CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

761071Orig1s000

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

NDA/Serial Number:	761071 / 6 (Original Resubmission)
Drug Name:	GP2017
Indication(s):	Same indications as Humira
Applicant:	Sandoz
Dates:	Submitted: 10/30/2017 PDUFA: 10/30/2018
Review Priority:	Standard review
Biometrics Division:	Division of Biometrics III
Statistics Reviewer:	Kathleen Fritsch, PhD
Concurring Reviewer:	Mohamed Alosh, PhD
Medical Division:	Division of Dermatology and Dental Products / Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Mark Borigini, MD / Nikolay Nikolov, MD (DPARP) Gary Chiang, MD / David Kettl, MD (DDDP)
Project Manager:	Nina Ton

Keywords: Biosimilar

Table of Contents

1 Executive Summary

Study 301 is a comparative clinical study of GP2017 versus Humira in subjects with moderate to severe psoriasis. The primary endpoint was the proportion of subjects at Week 16 with PASI 75 response. Secondary endpoints included the percent change in PASI up to Week 16 and success on the Investigator's Global Assessment (IGA). The study was conducted in the US, Bulgaria, France, and Slovakia. Subjects in Europe (EU) received EU-approved Humira, while subjects in the US received US-licensed Humira.

The PASI 75 response rates differed by geographical region, primarily due to differences on the Humira arm (53% in the US and 68% in the EU), while subjects receiving GP2017 in the US or EU were similar (approximately 58% in both the US and EU). Only 17% of the data was collected on European subjects. The differences between regions were smaller when comparing the percent change in PASI at Week 16 (reductions of 79% for GP2017 in the US vs. 83% in the EU, and reductions of 77% for Humira in the US vs. 87% in the EU).

Two centers were identified as having unusual results (Center 1268 in the US, and Center 1001 in Slovakia). Both centers exhibited a lower than expected variability in response, and Center 1268 was also notable for the fact that none of the 33 subjects at the center were classified as PASI 75 responders. A clinical inspection of the data at Center 1268 was requested, but the site could not be inspected because the investigator reported that all files related to the study had been lost in an accidental fire. No inspections were requested for Center 1001, because a sufficient amount of data was available from US sites.

Because of the questions of the comparability of the US and EU results, and the questions surrounding the reliability of the data collected at Center 1268 in the US, it may be appropriate to consider analyses based only on US subjects and those based on US subjects, excluding Center 1268. The results of these two subgroup analyses are consistent with the overall analyses, and the results of the subgroup analyses also meet the protocol-specified criteria of having the 90% confidence intervals fall within the pre-specified similarity margin of ±18%. Table 1 presents the results for the overall population (all subjects US + EU), the subset of US subjects, and the subset of US subjects excluding Center 1268. In each case, the results are generally consistent and the 90% confidence intervals are contained within the pre-specified similarity criterion. The secondary endpoints of percent change in PASI and IGA success are consistent with the results of the primary endpoint analysis.

Table 1 – PASI 75 Response Rates at Week 16

	GP2017	Humira	Difference	90% CI
Per Protocol Set				
Overall	N=197	N=196		
	66.8%	65.0%	1.8%	(-6.0, 9.7)
US	N=157	N=157		
	68.0%	62.6%	5.3%	(-3.5, 14.1)
US excluding Center 1268	N=143	N=143		
	74.5%	68.9%	5.6%	(-4.5, 15.7)
Full Analysis Set				
Overall	N=231	N=234		
	58.1%	55.9%	2.2%	(-5.4, 9.7)
US	N=188	N=190		
	57.9%	53.2%	4.7%	(-3.6, 13.1)
US excluding Center 1268	N=174	N=171		
	62.6%	59.1%	3.5%	(-4.9, 12.0)

In the protocol, the applicant originally proposed to provide analyses for the US regulatory submission comparing all GP2017 subjects (US + EU) versus only US Humira subjects, and the applicant submitted these analyses with the original application in 2016. For European regulatory submissions, the applicant planned to use analyses based on all collected data. After the applicant withdrew the original submission because not all facilities were ready for inspection, FDA provided advice regarding the clinical study analyses in the Acknowledge Withdrawal letter. FDA advised the applicant that excluding part of the Humira arm (EU study site subjects) while including all GP2017 subjects (US and EU study sites) breaks the connection to the randomization and could introduce bias, and requested that the applicant submit analyses using all subjects (consistent with what had been planned for European submissions) along with subgroup analyses for US and EU subjects. The applicant submitted the requested analyses with the resubmission of the application in 2017.

2 Introduction

2.1 Overview

GP2017 is being developed as a proposed biosimilar to US-licensed Humira (adalimumab) under section 351(k) of the Public Health Service (PHS) Act. Section 351(i) of the PHS Act defines biosimilarity to mean "that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components" and that "there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity, and potency of the product." As part of their development program, the applicant has conducted a comparative clinical study of GP2017 versus US-licensed and EU-approved Humira in subjects with moderate to severe plaque psoriasis (Study GP17-301). The applicant also conducted four pharmacokinetic studies. BLA 761071 was originally submitted on

8/25/2016. Because some facilities were not available for inspection within an appropriate timeframe, the applicant requested withdrawal of the application on 10/21/2016. The BLA was resubmitted on 10/30/2017.

2.1.1 Comparative Clinical Study

Study 301 was conducted in the US, Bulgaria, France, and Slovakia in subjects with moderate to severe psoriasis. The design details for Study 301 are summarized in Table 2.

Table 2 – Comparative Clinical Study Overview

Study Number	GP17-301	
Study Design	• Treatment Period 1 (Randomization to Week 17) - GP2017 vs. Humira (US-licensed Humira at US sites and EU-approved Humira at EU sites) • Treatment Period 2 (Week 17 to Week 35) - Among subjects achieving at least PASI 50 at Week 16, subjects from both treatment arms were randomized 2:1 to remain on initial treatment or alternate between the two treatments over 6-week periods • Extension Period (Week 35 to Week 51) – Subjects received the same treatment as in Treatment Period 1	
Inclusion criteria	Adult subjects at least 18 years of age with active, clinically stable plaque psoriasis with at least 10% BSA, PASI ≥ 12, and IGA ≥ 3. No previous exposure to adalimumab was permitted.	
Treatment regimen	Loading dose of 80 mg at Week 0, followed by 40 mg at Week 1 and every other week thereafter.	
Primary endpoint	PASI 75 at Week 16	
Secondary endpoint	Average percent change from baseline in PASI up to Week 16	
Treatment arms and Sample Size	GP2017: 231 (US: 188/ EU: 43)	Humira: 234 (US: 190 / EU: 44)
Study location	US, Bulgaria, France, and Slovakia	

PASI = Psoriasis Area Severity Index, BSA =Body Surface Area, IGA = Investigator's Global Assessment

2.1.2 Regulatory History

The applicant submitted BLA 761071 for GP2017 on 8/25/2016. During the filing review, the product quality team noted that not all facilities were ready for inspection within the expected time frame during the review cycle. After being advised of this issue, the applicant withdrew the application on 10/21/2016.

Although not filing issues, the statistical reviewer identified issues related to the statistical analyses for the applicant to address upon resubmission of the BLA. In accordance with the amended protocol, the submitted study report for Study 301 focused on analyses comparing GP2017 from all sites (US and EU) versus Humira from only US sites. The study report did not include analyses for the total study population nor analyses for subjects randomized to EU-Humira at the EU sites. Additionally, the datasets did not include an indicator as to which observations were included in the primary analysis or

sufficient detail regarding missing data handling. Because the applicant withdrew the original application shortly after submission, these database issues were not addressed in the original review cycle. However, the applicant was advised in FDA's Acknowledge Withdrawal letter to address the issues related to the submitted analyses and datasets if they resubmit the BLA.

The applicant resubmitted the BLA on 10/30/2017 after addressing the manufacturing, analysis, and dataset issues.

Pre-IND/IND Development

The design and statistical analysis of the comparative clinical study (Study 301) was discussed with the applicant at a Type 2 Biosimilar Biologic Product Development (BPD) meeting held on 1/14/2013. At the meeting, the following characteristics of Study 301 were discussed.

- FDA recommended enrolling subjects with PASI $\geq$ 12 rather than PASI $\geq$ 10 (along with BSA $\geq$ 10% and IGA $\geq$ 3)
- FDA stated that a primary endpoint of PASI 75 at Week 16 with a similarity margin of 18% is acceptable.
- The applicant proposed to increase the duration of switch/transition periods from 4 to 6 weeks.
- The applicant stated that they would treat subjects at US sites with US-licensed Humira.

The applicant opened the IND on 11/6/2013 with the protocol for Study 301 (Amendment 1). The applicant began enrolling subjects in Europe on 12/18/2013. The following comments related to the study design and statistical analysis were conveyed to the applicant on 3/17/2014 in a Study May Proceed letter.

- Clarify whether and how the stratification factors will be used in the primary analysis
- The primary analysis population has been specified as the per protocol set (PPS). For equivalency analyses, both the PPS and full analysis set (FAS) populations should be used.
- Clarify how center effects or treatment-by-center interactions will be assessed
- Provide details for the key secondary endpoint analyses.
- Provide details regarding the missing data handling proposals.

Meanwhile, the applicant revised the protocol (Amendment 2, dated 2/13/2014) and submitted it to the IND on 2/28/2014. The key design modifications included: (1) adding stratification by region (subjects in US enrolled after implementation of Amendment 2), and (2) clarifying that US-licensed Humira would be used in US and EU-approved Humira would be used outside the US. The first US subject was enrolled on 5/29/2014.

Amendment 3 was finalized on 9/10/2014, and was implemented after 94 subjects enrolled. This revision was submitted to the IND on 9/24/2014. The key modifications included: (1) adding an Extension Period (Weeks 35-51), (2) capping EU enrollment at 90 subjects, (3) clarifying that analyses for EMA will use all subjects, while analyses for

FDA will use all GP2017 subjects (EU + US) but only US-Humira subjects, and (4) adding stratification factors to the statistical models.

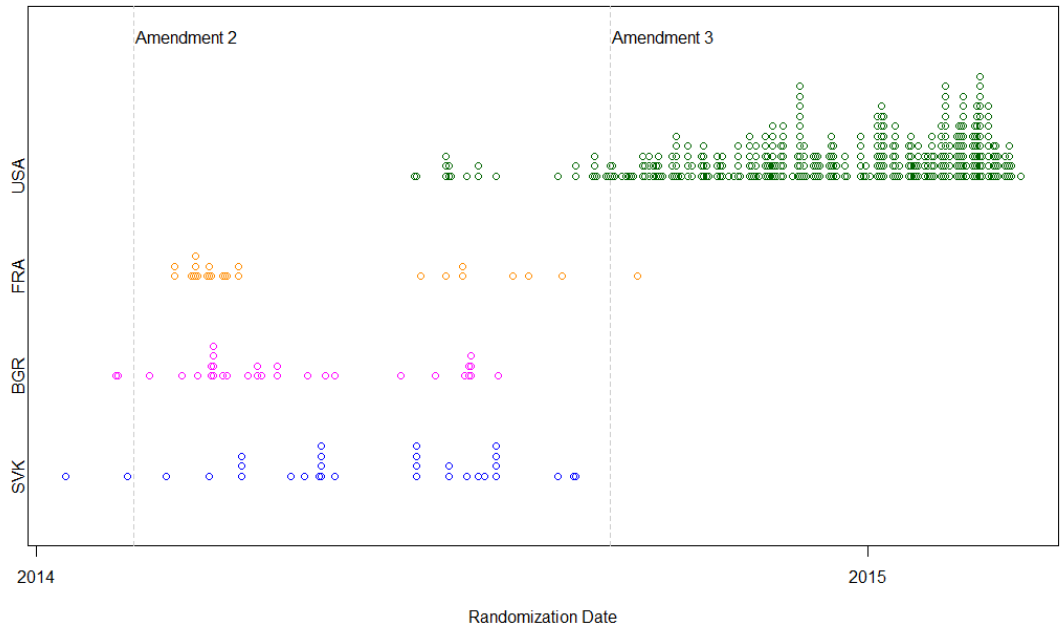
The applicant also submitted a response to the Study May Proceed letter on 5/14/2015 outlining the changes that had been made in Amendment 3. In addition to highlighting the changes to the planned analysis for the US application (comparing US-Humira subjects versus all GP2017 subjects) and the inclusion of the stratification factors of prior systemic therapy and body weight in the analysis models, the applicant clarified that because the study would enroll only a limited number of subjects per site, center effects or interactions would not be assessed. In addition, the sponsor noted that the primary method of handling missing data would be to classify subjects with missing data as non-responders.

The statistical reviewer noted that excluding part of the adalimumab arm (EU subjects) while including all GP2017 subjects (US and EU) breaks the connection to the randomization and could introduce bias; therefore, such an analysis is not recommended. The reviewer also recommended conducting sensitivity analyses regarding missing data handling (statistical review dated 9/12/2015). However, these comments were not conveyed to the applicant at the time. This information was conveyed to the applicant in the letter acknowledging withdrawal of the BLA in 2016.

Protocol 301 was amended one more time (Amendment 4) on 10/7/2015 after all subjects had completed Treatment Period 1. This amendment was submitted to the IND on 5/11/2016. The key modification was to note that the analyses for Treatment Period 2 and Extension Period will be conducted after all subjects complete the Extension Period, rather than producing study reports after each period.

The time course of randomization date relative to amendment date by country is presented in Figure 1. All but one European subject were enrolled prior to the implementation of Amendment 3, while the majority of US subjects were enrolled after the implementation of Amendment 3.

