EP2006[a] vs. US[b]	15 vs.15	-5.47	(-8.67, -2.27)	(-9.32, 9.32)	Yes
EP2006[a] vs. EU	15 vs. 34	-2.47	(-5.47, 0.54)	(-10.07, 10.07)	Yes
EU vs. US[b]	34 vs. 15	-3.12	(-6.34, 0.10)	(-9.32, 9.32)	Yes

[a]: EP2006 data includes nine Clinical lots and six Commercial lots;

[b]: US-Neupogen includes 10 US-licensed Neupogen in PFS and five US-licensed Neupogen in Vial.

IV.3. THE AGENCY'S STATISTICAL EQUIVALENCE TESTING ON CONTENT

Content is measured as the percentage (%) of the actual protein concentration (mcg/mL) relative to the target value of 600 mcg/mL. A total of 20 lots of EP2006 drug product (13 Clinical lots and seven commercial lots), 12 lots of US-licensed Neupogen, and 49 lots of EU-approved Neupogen were analyzed. Those data are provided by Sandoz in their response to IR submitted on November 25, 2014, including seven independent commercial lots and two additional clinical lots of EP2006 drug product. Please note, only the same strength as US-licensed Neupogen (300 mcg /0.5 mL and 480 mcg/0.8 mL) is considered in the Agency's evaluation.

The plot and descriptive statistics of the content data are provided in Figure 3 and

Table 7, respectively.

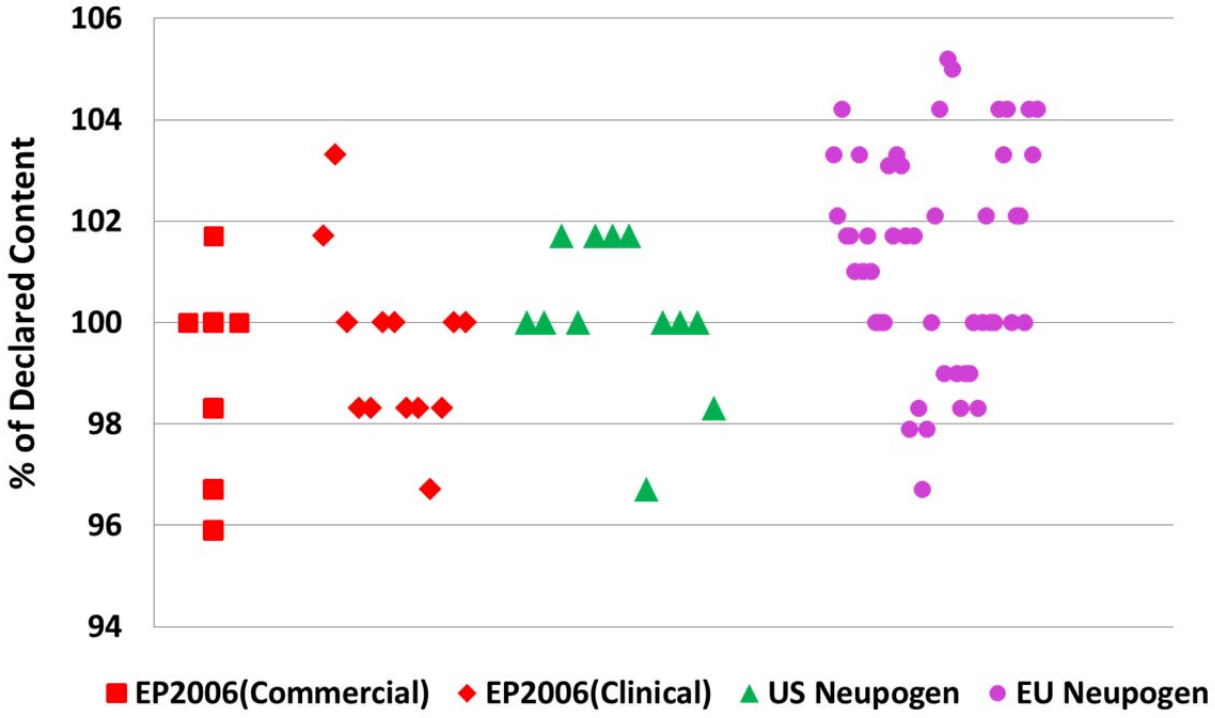


Figure 3– The Agency's Plot on Content (%) Data

Figure 3 Plots the Content data used for Statistical Equivalence Testing. In the plot, each data point is the reported value from each lot. The solid squares in red are the content (%) data from seven lots of EP2006 commercial drug product; the diamonds in red are the content (%) data from 13 lots of EP2006 clinical drug product; The green triangles and purple solid dots are the content (%) data from 20 lots of US-licensed Neupogen and 49 lots of EU-approved Neupogen, respectively.

Table 7 - Descriptive Statistics for Content (%) Data from the Agency's Analysis;

Product	# of Batches	Min	Max	Mean	Standard Deviation	CV (%)[a]
US-licensed Neupogen	12	96.7	101.7	100.15	1.51	1.51%
EU-approved Neupogen	49	96.7	105.2	101.33	2.14	2.11%
EP2006 Clinical	13	96.7	103.3	99.48	1.72	1.73%
EP2006 Commercial	7	95.9	101.7	98.94	2.07	2.09%

[a] CV(%) is coefficient of variability and is computed as the percentage of the sample standard deviation relative to the sample mean value;

The Agency conducted statistical equivalence testing in mean values for above content data. Again, statistical equivalence in mean values is established if the obtained 90% confidence interval of the mean difference falls inside the equivalence margin of $(-1.5\sigma_c, 1.5\sigma_c)$, where σ_c is the standard deviation of the comparator which can be either US-licensed Neupogen or EU-approved Neupogen, depending on the specific comparison being conducted. The summarized results are presented in Table 8 below.