Figure 1 – Randomization Date by Country



Source: reviewer analysis

2.2 Data Sources

This reviewer evaluated the applicant's clinical study report, clinical summaries, and proposed labeling. The submission was submitted in eCTD format and was entirely electronic. Both SDTM and analysis datasets were submitted. The analysis datasets for Study 301 used in this review are archived at \\cdsesub1\evsprod\bla761071\0005\m5\datasets\gp17-301\analysis\adam\datasets

3 Statistical Evaluation

3.1 Data and Analysis Quality

In the original BLA review cycle, the submitted study report for Study 301 focused on analyses comparing GP2017 from all sites (US and EU) versus Humira from only US sites. The study report did not include any whole-study analyses or any analyses including the subjects randomized to EU-Humira at the EU sites. Additionally, the datasets did not include an indicator as to which observations were included in the primary analysis or sufficient detail regarding missing data handling. Because the applicant withdrew the original application shortly after submission, these database issues were not addressed in the original review cycle. However, in the Acknowledge Withdrawal letter FDA included the following comments and requests related to the submitted analyses and datasets:

1. Although you have submitted the observed data and endpoints based on the PASI score in the ADXE.xpt dataset for Study GP17-301, the dataset does not include all of the necessary information to conduct the primary and key secondary analyses. In particular, the dataset does not include information regarding which

observations were included in the Week 16 analysis of PASI 75 or how missing data was handled for the per protocol set and full analysis set. Therefore,

a. Submit the statistical analysis programs used to create analysis-ready ('One PROC Away') datasets that can be used to analyze the primary and key secondary endpoints. In particular, ensure that the programs address which visit is used for the primary Week 16 PASI 75 analysis (especially when it is not the nominal Week 16 visit) and how missing data is handled for both the per protocol set and full analysis set.
b. In addition to the programs, submit the analysis-ready datasets to permit further analysis of the PASI 75 and percent change in PASI data. The analysis-ready datasets should include imputed data, along with appropriate flagging variables to indicate which observations were imputed. These datasets should also include variables needed for the analyses, such as stratification variables (original and actual), demographic variables, per protocol/full analysis population flags, and flags to indicate which observations are used in the Week 16 analyses. To keep the size of the analysis-ready datasets manageable, they should focus on PASI 75, PASI, and percent change in PASI endpoints.
c. Submit the statistical analysis programs used to create the estimates and confidence intervals for the primary and key secondary analyses (e.g. logistic regression, MMRM, and delta method calculations).

2. In Study GP17-301, you have only provided efficacy analyses that compare subjects treated at US study sites with US-licensed Humira vs. subjects treated with GP2017 at both US and EU study sites. You did not include subjects treated with EU-approved Humira at EU sites in the efficacy analyses. Excluding part of the Humira arm (EU study site subjects) while including all of the GP2017 subjects (US and EU study sites) breaks the connection to the randomization and could introduce bias. Submit supplementary analyses for the primary and key secondary endpoints using the full population (i.e., subjects treated at EU and US sites combined, for both treatment arms), along with supportive subgroup analyses by region (i.e., for both the US and EU subgroups).

The applicant addressed these issues in the resubmission of their application on 10/30/2017 for the current review cycle, and no further requests for datasets, analyses, or programs were made.

3.2 Evaluation of Efficacy

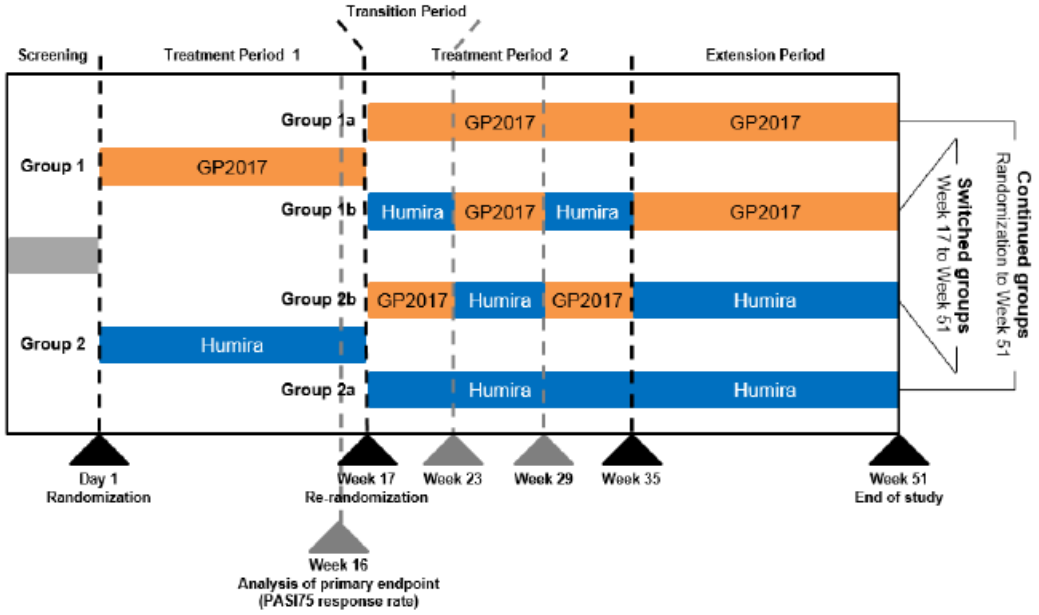
3.2.1 Study Design and Statistical Analysis

Study 301 was a randomized, double-blind comparative clinical study of GP2017 and Humira in subjects with moderate to severe plaque psoriasis. The study was conducted in the US, Bulgaria, France, and Slovakia. Subjects randomized to Humira in the US received US-licensed Humira and subjects randomized to Humira in Europe received EU-approved Humira. The study enrolled subjects age 18 years and older with PASI ≥ 12, Investigator's Global Assessment (IGA) ≥ 3 (moderate or severe) and total body surface area (BSA) ≥ 10%. Subjects were not allowed to have any previous exposure to adalimumab. Subjects were to have been diagnosed at least 6 months before

randomization, and were to have either previously received phototherapy or systemic therapy or be candidates for such therapy.

The study enrolled 465 subjects, 231 to GP2017 and 234 to Humira. Subjects were enrolled at 73 centers. Randomization was stratified by prior systemic therapy (none/any), region (EU/US), and body weight (< 90 kg / ≥ 90kg). Subjects received subcutaneous injection of 80 mg at Day 1, 40 mg at Week 1 and 40 mg every 2 weeks thereafter. The primary timepoint for assessment was Week 16. Subjects who achieved at least a PASI 50 response at Week 16 were re-randomized in a 2:1 ratio to either continue their initial treatment or switch treatments at 6-week intervals during Treatment Period 2 (Weeks 17 to 35). Re-randomization at Week 17 was stratified by region (EU/US) only. From Weeks 35 to 51 (Extension Period), subjects received the treatment initially assigned on Day 1. See Figure 2.

Figure 2 – Study Design



Source: pg 106 of gp17-301-report-body.pdf (2/9/2017 version)

Subjects were evaluated at screening, baseline, and Weeks 1, 3, 5, 7, 9, 11, 13, 15, 16, and 17 in Treatment Period 1, at Weeks 19, 21, 23, 25, 27, 29, 31, 33, and 35 in Treatment Period 2, and at Weeks 37, 39, 41, 43, 45, 47, 49, and 51 in the Extension Period. Because the appearance of the GP2017 and Humira syringes differed, unblinded study site personnel not involved in study assessments administered all study treatments.

Efficacy was assessed using the PASI scale, BSA, and IGA scale. The 5-point IGA scale is presented in Table 3.

Table 3 – Investigator's Global Assessment

Score	**Short description**	**Detailed description**
0	Clear	No signs of psoriasis. Post-inflammatory hyperpigmentation may be present
1	Almost clear	Normal to pink coloration of lesions, No thickening No to minimal (focal) scaling
2	Mild	Pink to light red coloration, Just detectable to mild thickening, Predominantly fine scaling
3	Moderate	Dull bright red, clearly distinguishable erythema, Clearly distinguishable to moderate thickening, Moderate scaling
4	Severe	Bright to deep dark red coloration, Severe thickening with hard edges, Severe/coarse scaling covering almost all or all lesions

Source: pg 130 of gp17-301-report-body.pdf (2/9/2017 version)

The protocol defined he full analysis set (FAS) as all randomized subjects to whom study treatment has been assigned. Subjects were to be analyzed according to the treatment randomized. The per protocol set (PPS) was defined as subjects who completed the study up to Week 16 and had no major protocol deviations. Subjects who discontinued due to unsatisfactory therapeutic effect, as long as they received at least 4 weeks/2 doses of treatment, were included in the PPS as non-responders. At the blinded data review meeting, the per protocol set was finalized to include subjects who met key inclusion/exclusion criteria, had a PASI score at the baseline visit, had at least 9 doses of study treatment, included the Week 15 dose, had a PASI assessment within 14 days after the Week 15 dose, and who were not unblinded prior to Week 16.

The final version of the protocol stated that the primary treatment comparison would be between all GP2017 subjects (EU+US) and Humira subjects from the US only. A subgroup analysis would be performed using only subjects from the US for both treatment arms. These analyses were submitted in the original application. In the resubmission, the applicant provided analyses for all subjects (EU + US for both treatment arms), along with US and EU subgroup analyses. This review will focus on the full study analyses and regional subgroup analyses, as these analyses preserve the randomization, rather than the analyses that compare all GP2017 subjects to the Humira subjects from the US only.

The primary endpoint was PASI 75 at Week 16. The statistical analysis plan (SAP) clarified that the PASI score to be used in this analysis was defined as the PASI score within 14 calendar days and closest to the Week 15 study dose date, and not necessarily the nominal Week 16 visit PASI score. The key secondary endpoint was the percent change from baseline in PASI score up to Week 16. The percent change in PASI up to Week 16 was to be analyzed in two ways: using a mixed-model repeated measures (MMRM) analysis and an analysis based on the average treatment effect (ATE). The primary analysis population was the PPS and the supportive analysis population was the FAS. Missing data handling was not an issue with the PPS, because, by definition, all

subjects in the PPS had completed the study through Week 16 and had a PASI assessment within 14 days after the Week 15 dose. In the FAS analysis, all subjects with missing Week 16 data were imputed as non-responders.

The difference in PASI 75 response was evaluated using covariate-adjusted 90% confidence intervals and a similarity margin of ± 18%. PASI 75 response was analyzed with logistic regression with terms for treatment group, body weight classification, and prior systemic therapy classification. The delta method was used to calculate the standard error for the confidence interval. The details of how the delta method was used to calculate the confidence intervals is presented in the Appendix. When the applicant conducted the blinded data review meeting, the study team noted that many subjects had discrepancies between the prior therapies recorded in the clinical database and the classification used in the randomization. Therefore, the applicant modified the SAP to state that the body weight stratum and prior therapy stratum classifications used in the logistic regression model were to be derived from the information in the clinical database, rather than the classification entered into the IRT (interactive response technology) system at randomization.

For the key secondary endpoint of percent change in PASI up to Week 16, the first analysis used a mixed-effect model repeated measures (MMRM) analysis with factors for treatment group, visit, treatment-by-visit interaction, body weight classification, and prior systemic therapy classification, and a covariate for baseline PASI score. The model used an unstructured covariance matrix. A 90% confidence interval for the difference in adjusted means was calculated. The second analysis calculated the weighted average treatment effect for each subject across Weeks 1, 3, 5, 7, 9, 11, 13, 15, and 16[1] (weighted by the distance between visits) and then analyzed the subject mean values with ANCOVA with terms for treatment group, body weight classification, prior systemic therapy classification, and baseline PASI as a covariate. Missing data was not imputed for the percent change in PASI analyses for either the PPS or FAS populations.

Additional secondary endpoints included PASI 50, PASI 75, PASI 90, and PASI 100 at each visit in Treatment Period 1, percent change in PASI at each visit in Treatment Period 1, IGA response (0 or 1) at each visit in Treatment Period 1, change from baseline in IGA at each visit in Treatment Period 1, Dermatology Life Quality Index (DLQI), EuroQol 5-Dimension Health Status Questionnaire (EQ-5D), and Health Assessment Questionnaire-Disability Index (HAQ-DI, in subjects with psoriatic arthritis only). See the Statistical Review by Rebecca Rothwell, Ph.D (dated 6/27/2018) for assessment of HAQ-DI in subjects with psoriatic arthritis.

3.2.2 Stratification Factors

Although the protocol stated that if the actual stratum differs from the assigned stratum in the Interactive Randomization Technology (IRT) system, then the assigned stratum will

[1] Note that the SAP states that the average treatment effect will be calculated over Weeks 2, 4, 8, and 12; however, this appears to by a typo as visits were conducted on Weeks 1, 3, 5, 7, 9, 11, 13, 15, and 16. The statistical programs use data from Weeks 1, 3, 5, 7, 9, 11, 13, 15, and 16.

be used in the analyses, during the blinded data review meeting (BDRM) the applicant identified 77 (16.6%) subjects who were misclassified in the prior therapy or weight stratum based on the weight and prior therapy information recorded in the case report form. In this group 69 subjects were misclassified regarding prior systemic therapy only, 6 subjects were misclassified regarding weight only, and 2 subjects were misclassified on both stratification factors. The misclassification of the 71 (15.3%) subjects into the prior systemic therapy stratum are presented in Table 4. Four of the 8 subjects misclassified on the weight stratum (<90 kg vs. ≥ 90 kg) weighed 89.81 kg (198 lbs) at baseline and were stratified into the ≥ 90 kg stratum rather than the < 90 kg stratum, and this appears to be a rounding issue. The other 4 subjects had weights ranging from 83.92 to 105.24 kg and were randomized using the incorrect stratum for unknown reasons. Because of the relatively high rate of misclassification based on prior systemic therapy, the applicant modified the statistical analysis plan (SAP) to state that the analyses would be conducted using the 'true' factor rather than the assigned factor, as the "correct true stratification is thought to be more appropriate." The applicant has not otherwise provided an explanation for the relatively high rate of misclassification of subjects regarding prior systemic therapy.

Table 4 – Misclassification into Prior Systemic Therapy Stratum

Randomization Stratum for prior systemic therapy	*Actual prior systemic therapy history*			
	GP2017		Humira	
	No	Yes	No	Yes
No	118/140 (84.3%)	22/140 (15.7%)	117/142 (82.4%)	25/142 (18.6%)
Yes	10/91 (11.0%)	81/91 (89.0%)	14/92 (15.2%)	78/92 (84.8%)

Source: Reviewer Analysis

3.2.3 Subject Disposition

Study 301 randomized 465 subjects: 231 to GP2017 and 234 to Humira. Approximately 14% of subjects discontinued treatment during Treatment Period 1. The most common reasons for study discontinuation were subject/guardian decision and lost to follow-up. The discontinuation rates were similar for the two arms. See Table 5. In Treatment Period 2, subjects who achieved at least PASI 50 at Week 16 were re-randomized to continue the initial treatment or switch between treatments at 6-week intervals. Approximately 9% of the 379 subjects who were re-randomized at Week 17 discontinued by Week 35. The most common reasons for discontinuation in Treatment Period 2 were subject/guardian decision and lack of efficacy. There are no clear patterns in the discontinuation data in Treatment Period 2. See Table 6.