Table 8– Equivalence Testing Results for Content (%) from the Agency's Analyses

	# of Lots	Mean Difference	90% Confidence Interval	Equivalence Margin	Statistical Equivalence?
EP2006[a] vs. US	20 vs.12	-0.86	(-1.87, 0.15)	(-2.26, 2.26)	Yes
EP2006[a] vs. EU	20 vs. 49	-1.91	(-2.98, -0.85)	(-3.23, 3.23)	Yes
EU vs. US	49 vs. 12	1.18	(0.27, 2.09)	(-2.26, 2.26)	Yes
EP2006 Clinical vs. US	13 vs. 12	-0.67	(-1.78, 0.43)	(-2.26, 2.26)	Yes
EP2006 Clinical vs. EU	13 vs. 49	-1.83	(-3.13, -0.53)	(-3.26, 3.26)	Yes
EP2006 Commercial vs. US	7 vs. 12	-0.83	(-2.33, 0.67)[b]	(-2.28, 2.28)	No
EP2006 Commercial vs. EU	7 vs. 49	-1.96	(-3.71, -0.20)[b]	(-3.23, 3.23)	No

[a]: EP2006 data includes 12 Clinical lots and seven Commercial lots;

[b]: a confidence level of 85.2% is used due to a small number of lots from EP2006 commercial process.

As can been seen, statistical equivalence is established among EP2006 drug product (Clinical + Commercial), US-licensed and EU-approved Neupogen. For the comparisons among EP2006 clinical drug product, US-licensed and EU-approved Neupogen, the same conclusion can be made. The seven lots of EP2006 commercial drug product does not pass the equivalence testing because the lower bounds of the confidence interval were outside the lower equivalence margins. Specifically, for the comparison between EP2006 commercial drug product and US-licensed Neupogen, the lower bound of the 90% confidence interval of the mean different is -2.33%, which is lower than the equivalence margin of -2.28% by 0.05%; for the comparison between EP2006 commercial drug product and EU-approved Neupogen, the lower bound of the 90% confidence interval of the mean different is -3.71%, which is lower than the equivalence margin of -3.23% by 0.48%. CMC reviewer of the Agency suggested that those differences (0.05% and 0.48%) are not considered as biologically meaningful, especially with such a limited number of lots of EP2006 commercial drug product. Thus, EP2006 drug product with combined data from both commercial and clinical lots is considered as appropriate for the analytical

similarity assessment for Content. With that, we can conclude that statistical equivalence in mean values of Content (%) is established among EP2006 drug product (Commercial + Clinical), US-licensed Neupogen, and EU-approved Neupogen.

V. CONCLUSION

Based on the Agency's assessment, the results from statistical equivalency testing of Bioactivity (%) and Content (%) support the demonstration that EP2006 drug product is highly similar to US-licensed Neupogen. The results of similarity between US-licensed and EU-approved Neupogen provide relevant information for bridging.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

XIAOYU DONG
01/30/2015

MEIYU SHEN
01/30/2015

YI TSONG
01/30/2015

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.:	125553
REFERENCE PRODUCT	NEUPOGEN® (filgrastim)
ESTABLISHED NAME	filgrastim
STRENGTH	300 mcg/0.5mL and 480 mcg/0.8 mL
DOSAGE FORM:	Injection or infusion
INDICATION:	Patients With Severe Chronic Neutropenia; Cancer Patients Receiving Myelosuppressive Chemotherapy; Cancer Patients Receiving Bone Marrow Transplant; Patients With Severe Chronic Neutropenia;Patients undergoing progenitor cell therapy
SPONSOR:	Sandoz Inc.
REVIEW FINISHED:	September 10, 2014
NAME OF STATISTICAL REVIEWER:	Xiaoyu (Cassie) Dong, Ph.D.
OBP REVIEWER	Maria Gutierrez Lugo

_____

Reviewer: Xiaoyu Dong, Mathematical Statistician, CDER/OTS/OB/DB VI

Concur: _____

Meiyu Shen, Ph.D., Acting TL, CDER/OTS/OB/DB VI
Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DB VI

Distribution: BLA 125553

CDER/OTS/OB/DB VI/ Yi Tsong
CDER/OTS/OB/DB VI/ Meiyu Shen
CDER/OTS/OB/ Lillian Patrician
CDER/OBP/ Maria Gutierrez Lugo

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY AND RECOMMENDATION *4*
II. SIMILARITY ASSESSMENT ON BIOACTIVITY (U/MG) *5*
II.1. Data *5*
II.2. Descrivptive statistics *7*
II.3. Priliminary comparisons *7*
II.4. Statistical equivalence testing *9*
III. SIMILARITY ASSESSMENT ON CONTENT (MG/ML) *10*
III.1. Data *10*
III.2. Descrivptive statistics *12*
III.3. Priliminary comparisons *13*
III.4. Statistical equivalence testing *14*
IV. CONCLUSION *15*

List of Tables

Table 1: Bioactivity Data of Sandoz's Similarity Evaluation 1: Biosimilarity of EP2006 with Reference Product Neupogen 6
Table 2: Bioactivity Data of Sandoz's Similarity Evaluation 2: Similarity of EP2006 with EU Comparator Product (Note: LEK is the biosimilar products used for EU approval a few years ago) 6
Table 3: Descriptive Statistics for Bioactivity (×E+08 U/mg) based on FDA CMC Statistics Reviewer's Analysis; 7
Table 4: Preliminary Comparisons for Bioactivity (×E+08 U/mg) Data Based on FDA CMC Statistics Reviewer's Analysis; Δμ is the mean difference between Test and Reference (Ref.); EAC = Equivalence Acceptance Criteria; σ_{ref} is the sample standard deviation of the reference. 8
Table 5: Statistical Equivalence Testing Results for Bioactivity (×E+08 U/mg) Based on FDA CMC Statistics Reviewer's Analysis; Note: Mean Difference = Test – Reference. EAC = Equivalence Acceptance Criteria; Width = $t_{1-\alpha,\ NB+NR-2} \times Spool1/NB + 1/NR$ 9
Table 6: Sandoz's Content Data of Similarity Evaluation 1: Biosimilarity of EP2006 with Reference Product Neupogen 11
Table 7: Sandoz's Content Data of Similarity Evaluation 2: Biosimilarity of EP2006 with Reference Product Neupogen 11
Table 8: Sandoz's Content Data of Similarity Evaluation 3: US and EU Neupogen 11
Table 9: Sandoz's Content Data from Batch Analysis for EP2006 Manufactured by IDT 12
Table 10: Descriptive Statistics for Content (mg / mL) Data Based on FDA CMC Statistics Reviewer's Analysis 12
Table 11: Preliminary Comparisons for Content (mg/mL) Data Based on FDA CMC Statistics Reviewer's Analysis; EAC = Equivalence Acceptance Criteria; Δμ is the mean difference between Test and Reference (Ref.); σ_{ref} is the sample standard deviation of the reference.. 13
Table 12: Statistical Equivalence Testing Results for Content (U/mg) Based on FDA CMC Statistics Reviewer's Analysis; Note: Mean Difference = Test – Reference; EAC = Equivalence Acceptance Criteria; Width = $t_{1-\alpha,\ NB+NR-2} \times Spool1/NB + 1/NR$ 14

I. EXECUTIVE SUMMARY AND RECOMMENDATION

Based on the statistical analyses of the available data, there is insufficient evidence to show equivalence between the proposed biosimilar product and the US-licensed product for Bioactivity and Content. To have more reliable assessment of similarity, we recommend that the sponsor should include more batches from the proposed biosimilar product and US reference product.

The summarized results of statistical equivalence testing are provided in Table A as below. Please see Section II and Section III for detailed analyses.

Table A – Summarized Results of Statistical Equivalence Testing for Bioactivity (×E+08 U/mg) and Content (mg/mL) based on FDA CMC Statistics Reviewer's Analysis; Mean Difference = Test – Reference; EAC = Equivalence Acceptance Criteria; Width = $t_{1-\alpha,\ NB+NR-2} \times S_{pool}\sqrt{1/N_B + 1/N_R}$

Bioactivity (×E+08 U/mg)										
Test	Ref	N_B	N_R	α[a]	Conf. Level [a]	Mean Diff	Width	Conf. Interval	EAC ($1.5\sigma_{ref}$)	Equivalence? [b]
GPG	US	6	4	9.3%	81.4%	-0.0833	0.0556	(-0.1389, -0.0278)	0.1225	No
LEK		5		9.3%	81.4%	-0.0800	0.0622	(-0.1422, -0.0178)		No
GPG+LEK		11		8.4%	83.2%	-0.0818	0.0451	(-0.1269, -0.0367)		No
GPG	All EU	6	6	9.3%	81.4%	-0.0833	0.0436	(-0.1270, -0.0397)	0.0949	No
LEK		5		9.3%	81.4%	-0.0800	0.0484	(-0.1284, -0.0316)		No
GPG+LEK		11		8.4%	83.2%	-0.0818	0.0362	(-0.1180, -0.0456)		No
All EU	US	6	4	9.3%	81.4%	0.0000	0.0660	(-0.0660, 0.0660)	0.1225	Yes
Content (mg/mL)										
Test	**Ref**	**N_B**	**N_R**	**α [a]**	**Conf. Level [a]**	**Mean Diff**	**Width**	**Conf. Interval**	**$1.5\sigma_{ref}$**	**Equivalence? [b]**
GPG	US	6	5	9.3%	81.4%	-0.0227	0.0094	(-0.0320, -0.0133)	0.0171	No
IDT		8		8.5%	83.0%	-0.0135	0.0074	(-0.0209, -0.0061)	0.0171	No
IDT	GPG	8	6	8.9%	82.2%	0.0092	0.0066	(0.0025, 0.0158)	0.0155	No
GPG	EU	6	6[c]	9.3%	81.4%	-0.0183	0.0111	(-0.0294, -0.0073)	0.0240	No
EU	US	12	5	8.8%	82.4%	-0.0018	0.0111	(-0.0129, 0.0092)	0.0171	Yes

[a] Confidence levels are computed based on the actual sample sizes. If the number of lots from either test or reference is less than six, the confidence level is computed based on six lots.

[b] Pass the statistical equivalence testing if the obtained Confidence Interval is completely covered by (-EAC, EAC).

[c] 6 samples were randomly selected from a total of 12 EU samples for TOST. The id numbers of the sample are 2, 3, 5, 6, 7 and 8 with Mean (EU TOST) = 0.6017, SD(EU TOST) = 0.0160; the other 6 samples are used to estimate σ_{ref} for EAC with Mean(EU STD) = 0.6067 and SD(EU STD) = 0.0163.