Table 5 – Disposition of Subjects in Treatment Period 1 (Randomization to Week 17)

	GP2017	Humira
Subjects Randomized	231	234
Discontinued Treatment Period 1	30 (13%)	33 (14%)
Subject/Guardian decision	15 (6%)	11 (5%)
Lost to follow-up	6 (3%)	4 (2%)
Lack of efficacy	4 (2%)	2 (1%)
Adverse event	3 (1%)	5 (2%)
Protocol deviation	2 (1%)	8 (3%)
Physician decision	--	2 (1%)
Pregnancy	--	1 (<1%)

Source: pg 173 of gp17-301-report-body.pdf (2/9/2017 version)

Table 6 – Disposition of Subjects in Treatment Period 2 (Week 17 to Week 35)

	Treatment in Period 1			
	GP2017 N=231		Humira N=234	
Completed Treatment Period 1	201		201	
	Treatment Sequence in Period 2			
	GP2017 to Switch	Continued GP2017	Humira to Switch	Continued Humira
Re-randomized	63	126	63	127
Completed Treatment Period 2	59 (94%)	112 (89%)	57 (90%)	116 (91%)
Discontinued Treatment Period 2	4 (6%)	14 (11%)	6 (10%)	11 (9%)
Subject/guardian decision	1 (2%)	7 (6%)	3 (5%)	1 (1%)
Lack of efficacy	3 (5%)	2 (2%)	2 (3%)	6 (5%)
Lost to follow-up	--	1 (1%)	1 (2%)	--
Adverse event	--	1 (1%)	--	4 (3%)
Death	--	1 (1%)	--	--
No-compliance with study treatment	--	1 (1%)	--	--
Pregnancy	--	1 (1%)	--	--

Source: pg 175 of gp17-301-report-body.pdf (2/9/2017 version)

Approximately 15% of subjects on each treatment arm were excluded from the per protocol analysis set (PPS). The PPS consisted of all subjects who completed the study up to Week 16 with no major protocol deviations; however, subjects who discontinued due to unsatisfactory therapeutic effect were included as non-responders, provided they received at least 2 injections. The reasons for exclusion were balanced across the treatment arms. Most subjects excluded from the PPS discontinued before Week 16 and did not have a Week 16 PASI assessment. See Table 7.

Table 7 – Per Protocol Analysis Set Exclusions

	GP2017 N=231	Humira N=234
Subjects meeting criteria for exclusion from PPS	38 (16%)	40 (17%)
Missing Week 16 PASI assessment	34 (15%)	35 (15%)
Not dose compliant	31 (13%)	33 (14%)
Discontinued prior to Week 16	29 (13%)	32 (14%)
Treatment deviation	5 (2%)	3 (1%)
Selection criteria not met	2 (1%)	2 (1%)
Other	--	1 (<1%)
Subjects meeting criteria for re-inclusion in the PPS	4 (2%)	2 (1%)
Discontinued due to lack of efficacy (after at least 2 injections)	4 (2%)	2 (1%)
Final number of subjects excluded from the PPS	34 (15%)	38 (16%)
Final number of subjects included in the PPS	197 (85%)	196 (84%)

Note: Subjects may have had more than one reason for exclusion
Source: pg 185 of gp17-301-report-body.pdf (2/9/2017 version)

3.2.4 Baseline Characteristics

The baseline demographics were generally balanced across the treatment groups in Study 301. The mean age was about 46 years, with 11% of subjects age 65 years and older. The majority of subjects were male (61%) and white (85%). See Table 8.

Table 8 – Baseline Demographics

	GP2017 N=231	Humira N=234
Age (years)		
Mean	45.6	46.9
Range	18 - 81	18 - 84
18 to 64 years	205 (89%)	209 (89%)
65 + years	26 (11%)	25 (11%)
Gender		
Female	89 (39%)	92 (39%)
Male	142 (61%)	142 (61%)
Race		
White	196 (85%)	201 (86%)
Black	14 (6%)	9 (4%)
Asian	3 (1%)	5 (2%)
Native American	4 (2%)	4 (2%)
Pacific Islander	--	1 (<1%)
Unknown	3 (1%)	--
Other	11 (5%)	14 (6%)
Weight (kg)		
Mean	92.76	90.95
<90 kg[a]	120 (52%)	127 (54%)
≥90 kg[a]	111 (48%)	107 (46%)

[a] Based on actual weight recorded at baseline
Source: pg 189 of gp17-301-report-body.pdf (2/9/2017 version) and reviewer analysis.

To be enrolled in the study, subjects were to have PASI ≥ 12, Investigator's Global Assessment (IGA) ≥ 3 (moderate or severe) and total body surface area (BSA) ≥ 10%. At baseline, subjects had a mean PASI score of about 20 and 29% BSA. Approximately 66% had an IGA score of moderate. More than half of subjects reported having no prior systemic therapy (both as defined by the randomization stratum and the information in the clinical database). As discussed in Section 3.2.2 about 15% of subjects had discrepancies between the prior therapy stratum and prior therapy information recorded in the case report form, with a higher proportion of subjects having had prior systemic therapy according to the clinical information collected on subjects than was reflected in the randomization stratum selected. The baseline disease characteristics were balanced across treatment arms. See Table 9.

Table 9 – Baseline Disease Characteristics

	GP2017 N=231	Humira N=234
PASI		
Mean (SD)	19.9 (8.55)	20.2 (7.68)
Range	12.0-58.8	11.7-53.4
BSA		
Mean (SD)	28.9 (17.09)	29.7 (15.61)
Range	10-90	8.75-85
IGA		
Moderate	152 (66%)	154 (66%)
Severe	79 (34%)	80 (34%)
Prior systemic therapy (randomization strata)		
No	140 (61%)	142 (61%)
Yes	91 (39%)	92 (39%)
Prior systemic therapy (clinical database)		
No	128 (55%)	131 (56%)
Yes	103 (45%)	103 (44%)

Source: pg 193-194 of gp17-301-report-body.pdf (2/9/2017 version) and reviewer analysis.

3.2.5 Primary Efficacy Endpoint

The primary efficacy endpoint was PASI 75 at Week 16. The primary analysis set was the PPS with the FAS as supportive. The difference in PASI 75 response was evaluated using covariate-adjusted 90% confidence intervals and similarity margins of ± 18%. PASI 75 response was analyzed with logistic regression with terms for treatment group, body weight classification, and prior systemic therapy classification. Based on discussions at the blinded data review meeting, the body weight and prior therapy classifications used in the logistic regression model were to be derived from the information in the clinical database, rather than the classification entered into the IRT (interactive response technology) system at randomization. The results on the PPS and FAS analysis sets were similar with an estimated treatment difference of approximately 2%, and 90% confidence intervals contained within ± 18% similarity margin, and thus met the pre-specified criteria. See Table 10.

Table 10 – PASI 75 at Week 16

Analysis Set	GP2017	Humira	Difference	90% CI
PPS	N=197 66.8%	N=196 65.0%	1.8%	(-6.0, 9.7)
FAS	N=231 58.1%	N=234 55.9%	2.2%	(-5.4, 9.7)

PPS= Per Protocol Set, FAS = Full Analysis Set
Source: pg 200 of gp17-301-report-body.pdf (2/9/2017 version) and reviewer analysis.

The applicant's original proposal in the protocol was to analyze the primary efficacy endpoint using all GP2017 data (from both US and EU) sites, but only Humira data from US sites (where subjects received US-licensed Humira). These results are presented in Table 11. As FDA noted in the letter acknowledging the applicant's withdrawal of the application following the original submission, excluding part of the Humira arm (EU study site subjects) while including all GP2017 subjects (US and EU study sites) breaks the connection to the randomization and could introduce bias. Importantly, such an analysis assumes that the GP2017 subjects in the EU are the same as the GP2017 subjects in the US, and that there are no country or site differences that could influence the results. However, as can be seen in Table 12, the subgroup results were varied with the PASI 75 response rate was higher on GP2017 than Humira at the US sites and the PASI 75 response rate was lower on GP2017 than Humira at the EU sites, reinforcing the concept that comparisons that respect the randomization are important. For comparisons against US-Humira, the subgroup analysis within US sites is more appropriate. As noted in Table 12, the treatment difference for PASI 75 response was 5.3% (favoring GP2017), and the 90% confidence intervals for the PPS and FAS remained within the ± 18% margins. Within the EU sites the treatment difference favored EU-Humira, but the sample size was small and the confidence interval was wide.

Table 11 – Protocol-Specified Analysis of PASI 75 for All GP2017 vs. US-Humira Only

Analysis Set	GP2017	US-Humira	Difference	90% CI
PPS	N=197 67.0%	N=157 62.4%	4.5%	(-3.8, 12.9)
FAS	N=231 58.1%	N=134 53.0%	5.1%	(-2.9, 13.1)

PPS= Per Protocol Set, FAS = Full Analysis Set
Source: pg 204 of gp17-301-report-body.pdf (2/9/2017 version)

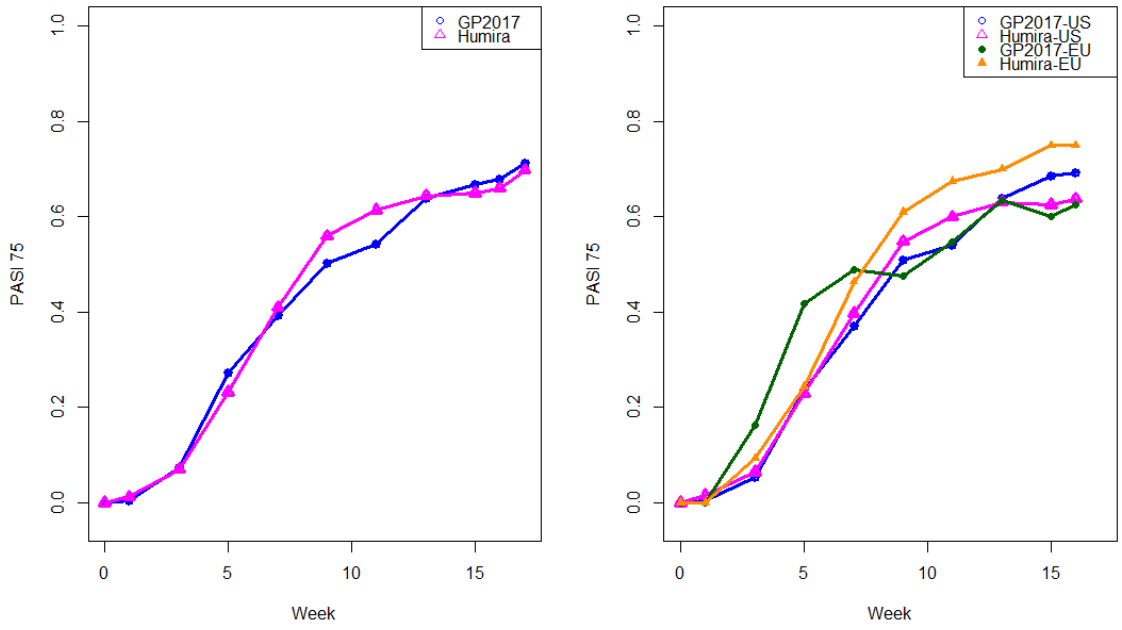
Table 12 – PASI 75 Response by Region

Analysis Set	GP2017	Humira	Difference	90% CI
US (PPS)	N=157	N=157		
	68.0%	62.6%	5.3%	(-3.5, 14.1)
EU (PPS)	N=40	N=39		
	60.4%	76.1%	-15.7%	(-30.5, -0.9)
US (FAS)	N=188	N=190		
	57.9%	53.2%	4.7%	(-3.6, 13.1)
EU (FAS)	N=43	N=44		
	58.0%	68.4%	-10.4%	(-26.0, 5.2)

PPS= Per Protocol Set, FAS = Full Analysis Set
Source: pg 204 of gp17-301-report-body.pdf (2/9/2017 version) and reviewer analysis

PASI 75 response rates by visit are presented in Figure 3. The patterns over time are similar to the results at Week 16 – the curves using the data from the EU and US combined are similar for GP2017 and Humira, but the curves broken down by region were more separated.

Figure 3 – PASI 75 Response by Week (Overall and by Region) – Observed Data



Source: Reviewer analysis

3.2.6 Secondary Endpoint

The key secondary endpoint was the percent change in PASI up to Week 16. The percent change in PASI up to Week 16 was to be analyzed in two ways: using a mixed-model repeated measures (MMRM) analysis and an analysis based on the average treatment effect (ATE). Both analyses evaluated the average percent change in PASI throughout Treatment Period 1. The protocol specified that each of these analyses for the mean difference in percent change in PASI would be compared with similarity margins of ±15.

For these analyses, the estimated differences ranged from -0.7 to 1.2 and all 90% confidence intervals were well within the specified margins. See Table 13.

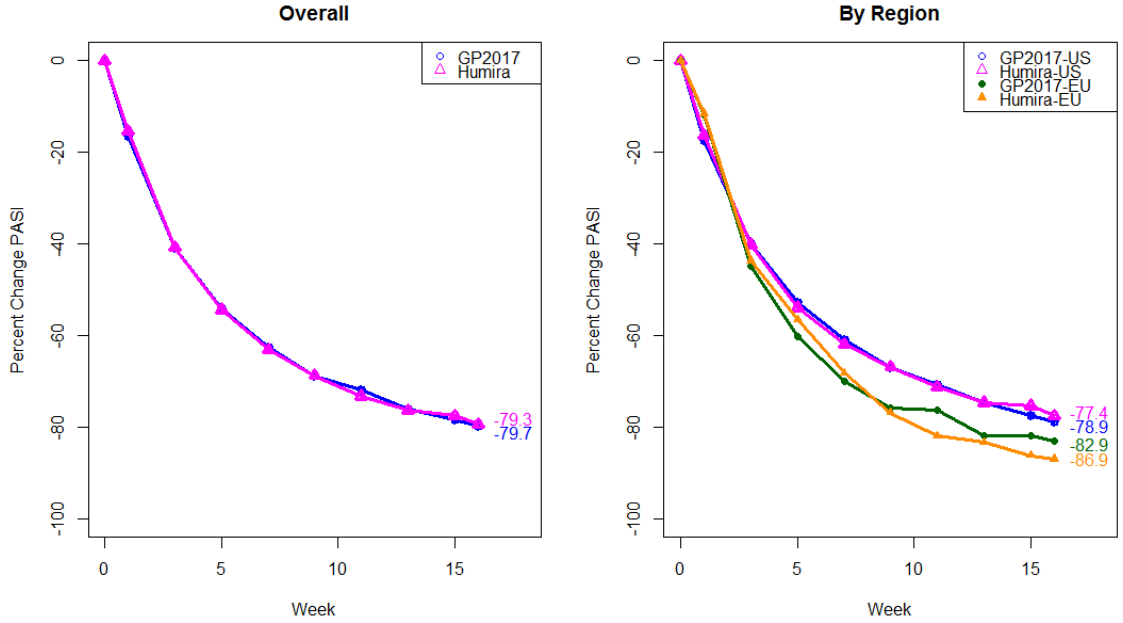
Table 13 – Secondary Endpoint – Percent Change in PASI up to Week 16

Analysis	GP2017	Humira	Difference[a]	90% CI[a]
PPS	N=197	N=196		
MMRM[a]	-60.7 (1.54)	-61.5 (1.55)	0.8	(-2.5, 4.2)
Avg. Trt. Effect[b]	-59.7 (1.59)	-60.8 (1.61)	1.2	(-2.1, 4.4)
FAS	N=231	N=234		
MMRM[a]	-60.1 (1.61)	-59.4 (1.61)	-0.7	(-4.2, 2.8)
Avg. Trt. Effect[b]	-58.0 (1.69)	-57.5 (1.68)	-0.5	(-3.9, 2.9)

[a] LS means, SE and 95% CI were estimated by a Mixed-Effects Repeated Measures (MMRM) model with treatment, visit, treatment-by-visit interaction, body weight strata, region and prior systemic therapy, as fixed factors and baseline PASI score as covariate.
[b] ATE is the weighted average of % change from baseline in PASI scores between Week 1 and Week 16 (weights based on the time interval between two consecutive visits). LS means, SE and 95% CI were estimated using an ANCOVA model with treatment, body weight strata, region and prior systemic therapy as fixed effects and baseline PASI score as covariate.
Source: pg 208 of gp17-301-report-body.pdf (2/9/2017 version) and reviewer analysis

The applicant's analysis of the percent change in PASI averaged the results across all visits through Week 16. Figure 4 presents the percent change in PASI values at each visit and by region (EU/US), including the values at Week 16. The percent change in PASI curves are smoother than the PASI 75 response curves, but otherwise the results are similar: the GP2017 and Humira curves are very similar in the overall population, but some differences between the US and EU outcomes are also evident in the percent change results.

Figure 4 – Percent Change in PASI (Overall and by Region)

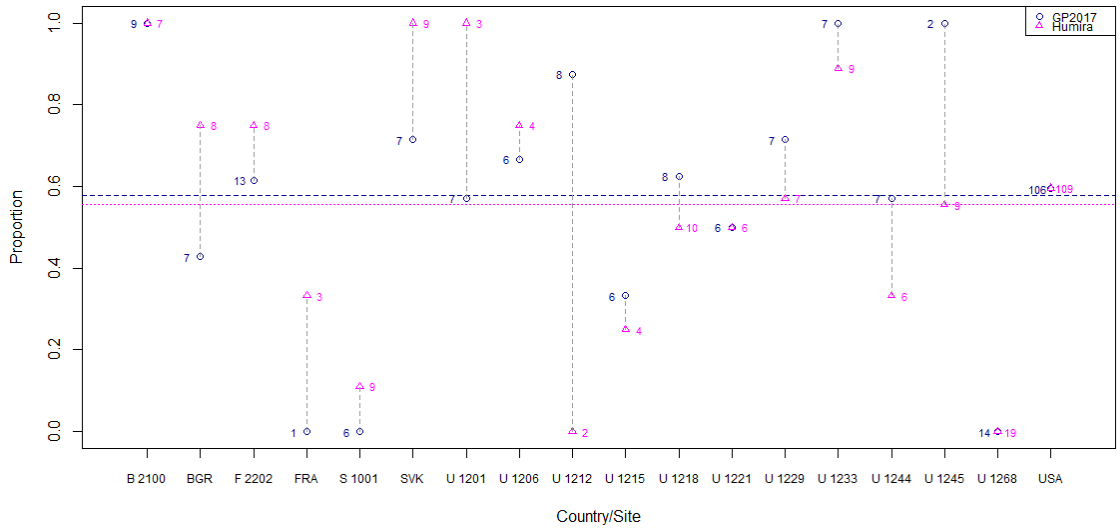


Source: Reviewer analysis.

3.2.7 Efficacy by Center

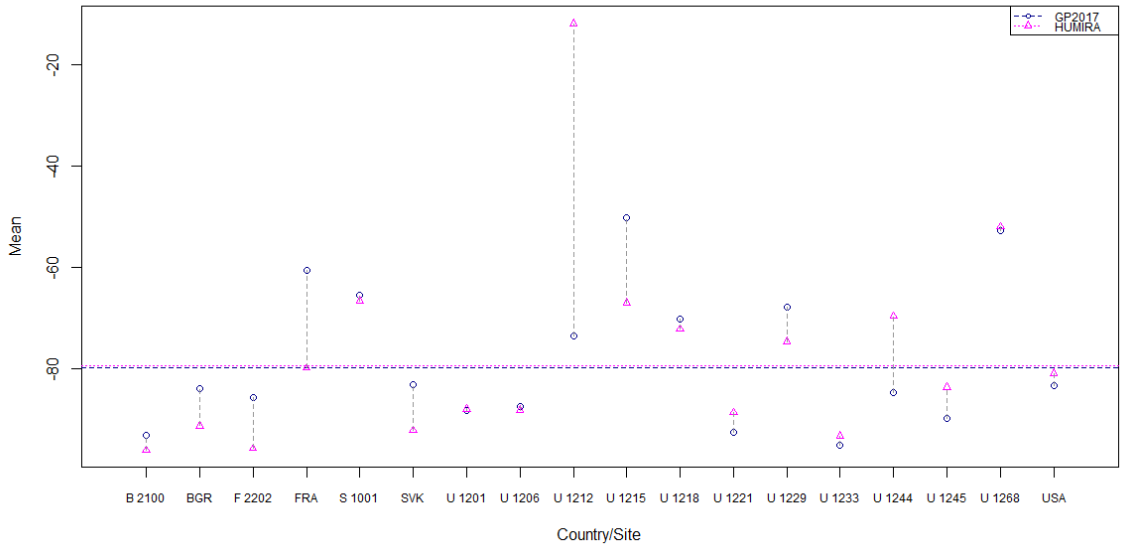
As noted in Section 3.2.5, GP2017 had a higher observed response rate than Humira in the US, but lower in the EU. Thus, response rates were also investigated by center. Because the study was conducted at 73 different centers, many centers enrolled few subjects. Thus, the by-center evaluation focused on the 14 centers that enrolled at least 10 subjects. The subjects from the remaining centers were grouped by country. The PASI 75 response rates by center/country are presented in Figure 5 and the percent change in PASI results are presented in Figure 6. Both figures show that the results at the larger centers generally follow the same pattern as the countries did. That is, the outcomes in the EU countries were generally more favorable on the Humira arm, while the outcomes in the US were generally more favorable on the GP2017 arm, though the results at the US centers were more varied for the percent change in PASI.

Figure 5 – PASI 75 Response by Center/Country



B/BGR = Bulgaria, F/FRA = France, S/SVK = Slovakia, U/USA = United States
Centers with at least 10 subjects are shown individually; the remaining centers are grouped within country
Source: reviewer analysis

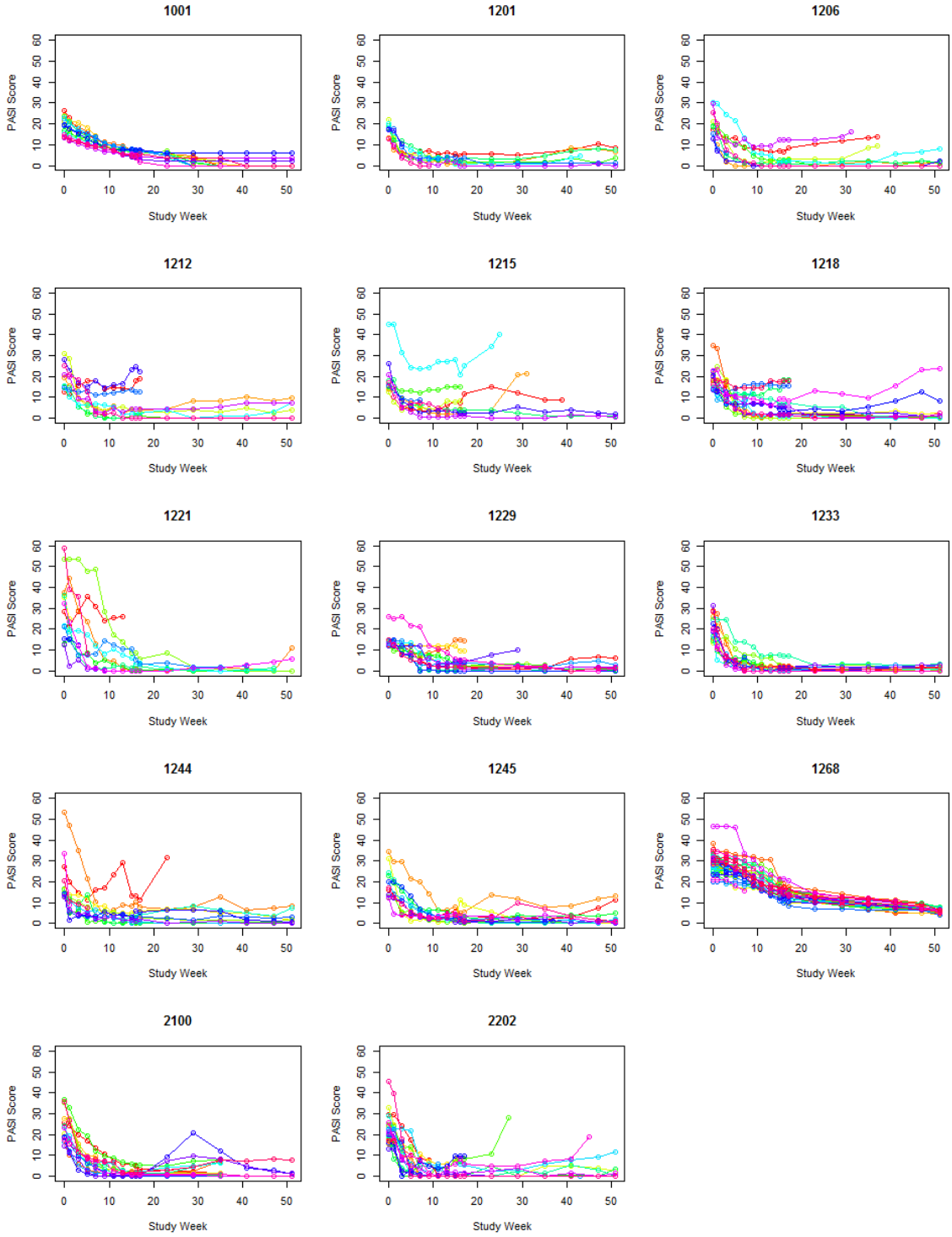
Figure 6 – Percent Change in PASI by Center/Country



B/BGR = Bulgaria, F/FRA = France, S/SVK = Slovakia, U/USA = United States
Centers with at least 10 subjects are shown individually; the remaining centers are grouped within country
Source: reviewer analysis

PASI scores were also examined for the individual subjects for the largest 14 centers (see Figure 7). Two centers stand out for having low variability and no subjects who experienced worsening PASI scores during the study. These centers are Center 1001 in Slovakia (15 subjects) and Center 1268 in the US (33 subjects). At Center 1001, the subjects' PASI scores improve smoothly during Treatment Period 1 and then either remained constant or improved during the remainder of the study. Subjects at other centers experience more variability in response and rarely have such smooth patterns. Subjects at Center 1268 also had below normal variability and none of the subjects reached a PASI score < 8 at any time during the study. Center 1268 also stands out for having no subjects on either arm achieve PASI 75 at Week 16 and all subjects achieving percent change in PASI values between -50 and -58 at Week 16. The applicant also identified Center 1268 as having unusual results, and included supplemental analyses excluding Center 1268 in the clinical study report. Because of the unusual responses at Center 1268, this center was recommended for FDA site inspection. The Office of Scientific Investigations communicated to this reviewer that all source documents at Center 1268 were destroyed in a fire accident, so the investigator cannot verify the data at the site. Centers 1218 and 1233 were additionally recommended for inspection. Most subjects in Study 301 were enrolled in US centers, and therefore inspection of Center 1001 in Slovakia was not requested. Results of the inspections at Centers 1218 and 1233 have not been communicated to this reviewer as of the day of this writing.

Figure 7 – PASI Scores at Larger Sites (N ≥ 10)



Source: reviewer analysis

Because of the unusual results at Center 1268 and the fact that source documents at this center were destroyed in a fire accident and could not be inspected, the PASI 75 results

excluding the subjects from Center 1268 (both US + EU subjects and US subjects only) are presented in Table 14. When Center 1268 is excluded from the analyses, the treatment effects ranged from 0.9% to 5.6%, but all 90% confidence intervals remain within the ±18% margins.

Table 14 – PASI 75 Results Excluding Center 1268

Analysis Set	GP2017	Humira	Difference	90% CI
Per Protocol Set				
All subjects (except Center 1268)	N=183 71.7%	N=182 70.2%	1.5%	(-6.6, 9.0)
US subjects (except Center 1268)	N=143 74.5%	N=143 68.9%	5.6%	(-4.5, 15.7)
Full Analysis Set				
All subjects (except Center 1268)	N=217 61.8%	N=215 60.9%	0.9%	(-6.6, 8.4)
US subjects (except Center 1268)	N=174 62.6%	N=171 59.1%	3.5%	(-4.9, 12.0)

Source: reviewer analysis

3.2.8 Stratification Factors

As noted in Section 3.2.2, during the blinded data review meeting the applicant identified 77 (16.6%) subjects who were misclassified in the prior therapy or weight stratum based on the weight and prior therapy information recorded in the case report form and modified the statistical analysis plan (SAP) to state that the analyses would be conducted using the 'true' factor rather than the assigned factor, as the "correct true stratification is thought to be more appropriate." Table 15 presents the analyses specified in the SAP (using source data) as well as the analyses consistent with the original protocol (using the randomization stratum). Whether the randomization stratum factors or source data factors are used in the analysis models has minimal impact on the results.