For Bioactivity (×E+08 U/mg), results of statistical equivalence testing from Table A are summarized below.

- Due to limited number of batches for biosimilar, US and EU products, the obtained confidence levels, (1-2α)100% are lower than 90% within a fixed EAC of $1.5\sigma_{ref}$;
- The sample means of Bioactivity of the biosimilar product manufactured at GPG and LEK are consistently lower than the US reference and EU comparator about 8%;
- Biosimilar product GPG fails to show statistical equivalence in Bioactivity to the US reference data because the confidence interval in mean difference, (-0.1389, -0.0278) is not covered by the acceptance criteria of (-0.1225, 0.1225). Likewise, GPG is also fails to show statistical equivalence to the EU comparator.
- EU comparator shows statistical equivalence to the US reference product.

For Content (mg/mL), results of statistical equivalence testing from Table A are summarized below.

- Due to limited number of batches for biosimilar, US and EU products, the obtained confidence levels, (1-2α)100% are lower than 90% within a fixed EAC of $1.5\sigma_{ref}$;
- Biosimilar product GPG fails to show statistical equivalence to the US reference because the confidence interval in mean difference, (-0.0320, -0.0133) is not covered by the acceptance interval of (-0.0171, 0.0171). Such a failure is mostly due to the sample mean difference of -0.0227 mg/mL. In addition, the upper confidence limit of GPG vs US is less than zero, indicating the biosimilar product may have a lower mean in Content than the US reference product. Similar observations are found for the comparison of IDT vs. US;
- Biosimilar product GPG also fails to show statistical equivalence to the EU comparator because the confidence interval of the mean difference, (-0.0294, -0.0073) is not covered by the acceptance interval of (-0.0240, 0.0240). Such a failure is mostly due to the mean difference of -0.0183 mg /mL. In addition, the upper confidence limit of GPG vs. EU is less than zero, indicating the biosimilar product may have a lower mean in Content than the EU comparator.
- EU comparator shows statistical equivalence to the US reference product.

II. SIMILARITY ASSESSMENT ON BIOACTIVITY (U/MG)

II.1. DATA

Per OBP reviewer's request, CMC Statistics reviewer conducted statistical equivalence testing for Bioactivity (U/mL) based on the available data. The data set consists of two parts, Similarity Evaluation 1 and Similarity Evaluation 2. In Similarity Evaluation 1, there are six biosimilar

drug product lots (GPG), four US reference product lots and two EU comparator product lots as shown in Table 1. In Similarity Evaluation 2, there are six biosimilar drug substance (DS) lots, five biosimilar lots (LEK), and four EU comparator lots as shown in Table 2. In addition, there are multiple strengths (filing volume) of the product, including 300 mcg/0.5 mL, 480 mcg/0.8 mL, and 480 mcg/0.5 mL. Based on the discussion with the OBP reviewer, we combined all the three strengths for the analysis considering that Bioactivity should be independent on the strength.

Table 1: Bioactivity Data of Sandoz's Similarity Evaluation 1: Biosimilarity of EP2006 with Reference Product Neupogen

Batch	Manufacturer	PPM	Strength (mcg/mL)	Bioactivity (U/mg)
V200001	GPG	PFS	480/0.8	1.0E+08
V201002	GPG	PFS	480/0.8	1.0E+08
V201102	GPG	PFS	480/0.8	1.1E+08
V201001	GPG	PFS	300/0.5	1.0E+08
V201101	GPG	PFS	300/0.5	1.0E+08
V200201	GPG	PFS	300/0.5	1.0E+08
1014928	US	PFS	480/0.8	1.1E+08
1025269	US	PFS	480/0.8	1.2E+08
1020649	US	PFS	300/0.5	1.0E+08
1021957	US	PFS	300/0.5	1.1E+08
1026606	EU	PFS	480/0.8	1.0E+08
1025051	EU	PFS	300/0.5	1.2E+08

Table 2: Bioactivity Data of Sandoz's Similarity Evaluation 2: Similarity of EP2006 with EU Comparator Product (Note: LEK is the biosimilar products used for EU approval a few years ago)

Batch	Manufacturer	PPM	Strength (mcg/mL)	Bioactivity (U/mg)
48200402	Sandoz	DS	-	1.1E+08
48200403	Sandoz	DS	-	1.0E+08
48200404	Sandoz	DS	-	1.1E+08
48200405	Sandoz	DS	-	1.0E+08
48200406	Sandoz	DS	-	1.1E+08
48200407	Sandoz	DS	-	1.1E+08
A03941609F	LEK	PFS	480/0.5	1.0E+08
A00657409G	LEK	PFS	300/0.5	1.0E+08
A00675111G	LEK	PFS	480/0.5	1.0E+08
A00675011G	LEK	PFS	300/0.5	1.0E+08
A00675211G	LEK	PFS	480/0.5	1.1E+08
N0875AA	EU	PFS	480/0.5	1.1E+08
N1144AE	EU	PFS	300/0.5	1.1E+08
N1113AG	EU	PFS	300/0.5	1.1E+08
N1114AJ	EU	PFS	480/0.5	1.1E+08

II.2. DESCRIVPTIVE STATISTICS

The descriptive statistics of Biosimilar (GPG, LEK and Sandoz DS), US reference, and EU comparator batches are summarized in Table 3 below. In particular, number of batches, minimum, maximum, mean, standard deviation (Std. Dev.) and observed coefficient of variation ($\hat{C}_v$) are computed. Here, $\hat{C}_v$ is ratio of sample standard deviation to the sample mean.