Table 15 – PASI 75 at Week 16 by Stratification Classification

Analysis Set	GP2017	Humira	Difference	90% CI
PPS (clinical database)	N=197 66.8%	N=196 65.0%	1.8%	(-6.0, 9.7)
PPS (randomization stratum)	N=197 66.8%	N=196 65.0%	1.8%	(-6.0, 9.6)
FAS (clinical database)	N=231 58.1%	N=234 55.9%	2.2%	(-5.4, 9.7)
FAS (randomization stratum)	N=231 58.0%	N=234 56.0%	2.0%	(-5.5, 9.5)

PPS= Per Protocol Set, FAS = Full Analysis Set
Source: reviewer analysis

3.2.9 Missing Data Handling for the Primary Endpoint

Missing data was not imputed for the per protocol analysis set, as subjects with missing data were excluded from the PPS. For the primary endpoint of PASI 75 in the full

analysis set, PASI 75 response missing data was imputed as non-response. The applicant did not propose any alternate methods for handling missing data as sensitivity analyses. Thus, to assess whether the handling of missing data had any impact on the results, this reviewer conducted sensitivity analyses where all the subjects with missing data on one arm were treated as failures and all subjects with missing data on the other arm were treated as successes. Missing data was balanced across the treatment arms with 15% of subjects on each arm having missing data at Week 16. Treating missing data in such an extreme manner causes large swings in the estimated treatment difference, swinging the point estimate about 15% in either direction. See Table 16. However, as noted in Section 3.2.3, the number of subjects who dropped out and their reasons for discontinuing were similar for the two arms. Although the extreme method of handling missing data causes the confidence intervals to extend beyond the similarity margins, such extreme assumptions regarding the missing data are unrealistic. The conclusions would be consistent with the primary analysis under a variety of more realistic assumptions about the missing data.

Table 16 – Sensitivity Analyses for PASI 75 Response Rates

Population	GP2017 N=231	Humira N=234	Difference	90% CI
GP2017 Missing as Failure/ Humira Missing as Success	58.1%	70.8%	-12.7%	(-19.9, 5.6)
GP2017 Missing as Success/ Humira Missing as Failure	72.8%	55.9%	16.8%	(9.7, 24.0)

Source: reviewer analysis

3.2.10 Supportive Endpoint – Investigator's Global Assessment

Success on the Investigator's Global Assessment (IGA) was a supportive endpoint. The IGA was assessed on a 5-point scale from 0 = clear to 4 = severe. Success on the IGA was defined as a score of clear or almost clear with at least two grades reduction from baseline. The applicant did not impute missing data for the IGA success endpoint. The IGA success rates are lower than the PASI 75 response rates. While the point estimates for GP2017 were slightly higher than those for Humira for PASI 75 response at Week 16, the point estimates for GP2017 for IGA success are slightly lower than those for Humira. In general, the results of the IGA success analysis are consistent with the results for the primary endpoint. See Table 17.

Table 17 – IGA Success at Week 16

Analysis Set	GP2017	Humira	Difference
PPS (Observed)	N=191	N=192	-3.4%
	49.7%	53.1%	

PPS= Per Protocol Set
Source: pg 221 of gp17-301-report-body.pdf (2/9/2017 version)

3.2.11 Assay Sensitivity and Constancy

Study 301 was a comparative clinical study of GP2017 and Humira; it did not include a placebo arm. Thus, we need to evaluate whether the study has adequate assay sensitivity (the ability to detect meaningful differences if they were to exist). Three placebo-controlled trials of Humira have been published (Gordon (2006), Saurat (2008), and Menter (2008)). Each of these studies had PASI 75 as the primary endpoint. The key design criteria and results for the published Humira studies are presented in Table 18. The Gordon study had less restrictive inclusion criteria (BSA ≥ 5, no requirement on PASI), but the Saurat and Menter studies had similar inclusion criteria to Study 301 (BSA ≥ 10, PASI ≥ 10 or 12, and sPGA/IGA ≥ Moderate). The percent improvement in PASI (79) on the Humira arm in Study 301 was generally consistent with the results from the published studies at Week 12-16 (70-81). For PASI 75, the response rate on the Humira arm in Study 301 (56%) was similar to the response rate in the Gordon study (53%), but lower than the response rate observed in the Saurat and Menter studies (71%-80%). However, because of the low placebo response rate in the previous studies, assuming assay sensitivity appears to be reasonable for Study 301.

Table 18 – Study Characteristics and Results of Published Humira Studies

	Gordon (2006)	Saurat (2008)	Menter (2008)	Study 301 (GP2017)
Selected inclusion criteria	BSA ≥ 5	BSA ≥ 10 PASI ≥ 10 sPGA ≥ Mod	BSA ≥ 10 PASI ≥ 12 sPGA ≥ Mod	BSA ≥ 10 PASI ≥ 12 IGA ≥ Mod
Region/Country	US, Canada	Europe, Canada	US, Canada	US, Europe
Baseline PASI Mean *(Humira)*	PASI = 16.7	PASI = 20.2	PASI = 19.0	PASI = 20.1
% Imp. in PASI *Humira* *Placebo*	(Week 12) 70 14	(Week 16) 81 22	(Week 12) 76 15	(Week 16) 79 --
PASI 75 *Humira* *Placebo*	(Week 12) 53% (n=50) 4% (n=52)	(Week 16) 80% (n=108) 19% (n=53)	(Week 16) 71% (n=814) 7% (n=398)	(Week 16) 56% (n= 234) --

3.3 Evaluation of Safety

Refer to the clinical review.

4 Findings in Special/Subgroup Populations

4.1 Gender, Race, Age, and Geographic Region

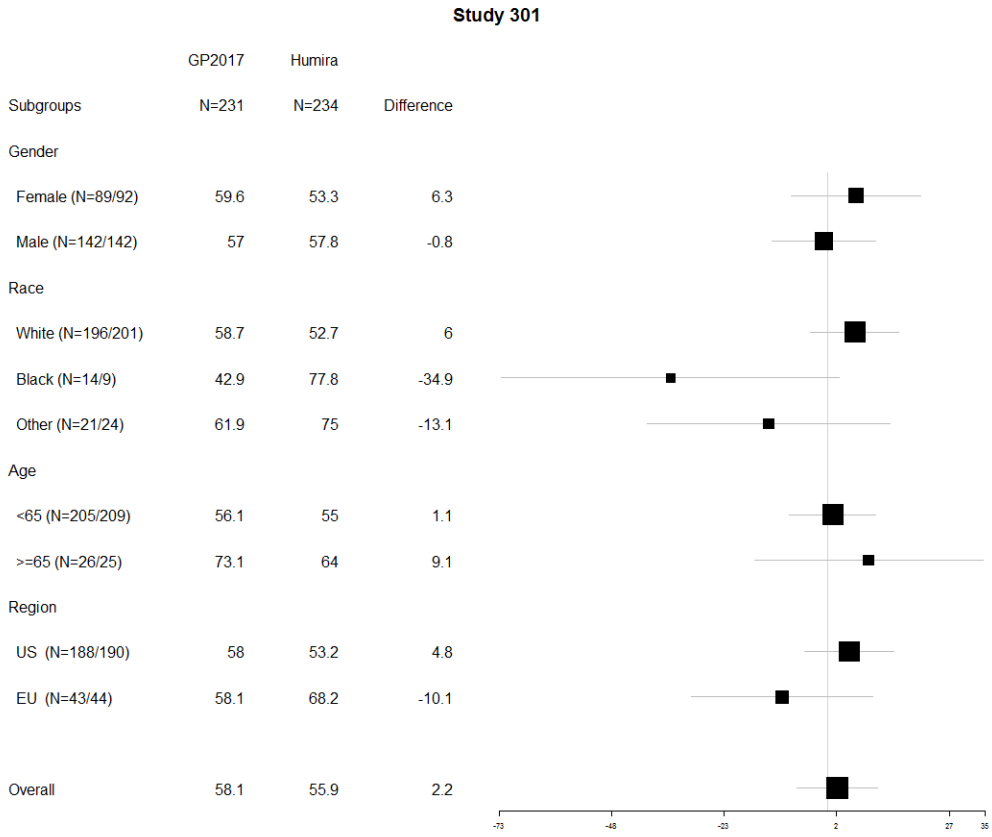
The PASI 75 response rates were consistent across gender. There were too few subjects that were not white, or age 65 and older to make meaningful comparisons. The response rates on the Humira arm differed between US and EU subjects. The impact of geographic region is further discussed in Section 3.2.7. See Table 19 and Figure 8.

Table 19 – PASI 75 Response Rates by Gender, Race, Age, and Geographic Region (FAS)

	GP2017 N=231	Humira N=234
Gender		
Female	53/89 (59.6%)	49/92 (53.3%)
Male	81/142 (57.0%)	82/142 (57.8%)
Race		
White	115/196 (58.7%)	106/201 (52.7%)
Black	6/14 (42.9%)	7/9 (77.8%)
Other	13/21 (61.9%)	18/24 (75.0%)
Age		
< 65	115/205 (56.1%)	115/209 (55.0%)
≥ 65	19/26 (73.1%)	16/25 (64.0%)
Region		
US	109/188 (58.0%)	101/190 (53.2%)
EU	25/43 (58.1%)	30/44 (68.2%)

Source: reviewer analysis

Figure 8 - PASI 75 Response Rates by Gender, Race, Age, and Geographic Region (FAS)



Source: reviewer analysis

4.2 Other Special/Subgroup Populations

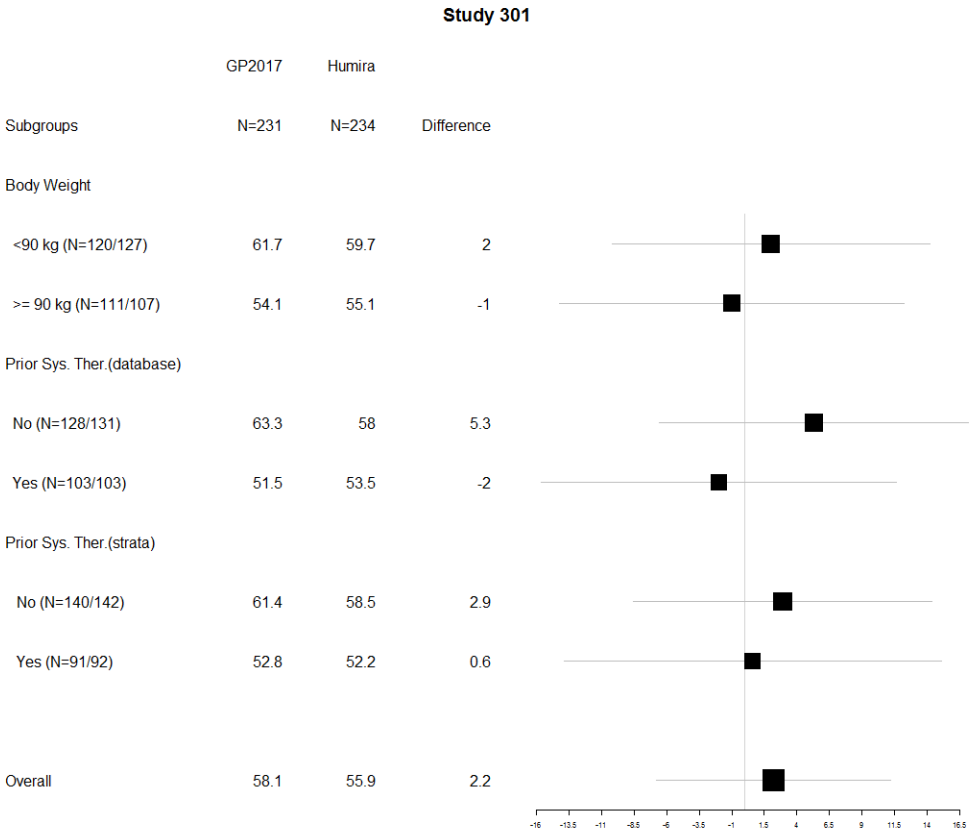
The randomization was stratified on body weight and prior systemic therapy. As noted in Section 3.2.2, during the blinded data review meeting the applicant identified 77 (16.6%) subjects who were misclassified in the prior therapy or weight stratum based on the weight and prior therapy information recorded in the case report form and modified the statistical analysis plan (SAP) to state that the analyses would be conducted using the 'true' factor rather than the assigned factor, as the "correct true stratification is thought to be more appropriate." Table 20 and Figure 9 present the PASI 75 response rates by body weight and the two classifications of prior systemic therapy (from the clinical database and from the randomization stratum). Results were generally consistent across the baseline classifications, thought the point estimates are farther apart using the clinical database classification for prior systemic therapy than using the randomization stratum.

Table 20 – PASI 75 Response Rates by Body Weight and Prior Systemic Therapy (FAS)

	GP2017 N=231	Humira N=234
Body Weight		
<90 kg	74/120 (61.7%)	72/127 (59.7%)
≥90 kg	60/111 (54.1%)	59/107 (55.1%)
Prior Systemic Therapy (clinical database)		
No	81/128 (63.3%)	76/131 (58.0%)
Yes	53/103 (51.5%)	55/103 (53.4%)
Prior Systemic Therapy (randomization strata)		
No	86/140 (61.4%)	83/142 (58.5%
Yes	48/91 (52.8%)	48/92 (51.2%)

Source: reviewer analysis

Figure 9 - PASI 75 Response Rates by Body Weight and Prior Systemic Therapy (FAS)



Source: reviewer analysis

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

In the protocol, the applicant originally proposed to provide analyses for the US regulatory submission comparing all GP2017 subjects versus only US Humira subjects. Subjects in Europe received EU-approved Humira, while subjects in the US received US-licensed Humira. For submission to European regulatory authorities, the applicant had planned to use analyses based on all collected data. FDA advised the applicant that excluding part of the Humira arm (EU study site subjects) while including all GP2017 subjects (US and EU study sites) breaks the connection to the randomization and could introduce bias, and requested that the applicant submit analyses using all the subjects (consistent with what had been planned for European submissions) along with subgroup analyses for US and EU subjects. The applicant submitted the requested analyses with the resubmission of the application.

Variability by region was observed for the PASI 75 at Week 16 response endpoint. The response rates differed more between the subjects receiving US-licensed or EU-approved Humira (53% vs. 68%), while subjects receiving GP2017 in the US or EU were similar (58% vs. 58%). Only 17% of the data was collected on European subjects. The differences between regions were smaller when comparing the percent change in PASI at Week 16 (reductions of 79% for GP2017 in the US vs. 83% in the EU, and reductions of 77% for Humira in the US vs. 87% in the EU).