Table 3: Descriptive Statistics for Bioactivity (×E+08 U/mg) based on FDA CMC Statistics Reviewer's Analysis;

Product	Evaluation	Manufacturer	# of Batches	Min	Max	Mean	Std. Dev (σ)	$\hat{C}_v$(%)
Biosimilar	1	GPG	6	1.0	1.1	1.017	0.0408	4.01%
	2	LEK	5	1.0	1.1	1.020	0.0447	4.38%
	2	Sandoz (DS)	6	1.0	1.1	1.067	0.0516	4.84%
	GPG + LEK		11	1.0	1.1	1.018	0.0405	3.98%
US-Licensed	1	US	4	1.0	1.2	1.100	0.0816	7.42%
EU-Approved	1	EU	2	1.0	1.2	1.100	0.1414	12.86%
	2	EU	4	1.1	1.1	1.100	0	0%
	All EU		6	1.0	1.2	1.100	0.0632	5.75%

There are several observations from Table 3.

- The number of lots for Biosimilar product, US reference, and EU comparator are limited, all less than 10.
- The maximum Bioactivity from Biosimilar PFS and DS lots is 1.1 E+08 U/mg, which is lower than the maximum value of both US and EU batches in Similarity Evaluation 1.
- The mean values of Bioactivity of Biosimilar batches are 1.017 E+08 U/mg for GPG, 1.020 E+08 U/mg for LEK, and 1.067 E+08 U/mg for Sandoz DS, which are lower than the mean values of US reference and EU comparator.
- The largest standard deviation, 0.0816 E+08 U/mg comes from four US reference batches, corresponding to 7.42% of the sample mean value (i.e. $\hat{C}_v$ = 7.42%). Please note, we didn't consider the Std. Dev. of two EU batches here because the number of batches is too small.

II.3. PRILIMINARY COMPARISONS

Before we conducted formal statistical equivalence testing, we performed some preliminary comparisons among Biosimilar, US reference and EU comparator. The results are summarized in Table 4 below.

Table 4: Preliminary Comparisons for Bioactivity (×E+08 U/mg) Data Based on FDA CMC Statistics Reviewer's Analysis; Δμ is the mean difference between Test and Reference (Ref.); EAC = Equivalence Acceptance Criteria; σ_{ref} is the sample standard deviation of the reference.

Test	Ref.	N_B	N_R	μ_{Ref}	$\Delta\mu_{Test-Ref}$	$\Delta\mu$ / μ_{Ref} (%)	σ_{ref}	EAC ($1.5\sigma_{ref}$)	EAC/μ_{Ref} (%)
GPG	US	6	4	1.100	-0.0833	-7.57%	0.0816	0.1225	11.14%
LEK		5			-0.0800	-7.27%			
GPG+LEK		11			-0.0818	-7.44%			
GPG	All EU	6	6	1.100	-0.0833	-7.57%	0.0632	0.0949	8.63%
LEK		5			-0.0800	-7.27%			
GPG+LEK		11			-0.0818	-7.44%			
All EU	US	6	4	1.100	0.0000	0	0.0816	0.1225	11.14%

Based on Table 4, preliminary comparisons between Biosimilar and US reference batches are summarized below:

- Sample mean for Bioactivity of the GPG batches is lower than that of the US batches. In particular, the observed mean difference in Bioactivity between GPG and US reference is -0.0833 U/mg. That is, the sample mean for Bioactivity of GPG is about 7.57% lower than the US reference data.
- The statistical equivalence acceptance criteria (EAC) for the mean difference is 1.5 σ_{ref} = 0.1225 U/mg, corresponding to 11.14% of the sample mean of the US batches.
- Similar results are observed in LEK vs. US and (GPG+LEK) vs. US;

Based on Table 4, preliminary comparisons between Biosimilar and EU comparator batches are summarized below:

- Sample mean for Bioactivity of the GPG batches is also lower than that of the EU comparator batches. In particular, the observed mean difference in Bioactivity between GPG and EU batches is -0.0833 U/mg. That is, the mean Bioactivity of GPG is about 7.57% lower than that of the EU comparator batches.
- The statistical EAC for the mean difference is $1.5\sigma_{ref}$ = 0.0949 U/mg, corresponding to 8.63% of the sample mean from the EU batches.
- Similar results are observed in LEK vs. EU and (GPG+LEK) vs. EU;

Based on Table 4, preliminary comparisons between US Reference and EU comparator batches are summarized below:

- There is no observed mean difference between US reference and EU comparator batches;
- The statistical EAC for the mean difference is 1.5 σ_{ref} = 0.1225 U/mg, corresponding to 11.14% of the sample mean from the US batches.

II.4. STATISTICAL EQUIVALENCE TESTING

Statistical equivalence testing in terms of the mean difference is formulated as the following hypothesis.

$$H_0: \mu_{test} - \mu_{ref} \leq \text{-EAC or } \mu_{test} - \mu_{ref} \geq \text{EAC}$$

versus

$$H_a: \text{- EAC} < \mu_{test} - \mu_{ref} \leq \text{EAC}$$

where μ_{test} and μ_{ref} are the mean responses of the test and reference products, respectively. Statistical equivalence is concluded if the (1-2α)100% two-sided confidence interval (CI) of the mean difference is completely within (-EAC, EAC), where EAC equals to $1.5\sigma_{ref}$ and CI is used to estimate the true mean difference. The value of α is adjusted by the sample size with a lower confidence level (1-2α < 90%) if the numbers of biosimilar and reference batches are both less than 10.

$$\begin{aligned}\text{Confidence Interval} &= \text{Mean Diff.} \pm \text{Width} \\ &= \text{Mean Diff.} \pm t_{1-\alpha,\ NB+NR-2} \times S_{pool}\sqrt{\frac{1}{N_B}+\frac{1}{N_R}} \end{aligned} \quad \textbf{(1)}$$

where $S_{pool}^2 = \frac{(N_R-1)S_R^2+(N_T-1)S_T^2}{N_R+N_T-2}$ is the pooled sample variance with equal variance assumption.

By applying the above approach, the obtained (1-2α)100% confidence interval of the mean difference for each comparison is provided in Table 5 and compared against (-EAC, EAC).

Table 5: Statistical Equivalence Testing Results for Bioactivity (×E+08 U/mg) Based on FDA CMC Statistics Reviewer's Analysis; Note: Mean Difference = Test – Reference. EAC = Equivalence Acceptance Criteria; Width = $t_{1-\alpha,\ NB+NR-2} \times S_{pool}\sqrt{1/N_B + 1/N_R}$

Test	Ref	N_B	N_R	α[a]	Conf. Level [a]	Mean Diff	Width	Conf. Interval	EAC ($1.5\sigma_{ref}$)	Equivalence? [b]
GPG	US	6	4	9.3%	81.4%	-0.0833	0.0556	(-0.1389, -0.0278)	0.1225	No
LEK		5		9.3%	81.4%	-0.0800	0.0622	(-0.1422, -0.0178)		No
GPG+LEK		11		8.4%	83.2%	-0.0818	0.0451	(-0.1269, -0.0367)		No
GPG	All EU	6	6	9.3%	81.4%	-0.0833	0.0436	(-0.1270, -0.0397)	0.0949	No
LEK		5		9.3%	81.4%	-0.0800	0.0484	(-0.1284, -0.0316)		No
GPG+LEK		11		8.4%	83.2%	-0.0818	0.0362	(-0.1180, -0.0456)		No
All EU	US	6	4	9.3%	81.4%	0.0000	0.0660	(-0.0660, 0.0660)	0.1225	Yes

[a] Confidence levels are computed based on the actual sample sizes. If the number of lots from either test or reference is less than six, the confidence level is computed based on six lots.

[b] Pass the statistical equivalence testing if the obtained Confidence Interval is completely covered by (-EAC, EAC)

There are several observations from Table 5.

- Due to limited number of batches for biosimilar, US and EU products, the confidence level $(1-2\alpha)100\%$ is lower than 90% within a fixed EAC of $1.5\sigma_{ref}$;
- Biosimilar product (GPG, LEK, GPG+LEK) fails to show statistical equivalence to the US reference product because their confidence intervals are not covered by the acceptance criteria of (-0.1225, 0.1225). Such a failure is mostly due to the relative large mean difference which is about 8% of the reference mean. In addition, the upper confidence limits are all less than zero, indicating the biosimilar product may have lower mean Bioactivity than the US reference product.
- Biosimilar product (GPG, LEK, GPG+LEK) fails to show statistical equivalence to the EU comparator because their confidence intervals are not covered by the acceptance criteria of (-0.0949, 0. 0949). Such a failure is mostly due to the relative large mean difference which is also about 8% of the mean value from the EU comparator. In addition, the upper confidence limits are all less than zero, indicating the biosimilar product may have lower mean Bioactivity than the EU comparator.
- EU comparator shows statistical equivalence to the US reference product.

III. SIMILARITY ASSESSMENT ON CONTENT (MG/ML)

III.1. DATA

In this section, similarity assessments in RP-HPLC Content (mg/mL) among Biosimilar product, US-licensed and EU-approved Neupogen are conducted using statistical equivalence testing. The data set consists of four parts, Similarity Evaluations 1~3 and batch analysis (LEK) as shown in Tables 6 ~ 9. Please note, only strengths of 300 mcg /0.5 mL and 480 mcg/0.8 mL are considered in our evaluation.

Table 6: Sandoz's Content Data of Similarity Evaluation 1: Biosimilarity of EP2006 with Reference Product Neupogen

Sample type	Batch Number	Strength [mcg/mL]	PPM	Manufacturer (EP2006)	RP-HPLC Sum [%]	RP-HPLC LSI [%]	RP-HPLC content [mg/mL]
Similarity evaluation 1: Biosimilarity of EP2006 with reference product Neupogen							
EP2006	V201001	300/0.5	PFS	GPG	1.5	0.5	0.57
EP2006	V201101	300/0.5	PFS	GPG	1.1	0.6	0.58
EP2006	V200201	300/0.5	PFS	GPG	1.1	0.5	0.59
EP2006	V200001	480/0.8	PFS	GPG	1.1	0.5	0.60
EP2006	V201002	480/0.8	PFS	GPG	1.4	0.5	0.58
EP2006	V201102	480/0.8	PFS	GPG	1.1	0.6	0.58

Table 7: Sandoz's Content Data of Similarity Evaluation 2: Biosimilarity of EP2006 with Reference Product Neupogen

Similarity evaluation 2: Similarity exercise of EP2006 with EU comparator product							
EP2006	A03941609F	960/1.0	vial	LEK	1.8	1.0	0.97
EP2006	A00657409G	300/0.5	PFS	LEK	1.9	0.7	0.61
EP2006	A00675111G	480/0.5	PFS	LEK	2.0	0.7	0.97
EP2006	A00675011G	300/0.5	PFS	LEK	1.5	0.6	0.62
EP2006	A00675211G	480/0.5	PFS	LEK	1.7	0.7	0.97