Two centers were identified as having unusual results (Center 1268 in the US, and Center 1001 in Slovakia). Both centers exhibited a lower than expected variability in response, and Center 1268 was also notable for the fact that none of the 33 subjects at the center were classified as PASI 75 responders. A clinical inspection of the data at Center 1268 was requested, but the site could not be inspected because the investigator reported that all files related to the study had been lost in an accidental fire. No inspections were requested for Center 1001, because a sufficient amount of data was available from US sites.

Because of the questions of the comparability of the US and EU results, and the questions surrounding the reliability of the data collected at Center 1268 in the US, it may be appropriate to consider analyses based only on US subjects and those based on US subjects, excluding Center 1268. The results of these two subgroup analyses are consistent with the overall analyses, and the results of the subgroup analyses also meet the protocol-specified criteria of having the 90% confidence intervals fall within ±18%.

During the blinded data review meeting the applicant identified 77 (16.6%) subjects who were misclassified in the prior therapy or weight stratum based on the weight and prior therapy information recorded in the case report form and modified the statistical analysis plan (SAP) to state that the analyses would be conducted using the ‘true’ factor rather than the assigned factor. Results are similar whether the ‘true’ factors or the stratification factors are used.

5.2 Conclusions and Recommendations

Study 301 is a comparative clinical study of GP2017 versus Humira in subjects with moderate to severe psoriasis. The primary endpoint was the proportion of subjects at Week 16 with PASI 75 response. Table 21 presents the results for the overall population (all subjects), the subset of US subjects, and the subset of US subjects excluding Center 1268. In each case, the results are generally consistent and the 90% confidence intervals are contained within ±18%, the pre-specified similarity criterion. The secondary endpoints of percent change in PASI and IGA success are consistent with the results of the primary endpoint analysis.

Table 21 – PASI 75 Response Rates at Week 16

	GP2017	Humira	Difference	90% CI
PPS				
Overall	N=197 66.8%	N=196 65.0%	1.8%	(-6.0, 9.7)
US	N=157 68.0%	N=157 62.6%	5.3%	(-3.5, 14.1)
US excluding Center 1268	N=143 74.5%	N=143 68.9%	5.6%	(-4.5, 15.7)
FAS				
Overall	N=231 58.1%	N=234 55.9%	2.2%	(-5.4, 9.7)
US	N=188 57.9%	N=190 53.2%	4.7%	(-3.6, 13.1)
US excluding Center 1268	N=174 62.6%	N=171 59.1%	3.5%	(-4.9, 12.0)

Appendix

The applicant used the following procedure (excerpted from the SAP) to calculate confidence intervals for the PASI 75 endpoint using the delta method and estimates from the logistic regression model.

> With a data set of n patients, binary response vector $Y = (y_1, y_2, \ldots, y_n)'$, and, a logistic regression model assumes $\text{logit}[P(y_i = 1|x_i)] = \beta' x_i$, where $\text{logit}(p) = \ln[p/(1 - p)]$. If b denote the maximum likelihood estimate (MLE) of β, its estimated variance-covariance matrix is V and $X = (x_1, x_2, \ldots, x_n)'$ denote the covariate matrix.
>
> A new covariate matrix X_t from X by adjusting the column corresponding to treatment assignment will have to be created such that all patients are in the treated group. The vector of estimated probabilities of response to treatment, P_t, will be calculated from X_t and $b[P_t = \text{logit}-1(X_t\, b)]$. Similarly, assuming that each patient is assigned to control the above steps are repeated to get X_c and P_c. The estimated difference in proportions is $d = \Sigma_i(P_{ti} - P_{ci})/n$, where P_{ti} and P_{ci} are the ith elements of P_t and P_c respectively. At is defined as a vector with elements A_{ti} =

$P_{ti}(1 - P_{ti})$. Similarly, A_c is defined with $A_{ci} = P_{ci}(1 - P_{ci})$. The delta method is then used to estimate the standard error of the estimation:

$dt = (A_t'X_t)/n$
$dc = (A_c'X_c)/n$
$SE(d) = \sqrt{(dtVdt' + dcVdc' - 2dcVdt')}$
The confidence interval of the estimation is obtained by $d \pm Z(1-\alpha/2)SE(d)$.

Source: pg 41 of gp17-301-statistical.pdf

References

Gordon KB et al, J Am Acad Dermatol. 2006 Oct; 55(4): 598-606
Saurat JH et al, Br J Dermatol. 2008; 158: 558-66
Menter A et al, J Am Acad Dermatol. 2008; 58(1): 106-15

Signatures/Distribution List

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.
Date: 6/29/2018

Statistical Team Leader: Mohamed Alosh, Ph.D.
Date: 6/29/2018

cc:
DDDP/Marcus
DDDP/Kettl
DDDP/Chiang
DPARP/Nikolov
DPARP/Borigini
DPARP/Lee
OBIO/Patrician
DBIII/Johnson
DBIII/Alosh
DBIII/Fritsch

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

/s/

KATHLEEN S FRITSCH
06/29/2018

MOHAMED A ALOSH
06/29/2018

STATISTICAL REVIEW AND EVALUATION

Biometrics Division: VI

BLA NO.	761071
DATE RECEIVED BY THE CENTER	11/02/2017
DRUG NAME	GP 2017
DOSAGE FORM	Injection
STRENGTH	40 mg, (b) (4)
INDICATION	Biosimilar to US-Licensed Humira, all approved indications of adalimumab except for JIA in patients between 2 and 4 years, pediatric Crohn's disease, hidradenitis suppurativa, and uveitis, which are still covered by orphan exclusivity at the time of this BLA submission.
SPONSOR	Sandoz
REVIEW FINISHED	06/28/2018
STATISTICAL REVIEWER	Tianhua Wang, Ph.D.
SECONDARY REVIEWER	Meiyu Shen, Ph.D.
PROJECT MANAGER	Phuong Nina Ton

_____

Tianhua Wang, Ph.D., Mathematical Statistician, CDER/OTS/OB/DB VI
Meiyu Shen, Ph.D., Lead Mathematical Statistician, CDER/OTS/OB/DB VI

Concur: _____

Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DB VI

CC List:

Meiyu Shen, Ph.D., Lead Mathematical Statistician, CDER/OTS/OB/DB VI
Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DB VI
Lillian Patrician, CDER/OTS/OB
Phuong Nina Ton, OMPT/CDER/OND/ODEII/DPARP

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY AND RECOMMENDATION

The CMC statistical reviewer in the Office of Biostatistics analyzed the comparative results of 2 critical quality attributes: TNF−α Target Binding by SPR and Apoptosis Inhibition Bioactivity, which were recommended for equivalence testing analysis by the Office of Biotechnology. Tier 1 statistical equivalence testing was conducted using equivalence margins of $\pm 1.5\sigma_R$, where R represents the reference product US-licensed Humira or comparator product EU-sourced Humira. Fourtee independent GP 2017 drug product or substance lots, 18 US-licensed Humira lots, and 18 EU-sourced Humira lots were used for equivalence testing for TNF−α Target Binding by SPR. The results are summarized in Table 1.

Table 1: Equivalence testing results for the TNF−α Target Binding by SPR

Comparison	# of lots	Mean Difference, %	90% Confidence Interval for Mean Difference, %	Equivalence Margin, %	Pass the Equivalence Testing?
GP 2017 vs. US	(14, 18)	3.45	(-2.04, +8.92)	(-15.54, +15.54)	Yes
GP 2017 vs. EU	(14, 18)	4.17	(-0.67, +9.02)	(-12.11, +12.11)	Yes
EU vs. US	(18, 18)	-0.72	(-5.97, +4.52)	(-15.54, +15.54)	Yes

Fifteen independent drug product or substance lots, 16 US-licensed Humira lots and 21 EU-sourced Humira lots were used for equivalence testing for Apoptosis Inhibition Bioactivity. The results are summarized in Table 2.

Table 2: Equivalence testing results for the Apoptosis Inhibition Bioactivity

Comparison	# of lots	Mean Difference, %	90% Confidence Interval for Mean Difference, %	Equivalence Margin, %	Pass the Equivalence Testing?
GP 2017 vs. US	(15, 16)	-0.22	(-6.20, +5.76)	(-11.36, +11.36)	Yes
GP 2017 vs. EU	(15, 21)	-2.04	(-8.48, +4.40)	(-16.47, +16.47)	Yes
EU vs. US	(16, 21)	1.82	(-3.34, +6.98)	(-11.36, +11.36)	Yes

*The 90% confidence interval is adjusted by the sample size imbalance.

As shown in Tables 1 and 2, the results from the statistical equivalence testing of TNF−α Target Binding by SPR and Apoptosis Inhibition Bioactivity support a demonstration that the proposed

biosimilar GP 2017 is highly similar to US-licensed Humira. The results also support the analytical portion of the scientific bridge to justify the relevance of EU-sourced Humira data from the comparative clinical study.

2 INTRODUCTION

The applicant originally submitted BLA 761071 on August 25, 2016. Because some facilities were not available for inspection within an appropriate timeframe, the applicant requested withdrawal of the application on October 21, 2016. On October 30, 2017, Sandoz re-submitted to the U.S. Food and Drug Administration (FDA) a 351(k) BLA which included an analytical similarity assessment of comparing the Tier 1 quality attributes for GP 2017, US-licensed Humira, and EU-sourced Humira.

In the orginal submission, regarding the equivalence testing for TNF−α Target Binding by SPR, Sandoz used 15 GP 2017 Drug Product (DP) batches, 18 US-Licensed Humira batches (selected from 36 available US-Licensed Humira batches) and 18 EU-sourced Humira batches (selected from 43 available EU sourced Humira batches). These 15 GP 2017 DP batches were manufactured from 9 different Drug Substance (DS) batches. FDA's expectation for the analytical similarity assessment is that for a given statistical evaluation, each assay result used in that evaluation be derived from an independent DP batch or DS batch. FDA considers an "independent" batch to be a single DP batch produced from a single DS batch, or a single DS batch where no subsequent DP batch is included in the analysis. Additionally, FDA does not consider different DP batches produced from the same DS batch to be independent. When there are multiple DP batches from a single DS batch, the first manufactured DP batch is recommended to be included in the equivalence test. Besides, some manufactured DP batches from some DS batches (such as B079500 and B083248) were not included in the Tier 1 equivalence testing for TNF−α Target Binding by SPR. On April 18, 2018, FDA sent out an Information Request letter to Sandoz. The letter requested Sanodz to re-evaluate the equivalence testing for TNF−α Target Binding by SPR including results derived using only one, and preferably the first, GP2017 DP batch manufactured from a given GP2017 DS batch, and also provide their selection criteria for the batches of US-licensed Humira and the batches of EU-sourced Humira included in the Tier 1 TNF−α Target Binding by SPR equivalence testing, and the scientific justification used to support their selection criteria.

Regarding the equivalence testing for Apoptosis Inhibition Bioactivity, Sandoz used 8 GP2017 DP batches plus 2 DS batches B170052 and B291059, 16 US-Licensed Humira batches (selected from 36 available US-Licensed Humira batches) and 21 EU sourced Humira batches (selected from 43 available EU sourced Humira batches). Some manufactured DP batches from some DS batches (such as B079500, B083248, B170052) were not included in the Tier 1 equivalence testing for Apoptosis Inhibition Bioactivity. On April 18, 2018, FDA sent out an Information Request letter to Sandoz. The letter requested Sanodz to re-evaluate the equivalence testing for Apoptosis Inhibition Bioactivity including only the first GP2017 DP batch manufactured from a given GP2017 DS batch, provide their selection criteria for the batches of US-licensed Humira and the batches of EU sourced Humira included in the Tier 1 Apoptosis Inhibition Bioactivity equivalence testing, and the scientific justification used to support their selection criteria.

Sandoz reponded the above requests on May 09, 2017 as below:

a) Regarding the GP 2017 lots used for the equivalence testing for TNF−α Target Binding by SPR, Sandoz acknowledges the agency's request and herewith would like to provide the statistical re-evaluation of TNF-α target binding by SPR based on independent GP2017 DP and DS batches. As requested by the agency all non-independent GP2017 DP data sets were excluded from the assessment. Exclusion of batches was based on the manufacturing date of the respective GP2017 DP batches. To obtain a sufficient basis for equivalence testing, target binding data of five independent GP2017 DS batches was included.

b) Regarding the GP 2017 lots used for the equivalence testing for Apoptosis Inhibition Bioactivity. DS batches B079500, B083248 and B170052 in the equivalence testing of apoptosis inhibition activity. Sandoz herewith would like to clarify that DS batches B079500, B083248 and B170052 are already expired and none of these batches were measured for apoptosis inhibition activity during shelf life of the product. As data measured on expired material is not considered representative, the DP batches derived from DS batches B079500, B083248 and B170052 will not be considered for the statistical re-evaluation of apoptosis inhibition activity. Instead, Sandoz included all available data of independent DS batches generated so far (5 additional DS batches).

c) Regarding the selection criteria and scientific justification of the critera of the reference US and EU DP lots, in general Sandoz' development strategy for GP2017 was to

investigate a representative number of originator batches for every quality attribute. The number of batches included in the analytical biosimilarity evaluation depended on different factors such as availability of orthogonal or supportive assays as well as method complexity or necessity for statistical analysis. The quality attribute TNF-α binding is not only addressed by one method but by orthogonal methods as well. In the early stage of development the TNF-α neutralization reporter gene assay (RGA) was chosen as the most reliable read out for target binding due to the assay's superior reproducibility compared to other target binding methods such as surface plasmon resonance (SPR). Consequently the reporter gene assay became part of release specifications of GP 2017 and an extensive amount of EU- and US-Humira batches was measured. As orthogonal methods such as SPR and apoptosis inhibition characterize the same mechanism of action (soluble TNF-α binding and neutralization) as the reporter gene assay, measurements of a lower amount of originator batches was considered justified. Moreover, the apoptosis inhibition assay was included only at a later stage during development (in 2014). To account for reference product variability, Humira US and EU batches were regularly purchased and tested during the entire development of GP2017. As indicated by the relatively uniform distribution of the values over time, Humira batches included in the analyses were sampled randomly and are therefore considered representative for the reference product. Most investigated Humira US and EU batches were analyzed in a range spanning approximately 20 - 80 weeks before expiry, i.e. a major part of the reference product shelf life is covered by the analyses.

FDA CMC statistical reviewer and the Reviewer from Office of Biotechnology both agreed with Sandoz's response. The Agency carefully evaluated the data for the TNF−α Target Binding by SPR and Apoptosis Inhibition Bioactivity provided in the BLA submission. Our comments regarding Sandoz's equivalence testing (Tier 1 approach) is provided in Section 3, and our independent statistical equivalence testing analyses are presented in Section 4.