Table 8: Sandoz's Content Data of Similarity Evaluation 3: US and EU Neupogen

	Product	Batch	Strength (mcg/mL)	PPM	RP-HPLC Content (mg/mL)
1	Neupogen US	1027491	300/0.5	PFS	0.61
2	Neupogen US	1009162	300/0.5	PFS	0.6
3	Neupogen US	1020649	300/0.5	PFS	0.62
4	Neupogen US	P104490	300/0.5	PFS	0.59
5	Neupogen US	1023892	480/0.8	PFS	0.61
1	Neupogen EU	1029837	300/0.5	PFS	0.58
2	Neupogen EU	1029442	300/0.5	PFS	0.58
3	Neupogen EU	1029228	300/0.5	PFS	0.59
4	Neupogen EU	N0527AA	300/0.5	PFS	0.63
5	Neupogen EU	N1144AE	300/0.5	PFS	0.6
6	Neupogen EU	N1113AG	300/0.5	PFS	0.6
7	Neupogen EU	N1204AJ	300/0.5	PFS	0.62
8	Neupogen EU	1026519	300/0.5	PFS	0.62
9	Neupogen EU	1027142	300/0.5	PFS	0.61
10	Neupogen EU	1026494	300/0.5	PFS	0.61
11	Neupogen EU	1023368	300/0.5	PFS	0.61
12	Neupogen EU	1024772	300/0.5	PFS	0.60

Table 9: Sandoz's Content Data from Batch Analysis for EP2006 Manufactured by IDT

	Product	Batch	Manufacturer	Strength (mcg/mL)	PPM	RP-HPLC Content (mg/mL)
1	EP2006	#030806	IDT	300/0.5	PFS	0.60
2	EP2006	#040906	IDT	300/0.5	PFS	0.59
3	EP2006	#050906	IDT	300/0.5	PFS	0.59
4	EP2006	#150210	IDT	300/0.5	PFS	0.59
5	EP2006	#140210	IDT	300/0.5	PFS	0.59
6	EP2006	#050409	IDT	300/0.5	PFS	0.60
7	EP2006	#220810	IDT	300/0.5	PFS	0.58
8	EP2006	#111007	IDT	300/0.5	PFS	0.60

III.2. DESCRIVPTIVE STATISTICS

The descriptive statistics of EP2006 (GPG, LEK and IDT), US reference, and EU comparator are summarized in Table 10 below. In particular, number of batches, minimum, maximum, mean, standard deviation (Std. Dev.) and observed coefficient of variation ($\hat{C}_v$) are computed. Again, $\hat{C}_v$ is ratio of sample standard deviation to the sample mean.

Table 10: Descriptive Statistics for Content (mg / mL) Data Based on FDA CMC Statistics Reviewer's Analysis

Manufacturer	# of Batches	Min	Max	Percent < 0.60 mg/mL	Mean	Std. Dev	$\hat{C}_v(\%)$
GPG	6	0.57	0.60	83%	0.5833	0.0103	1.77%
LEK	2	0.61	0.62	0	0.6150	0.0071	1.15%
IDT	8	0.58	0.60	63%	0.5925	0.0071	1.19%
US	5	0.59	0.62	20%	0.6060	0.0114	1.88%
EU	12	0.58	0.63	25%	0.6045	0.0156	2.59%

There are several observations from Table 10.

- There are a total of 16 EP2006 batches, including six batches from GPG, two batches from LEK and eight batches from IDT;
- 83% (5 out of 6) of GPG batches and 63% (5 out of 8) of IDT batches have Content less than the target value of 0.60 mg/mL; in contrast, only 20% of US batches (1 out of 5) and only 25% of EU batches (3 out of 12) have content less than 0.60 mg/mL. This shows that the content of EP2006 tends to be lower than that of US and EU products;
- The sample mean Content of GPG and IDT are 0.5833 mg/mL and 0.05943, which are lower than that of US reference product;
- The sample standard deviation (Std. Dev) of GPG is 0.0103 g/mL, comparable to that of US reference product with a value of 0.0114 mg/mL; while the sample Std. Dev of IDT is 0.0053 mg/mL, only about half of Std. Dev. of the US reference;
- The standard deviations of EP2006, US and EU batches are all less than 3% of the mean content (i.e. $\hat{C}_v < 3\%$)

III.3. PRILIMINARY COMPARISONS

Before we conducted formal statistical equivalence testing, we performed some preliminary comparisons among EP2006, US reference and EU comparator. The results are summarized in Table 11 below.

Table 11: Preliminary Comparisons for Content (mg/mL) Data Based on FDA CMC Statistics Reviewer's Analysis; EAC = Equivalence Acceptance Criteria; $\Delta\mu$ is the mean difference between Test and Reference (Ref.); σ_{ref} is the sample standard deviation of the reference.

Test	Ref.	N_B	N_R	μ_{Ref}	$\Delta\mu$	$\Delta\mu$ / μ_{Ref} (%)	σ_{ref}	EAC ($1.5\sigma_{ref}$)	EAC/ μ_{Ref} (%)
GPG	US	6	5	0.6060	-0.0227	-3.75%	0.0114	0.0171	2.82%
IDT		8			-0.0135	-2.23%			
IDT	GPG	8	6	0.5833	0.0092	1.58%	0.0103	0.0155	2.66%
GPG	EU	6	12[a]	0.6017	-0.0183	-3.04%	0.0163	0.0240	4.00%
EU	US	12	5	0.6060	-0.0018	-0.30%	0.0114	0.0171	2.82%

[a] 6 samples were randomly selected from a total of 12 EU samples for TOST. The id numbers of the sample are 2, 3, 5, 6, 7 and 8 with Mean (EU TOST) = 0.6017, SD (EU TOST) = 0.0160; the other 6 samples are used to estimate σ_{ref} for EAC with Mean(EU STD) = 0.6067 and SD(EU STD) = 0.0163.