3 APPLICANT'S STATISTICAL EQUIVALENCE TESTING

In this submission, Sandoz conducted Tier 1 statistical equivalence testing with the margin defined as $\left(-1.5\hat{\sigma}_R,\ +1.5\hat{\sigma}_R\right)$ for TNF−α Target Binding by SPR and Apoptosis Inhibition Bioactivity. Pairwise comparisions were used for the assessment of the Tier 1 quality attributes. Similarity is demonstrated if all the two-sided 90% confidence intervals of the difference

between means for GP 2017 vs. US-licensed Humira, GP 2017 vs. EU-sourced Humira, and EU-sourced Humira vs. US-licensed Humira are within the EAC of $\left(-1.5\hat{\sigma}_R,\ +1.5\hat{\sigma}_R\right)$, where R represents the reference product US-licensed Humira or the comparator product EU-sourced Humira in each of the 3 pairwise comparisions. Sandoz presumed unequal variances for the two-sided 90% confidence interval. Sandoz's statistical approach generally followed the agent's current recommendation for Tier 1 analytical similarity assessment. FDA CMC statistical reviewer also performed the independent analysis and confirmed the results in Section 4.

4 FDA STATISTICAL ANALYSES

To evaluate analytical similarity, the Agency recommends a tiered approach. That is, product quality attributes are assigned to three tiered based on their criticality. The quality attributes with potential highest risk in product quality, efficiency, safety and PK/PD are assigned to Tier 1, in which analytical similarity is assessed by statistical equivalence test. Quality attributes with lower impact are assigned to Tier 2 and their analytical similarity is evaluated by Quality Range approach. That is, a high percentage of the biosimilar data should be covered by (Mean – X*SD, Mean + X*SD) defined by the reference product. Here, the multiplier X should be justified by the scientific knowledge. The quality attributes with the lowest risk are assigned to Tier 3 and their analytical similarity is evaluated by side-by-side comparison using graphic display. This review focuses on the equivalence testing in Tier 1.

4.1 Statistical method

Let μ_T and μ_R be respectively the population means of the quality attribute for the test product and the population mean of the quality attribute for the US-licensed Humira product. Let σ_R be the standard deviation of the quality attribute of interest for the US-licensed Humira. In order to conclude the equivalence in the quality attribute of interest between the test product and the US-licensed Humira product, we aim to reject the null hypothesis of the following null and alternative hypotheses:

$$H_0: \mu_T - \mu_R \leq \theta_1 \quad \text{or} \quad \mu_T - \mu_R \geq \theta_2$$
$$H_1: \theta_1 < \mu_T - \mu_R < \theta_2$$

.

Here $\theta_1 = -1.5\sigma_R$, $\theta_2 = 1.5\sigma_R$, θ_1 and θ_2 are equivalence margins.

We reject H_0 if 90% confidence interval for the mean difference in the quality attribute of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. In other words, we conclude that the equivalence in the quality attribute of interest between the test product and the US-licensed Humira product if 90% confidence interval for the mean difference in the quality attribute of interest falls within $(-1.5\sigma_R, 1.5\sigma_R)$. This specific equivalence margin was set as 1.5 times the standard deviation of the quality attribute for the US-licensed Humira product to ensure an adequate power for the case in which a small but sufficient number of lots are available for testing. For example, the probability of rejecting H_0 in the above two one-sided tests procedure with the equivalence margin being $(-1.5\sigma_R, 1.5\sigma_R)$ is 87% if the true mean difference is $0.125\sigma_R$ for a sample size of 10 biosimilar lots and 10 US-licensed Humira lots. 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. First we estimate σ_R by the sample variability of the US-licensed Humira product and then in the statistical analysis, θ_1 and θ_2 are treated as a constant, not a random variable.

Let X_{Tj} be the observed value of the quality attribute of interest for Batch j of the test product (the proposed biosimilar product) and X_{Rj} be the observed value of the quality attribute of interest for Batch j of the US-licensed Humira product. Since the two products are manufactured by two manufacturers, two groups are independent. $\overline{X}_i = \frac{\Sigma_{j=1}^{n_i} X_{ij}}{n_i}$ and $S_i^2 = \frac{\Sigma_{j=1}^{n_i} (X_{ij} - \overline{X}_i)}{(n_i - 1)}$, where n_i is the number of lots in the ith product, $i = T,R$.

Under the unequal variance of the test product and the US-licensed Humira product, the $(1-2\alpha) \times 100\%$ confidence interval of the mean difference in the quality attribute of interest can be calculated as:

$$\left(\overline{X}_T - \overline{X}_R - t_\alpha(v)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R}},\ \overline{X}_T - \overline{X}_R + t_\alpha(v)\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R}}\right) \qquad (1)$$

Here $t_\alpha(v)$ is the $(1-\alpha)$ quantile and v is the degrees of freedom calculated by Satterthwaite's approximation.

If $n_R > 1.5n_T$, the $(1-2\alpha)\times 100\%$ confidence interval of the mean difference in the quality attribute of interest can be calculated as:

$$\left(\overline{X}_T - \overline{X}_R - t_\alpha(v^*)\sqrt{\frac{S_T^2}{n_T}+\frac{S_R^2}{n_R^*}},\ \overline{X}_T - \overline{X}_R + t_\alpha(v^*)\sqrt{\frac{S_T^2}{n_T}+\frac{S_R^2}{n_R^*}}\right) \qquad (2)$$

Here $n_R^* = \min(n_R, 1.5n_T)$ and $v^* = \dfrac{\left(\frac{S_T^2}{n_T}+\frac{S_R^2}{n_R^*}\right)^2}{\frac{1}{n_T-1}\left(\frac{S_T^2}{n_T}\right)^2+\frac{1}{n_R-1}\left(\frac{S_R^2}{n_R^*}\right)^2}$

If the number of biosimilar lots, n_T, is 50% more than the number of reference lots, n_R, we can apply a similar approach as above with $n_T^* = \min(n_T, 1.5n_R)$ for the confidence interval calculation. In the following analyses, we use $\alpha = 0.05$.

4.2 FDA statistical equivalence testing for TNF-alpha Target Binding by SPR (%).

The TNF-alpha Target Binding data distributions of GP 2017, US-licensed Humira and EU sourced Humira are displayed in Figure 1. Fourteen batches of GP 2017, 18 batches of US-licensed Humira, and 18 batches of EU-sourced Humira are included in the TNF-alpha Target Binding dataset for statistical equivalence testing. Descriptive statistics for the TNF-alpha Target Binding data of GP 2017, US-licensed Humira, and EU-sourced Humira are listed in Table 3.

Figure 1: Scatter plots of TNF-alpha Target Binding by SPR for US-licensed Humira, GP 2017, and EU-sourced Humira.

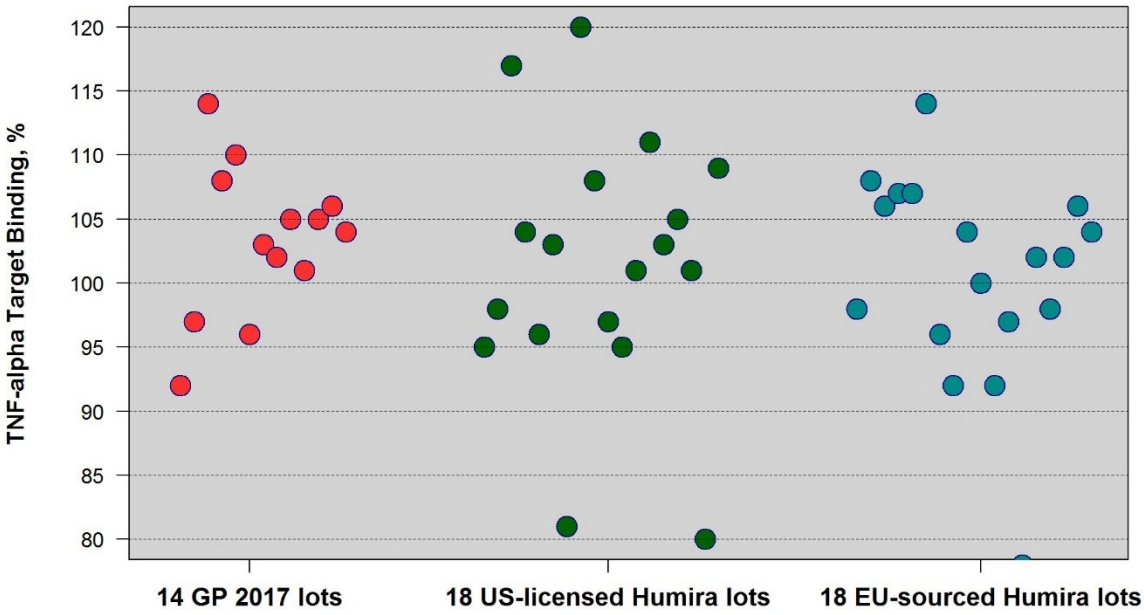


Table 3: Descriptive statistics for the TNF-alpha Target Binding data

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Min, %	Max, %
GP 2017	14	104.7	7.94	92	124
US-licensed Humira	18	101.33	10.36	80	120
EU-sourced Humira	18	100.61	8.07	78	114

Table 4 shows that the 90% confidence interval for the mean difference in the TNF-alpha Target Binding by SPR between GP 2017 and US-licensed Humira is (-2.04, +8.92)%. It falls entirely within the equivalence margin (-15.54, +15.54)%. Hence, the results of the TNF-alpha Target Binding by SPR for GP 2017 are equivalent to those for US-licensed Humira.

It also shows that the 90% confidence interval for the mean difference in TNF-alpha Target Binding by SPR between GP 2017 and EU-sourced Humira is (-0.67, +9.02)%. It falls within the equivalence margin (-12.11, +12.11)%. Therefore, the results of the TNF-alpha Target Binding by SPR for GP 2017 are equivalent to those for EU-sourced Humira.

The 90% confidence interval for the mean difference in TNF-alpha Target Binding by SPR between EU-sourced Humira and US-licensed Humira is (-5.97, +4.52)%, which falls entirely within the equivalence margin (-15.54, +15.54)%. Therefore, the results of the TNF-alpha Target Binding by SPR for EU-sourced Humira are equivalent to those for US-licensed Humira.

The statistical equivalence analyses support that the TNF-alpha Target Binding by SPR of GP 2017 is similar to that of US-licensed Humira. The results of all three pairwise comparisons support the analytical portion of the scientific bridge to justify the relevance of the data obtained from clinical studies that compared EU-sourced Humira and the GP 2017 product to support a demonstration of biosimilarity to US-licensed Humira.

Table 4: Results of equivalence testing for TNF-alpha Target Binding by SPR

Comparison	# of lots	Mean Difference, %	90% Confidence Interval for Mean Difference, %	Equivalence Margin, %	Pass the Equivalence Testing?
GP 2017 vs. US	(14, 18)	3.45	(-2.04, +8.92)	(-15.54, +15.54)	Yes
GP 2017 vs. EU	(14, 18)	4.17	(-0.67, +9.02)	(-12.11, +12.11)	Yes
EU vs. US	(18, 18)	-0.72	(-5.97, +4.52)	(-15.54, +15.54)	Yes

4.3 FDA statistical equivalence testing for Apoptosis Inhibition Activity (%)

Scatter plots of the Apoptosis Inhibition Activity for GP 2017, US-licensed Humira and EU-sourced Humira are shown in Figure 2. Fifteen batches of GP 2017, 16 batches of US-licensed Humira, and 21 batches of EU-sourced Humira are included in the Apoptosis Inhibition Activity dataset for statistical equivalence testing. Descriptive statistics for the Apoptosis Inhibition Activity data of GP 2017, US-licensed Humira, and EU-sourced Humira are listed in Table 5.

Figure 2: Scatter plots of Apoptosis Inhibition Activity for US-licensed Humira,

GP 2017, and EU-sourced Humira

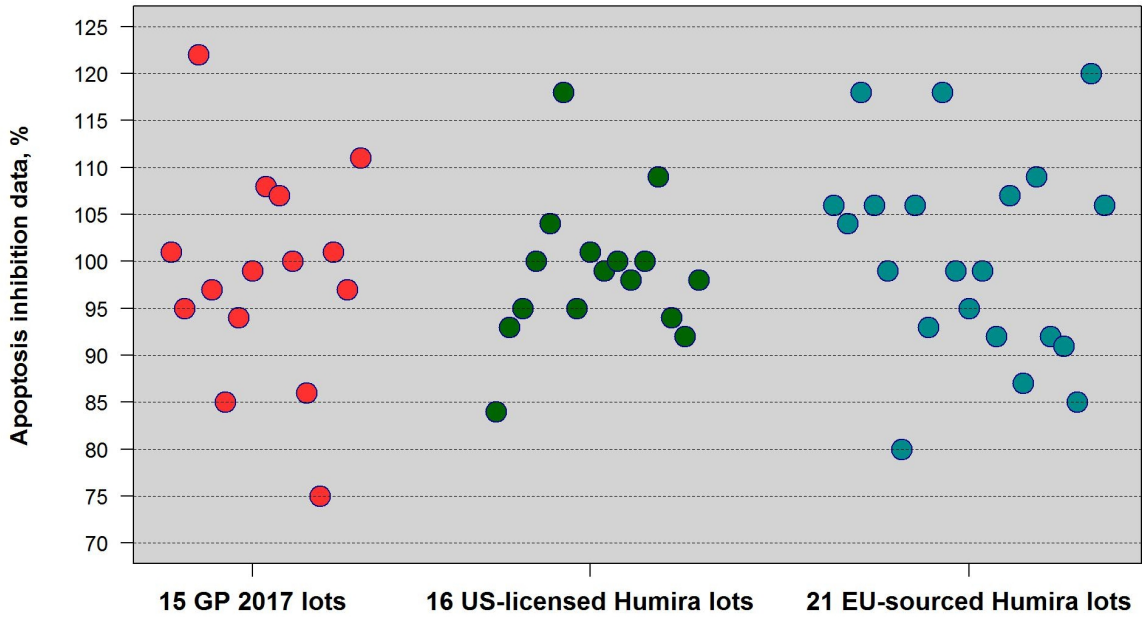


Table 5: Descriptive statistics for the Apoptosis Inhibition Activity data

Product	Number of lots	Sample mean, %	Sample standard deviation, %	Min, %	Max, %
GP 2017	15	98.53	11.38	75	122
US-licensed Humira	16	98.75	7.56	84	118
EU-sourced Humira	21	100.5	10.98	80	120

Table 6 shows that the 90% confidence interval for the mean difference in the Apoptosis Inhibition Activity between GP 2017 and US-licensed Humira is (-6.20, +5.76)%. It falls entirely within the equivalence margin (-11.36, +11.36)%. Hence the Apoptosis Inhibition Activity of GP 2017 is equivalent to the Apoptosis Inhibition Activity of US-licensed Humira.