Based on Table 11, preliminary comparisons between Biosimilar and US reference data are summarized below:

- Sample mean of Content of the biosimilar product GPG is lower than that of the US batches. In particular, the observed mean difference between GPG and US reference is -0.0227 mg/mL, which is about 3.75% of the mean from the US reference data;
- The statistical EAC for comparisons with US product as the reference is $1.5\sigma_{ref}$ = 0.0171 mg / mL, corresponding to 2.82% of the mean of the US reference data. Such an EAC is comparable to the specification of 0.57 – 0.63 mg/mL for content;
- Similar observations are found for IDT vs. US.

Based on Table 11, preliminary comparisons between Biosimilar and EU comparator data are summarized below:

- Sample mean of Content of GPG is also lower than that of the EU comparator. In particular, the observed mean difference in Content between GPG and EU data is -0.0183 mg/mL, which is about 3.04% of the mean from the US reference data.
- The statistical EAC for comparisons with EU product as the reference is $1.5\sigma_{ref}$ = 0.0245 mg/mL, corresponding to 4.07% of the mean Content from the EU data.

Based on Table 4, preliminary comparisons between US Reference and EU comparator data are summarized below:

- The observed mean difference between EU comparator and US reference is only -0.0018 mg/mL, which is much smaller than the mean difference between GPG and US;

III.4. STATISTICAL EQUIVALENCE TESTING

By applying statistical equivalence testing with proposed EAC values in Table 11, the obtained confidence interval of the mean difference for each comparison is provided in Table 12. Equivalence in means will be concluded if the obtained confidence interval is completely covered by (-EAC, EAC). Due to insufficient number of batches from LEK, LEK is not included for statistical equivalence testing.

Table 12: Statistical Equivalence Testing Results for Content (U/mg) Based on FDA CMC Statistics Reviewer's Analysis; Note: Mean Difference = Test – Reference; EAC = Equivalence Acceptance Criteria; Width = $t_{1-\alpha, NB+NR-2} \times S_{pool}\sqrt{1/N_B + 1/N_R}$

Test	Ref	N_B	N_R	Alpha [a]	Conf. Level [a]	Mean Diff	Width	Conf. Interval	$1.5\sigma_{ref}$	Equivalence? [b]
GPG	US	6	5	9.3%	81.4%	-0.0227	0.0094	(-0.0320, -0.0133)	0.0171	No
IDT		8		8.5%	83.0%	-0.0135	0.0074	(-0.0209, -0.0061)		No
IDT	GPG	8	6	8.5%	83.0%	0.0092	0.0066	(0.0025, 0.0158)	0.0155	No
GPG	EU	6	6[c]	9.3%	81.4%	-0.0183	0.0111	(-0.0294, -0.0073)	0.0240	No
EU	US	12	5	8.8%	82.4%	-0.0018	0.0111	(-0.0129, 0.0092)	0.0171	Yes

[a] If the number of lots from either test or reference is less than six, the confidence level is computed based on six lots.

[b] Pass the statistical equivalence testing if the obtained Confidence Interval is completely covered by (-EAC, EAC).

[c] 6 samples were randomly selected from a total of 12 EU samples for TOST. The id numbers of the sample are 2, 3, 5, 6, 7 and 8 with Mean (EU TOST) = 0.6017, SD(EU TOST) = 0.0160; the other 6 samples are used to estimate σ_{ref} for EAC with Mean(EU STD) = 0.6067 and SD(EU STD) = 0.0163.

There are several observations from Table 12.

- Due to limited number of batches for biosimilar, US and EU products, the confidence level is lower than 90% with a fixed EAC of $1.5\sigma_{ref}$;
- Biosimilar product GPG fails to show statistical equivalence to the US reference because the confidence interval (-0.0320, -0.0133) in mean difference is not covered by the acceptance criteria of (-0.0171, 0.0171). Such a failure is mostly due to the relative large mean difference of -0.0227 mg/mL.

  In addition, the upper confidence limit of GPG vs US is less than zero, indicating the biosimilar product may have a lower mean value of Content than the US reference product. Similar observations are found for the comparison of IDT vs. US;

- Likewise, Biosimilar product GPG fails to show statistical equivalence to the EU comparator because the confidence interval (-0.0294, -0.0073) in mean difference is not covered by the acceptance criteria of (-0.0240, 0. 0240). Such a failure is mostly due to the relative large mean difference of -0.0183 mg /mL.

  In addition, the upper confidence limit of GPG vs. EU is less than zero, indicating the biosimilar product may have a lower mean value of Content than the EU comparator.

- EU comparator shows statistical equivalence to the US reference product.

IV. CONCLUSION

In summary, there is insufficient evidence to show equivalence between the proposed biosimilar product and the US-licensed product for Bioactivity and Content based on the statistical analyses of the available data. To have more reliable assessment of similarity, we recommend that the sponsor should include more batches from the proposed biosimilar product and US reference product.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

XIAOYU DONG
09/10/2014

YI TSONG
09/11/2014

MEIYU SHEN
09/11/2014