It also shows that the 90% confidence interval for the mean difference in the Apoptosis Inhibition Activity between GP 2017 and EU-sourced Humira is (-8.48, +4.40)%. It falls within the equivalence margin (-16.47, +16.47)%. Therefore, the Apoptosis Inhibition Activity of GP 2017 is equivalent to the Apoptosis Inhibition Activity of EU-sourced Humira.

The 90% confidence interval for the mean difference in Apoptosis Inhibition Activity between EU-sourced Humira and US-licensed Humira is (-3.34, +6.98)%, which falls entirely within the equivalence margin (-11.36, +11.36)%. Therefore, the Apoptosis Inhibition Activity of EU-sourced Humira is equivalent to the Apoptosis Inhibition Activity of US-licensed Humira.

The statistical equivalence analyses support that Apoptosis Inhibition Activity of GP 2017 is similar to that of US-licensed Humira. The results of all three pairwise comparisons support the analytical portion of the scientific bridge to justify the relevance of the data obtained from clinical studies that compared EU-sourced Humira and the GP 2017 product to support a demonstration of biosimilarity to US-licensed Humira.

Table 6: Equivalence testing results for the Apoptosis Inhibition Activity

Comparison	# of lots	Mean Difference, %	90% Confidence Interval for Mean Difference, %	Equivalence Margin, %	Pass the Equivalence Testing?
GP 2017 vs. US	(15, 16)	-0.22	(-6.20, +5.76)	(-11.36, +11.36)	Yes
GP 2017 vs. EU	(15, 21)	-2.04	(-8.48, +4.40)	(-16.47, +16.47)	Yes
EU vs. US	(16, 21)	1.82	(-3.34, +6.98)	(-11.36, +11.36)	Yes

5 CONCLUSION AND RECOMMENDATION

The statistical equivalence analyses shown above regarding TNF-alpha Target Bindin by SPR and the Apoptosis Inhibition Activity of GP 2017 support a demonstration that GP 2017 is highly similar to US-licensed Humira. They also support the analytical portion of the scientific bridge to justify the relevance of EU-sourced Humira data from the comparative clinical study.

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

/s/

TIANHUA WANG
06/29/2018

MEIYU SHEN
06/29/2018

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

BLA/SN:	BLA 761071
Supplement #:	Original
Drug Name:	GP2017
Indication(s):	Biosimilar to Humira (adalimumab)
Applicant:	Sandoz
Date(s):	Submitted: 10/30/2017, BsUFA: 10/30/2018
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Rebecca Rothwell, Ph.D.
Concurring Reviewers:	Gregory Levin, Ph.D.
Additional Statistical Review Team:	Kathleen Fritsch, Ph.D./Mohamed Alosh, Ph.D. (DBIII)
Medical Division:	DDDP/DPARP
Clinical Team:	Mark Borigini, M.D./Nikolay Nikolov, M.D. (DPARP)
	Gary Chiang, M.D./David Kettl, M.D. (DDDP)
Project Manager:	Nina Ton

Keywords:

Biosimilar, active control

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The applicant, Sandoz GmbH, submitted Biologics License Application (BLA) 761071 to demonstrate biosimilarity of GP2017 to the reference product, Humira (adalimumab), based on the totality of evidence including analytical, nonclinical, and clinical data. The clinical program includes four pharmacokinetic (PK) studies and one clinical comparative efficacy and safety study in patients with plaque psoriasis. The primary clinical efficacy statistical review was performed by Dr. Kathleen Fritsch from the Division of Biometrics III, supporting the Division of Dermatology and Dental Products. The study population in the comparative clinical study included 98 subjects with a diagnosis of psoriatic arthritis (PsA). Within this subgroup, the change from baseline in the Health Assessment Questionnaire - Disability Index (HAQ-DI) was used to assess the patient's level of physical functional ability and activity restriction. This supportive statistical review focused on the analyses for this endpoint in the subset of subjects with concomitant PsA. Though sample sizes were small within this subset and assessments were limited to descriptive statistics, the effects of GP2017 compared to Humira on physical function as measured by HAQ-DI were similar.

2 INTRODUCTION

2.1 Background

Sandoz GmbH (Sandoz) submitted BLA 761071 under section 351(k) of the Public Health Service (PHS) Act to support GP2017 as a biosimilar to US-licensed Humira (adalimumab). Adalimumab is a monoclonal antibody approved for the treatment of rheumatoid arthritis (RA), polyarticular juvenile idiopathic arthritis (JIA), PsA, ankylosing spondylitis (AS), Crohn's disease, ulcerative colitis (UC), plaque psoriasis (PsO), pediatric Crohn's disease, hidradenitis suppurativa, and uveitis. The indications for GP2017 sought by Sandoz include all approved indications of adalimumab except for JIA in patients between 2 and 4 years, pediatric Crohn's disease, hidradenitis suppurativa, and uveitis, which are still covered by orphan exclusivity at the time of this BLA submission.

The recommended dose for RA, PsA, and AS is 40 mg every other week (q2w). Some patients with RA not receiving methotrexate may benefit from increasing the frequency to 40 mg every week. For use in PsO or uveitis, the recommended dose is 80 mg initial dose, followed by 40 mg q2w starting one week after initial dose. The recommended dose in Crohn's disease and UC is 160 mg on day 1, followed by 80 mg on day 15, and a maintenance dose of 40 mg q2w beginning on day 29. Patients with UC should only continue to receive adalimumab if they have shown evidence of clinical remission by eight weeks. For use in hidradenitis suppurativa, the recommended dose is an initial dose of 160 mg, a second dose of 80 mg two weeks later, and third and subsequent doses of 40 mg every week. Dosing in JIA and pediatric Crohn's disease is dependent on weight.

Approval of this biosimilar application will be based on the totality of evidence including analytical, nonclinical, and clinical data. The clinical program includes four PK studies and one clinical confirmatory efficacy and safety study in plaque psoriasis. Dr. Kathleen Fritsch from the Division of Biometrics III, supporting the Division of Dermatology and Dental Products

(DDDP), performed the primary statistical review of efficacy and safety. This supportive statistical review focuses on the analyses for the subset of subjects with concomitant PsA.

2.2 History of Product Development

The clinical development program for GP2017 was introduced to DDDP and the Division of Pulmonary, Allergy, and Rheumatology Products (DPARP) under IND 115732. The study design and study populations of the pivotal PK study GP17-101 and the pivotal comparative efficacy and safety study GP17-301 were discussed with the FDA during a type B pre-IND meeting on January 14, 2013. At this meeting, the FDA agreed that the applicant's proposal to demonstrate therapeutic comparability between GP2017 and the reference product Humira in terms of safety and efficacy by performing one comparative clinical study in adult patients with moderate-to-severe chronic plaque psoriasis was acceptable. However, the applicant would need to provide sufficient scientific justification for extrapolating clinical data to support a determination of biosimilarity for each condition of use. Such scientific justification should address the mechanism of action in each condition of use, pharmacokinetics of the product, differences in expected toxicities in each condition of use and patient population, and any other factor that may affect differences in safety or efficacy across conditions of use and patient populations.

The applicant originally submitted BLA 761071 on August 25, 2016. Because some facilities were not available for inspection within an appropriate timeframe, the applicant requested withdrawal of the application on October 21, 2016. In the withdrawal acknowledgement letter, FDA requested that in future re-submissions, the applicant address the statistical deficiencies identified during the preliminary review. In particular, this communication requested submission of all necessary information to conduct the primary and key secondary analyses and supplementary analyses for the primary and key secondary endpoints using the full population (both European and United States combined for both treatment arms).

2.3 Specific Studies Reviewed

The applicant submitted results from one completed clinical study to support similarity in clinical efficacy. GP17-301 was a multicenter, randomized, double-blind, active-controlled, comparative study in patients with moderate to severe chronic plaque-type psoriasis (Table 1). The study enrolled and randomized 465 subjects, all of whom received at least one dose of study medication. The study was conducted in 73 centers in the US (378 subjects, 81.3%), Bulgaria (31 subjects, 6.7%), France (25 subjects, 5.4%), and Slovakia (31 subjects, 6.7%). The study population in GP17-301 included 98 subjects with a diagnosis of PsA. Within this subgroup, HAQ-DI was used to assess the patient's level of physical functional ability and activity restriction. This review focuses only on the analyses for this subset of subjects.

Table 1 Summary of Trials to be Assessed in the Statistical Review

Trial ID	Design	Treatment/ Sample Size	Endpoint (Analysis)
GP17-301	MC, R, DB, AC comparative clinical trial	GP2017: 231 Humira: 234 (US:190/EU: 44) With PsA: GP2017: 52 Humira: 46 (US: 39/EU: 7)	Primary: PASI75 response at week 16 (logistic regression) In PsA subjects: Mean change from baseline to week 11 and week 16 in HAQ-DI (summary statistics only)

Abbreviations: MC: multi-center, R: randomized, DB: double-blind, PG: parallel-group,AC: active-controlled, PsA: psoriatic arthritis, PASI75: 75% improvement from baseline in Psoriasis Area and Severity Index

2.4 Data Sources

Data were submitted by the applicant to the CDER electronic data room in SAS transport format. Protocols, correspondence, data listings, program code, and study reports were accessed under the network path \\cdsesub1\evsprod\bla761071\0005\m5\.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The submitted datasets were of acceptable quality and were adequately documented. We could reproduce the results of all relevant analyses for this review. The clinical inspection of one study site, 1268, located in Florida, found that all source documents were destroyed in a fire accident such that the data at the site could not be confirmed. The primary statistical reviewer recommended removing the data from this site in our analyses. No subjects from the PsA subset were enrolled from this site and, therefore, the results presented here were not affected.

3.2 Study Design

GP17-301 was a multicenter, randomized, double-blind, active-controlled, comparative study in 465 patients with moderate to severe chronic plaque-type PsO. The primary objective of this study was to demonstrate equivalent efficacy of GP2017 and Humira in patients with moderate to severe chronic plaque-type psoriasis with respect to Psoriasis Area Severity Index 75% response (PASI75) rate are week 16. A secondary objective and the focus of this review, was the comparison of the functional ability, as measured by the HAQ-DI, in patients with a medical history of PsA.

The study enrolled subjects age 18 or older with moderate to severe psoriasis, defined as a PASI score of 12 or greater, an investigator's global assessment (IGA) score of three or greater, and body surface area (BSA) by plaque-type PsO of 10% or greater. Subjects had chronic plaque-type PsO for at least six months prior to randomization and had previously received phototherapy or systemic psoriasis therapy or were candidates for such therapies. Subjects could not have previous exposure to adalimumab and subjects could not continue use of prohibited psoriasis treatments, including topical or systemic corticosteroid or UV-therapy.

Subjects were randomized 1:1 to GP2017 or Humira (US-licensed Humira at US sites and EU-approved Humira at EU sites), stratified by body weight (greater than 90 kg or less than 90 kg), region (US or EU), and prior systemic therapy (none or any). Patients received a loading dose of 80 mg, followed by 40 mg doses every other week on each arm. The submission included a bridging argument from EU-Humira to US-Humira based on three-way analytical and three-way PK data. Therefore, data from both Humira treatment groups were pooled together for the efficacy and safety analyses. The study included two treatment periods (day 1 to week 17 and week 17 to 35) and an extension period (week 35 to 51). At week 16, subjects who achieved 50% response on the PASI score (PASI50 response) were re-randomized 2:1 to either continue their originally randomized treatment or to receive three alternating treatments with GP2017 or Humira for six consecutive weeks for weeks 17 to 35 (Figure 1). After week 35, subjects received their originally randomized treatment.

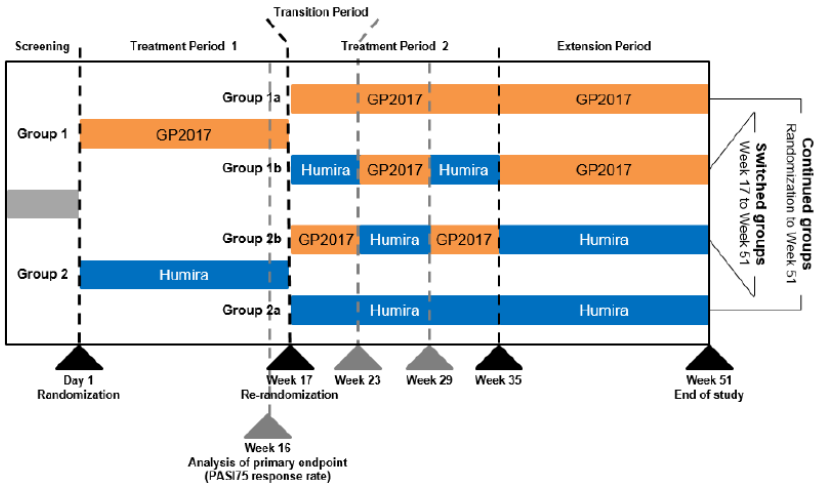


Figure 1 Study Design for GP17-301
Abbreviations: PSAI= psoriasis area severity index
Source: Clinical Study Report, p. 106

3.3 Endpoints

The primary endpoint in this study was the proportion of patients achieving a PASI75 response at week 16, evaluated using a pre-specified similarity margin of ± 18%. The study also included additional secondary and supportive endpoints based on the PASI and IGA scales. These endpoints were evaluated by Dr. Fritsch as primary statistical reviewer of efficacy and safety. Within the 98 subjects with a diagnosis of PsA, HAQ-DI assessments at week 11 and week 16 were used to assess the impact of chronic disease on the patient's level of functional ability and activity restriction. HAQ-DI assessments were also taken after re-randomization at week 17, 23, 29, 35, 41, 47, and 51. At each time point, HAQ-DI was evaluated as percent and actual change from baseline at that time point.

3.4 Statistical Methodologies

The secondary endpoint of change from baseline in HAQ-DI in treatment period 1 (baseline through week 17) was evaluated using both the per protocol analysis set (PPS), as well as the full analysis set (FAS), each restricted to the subset of patients with a diagnosis of PsA. The FAS

consisted of all randomized patients in the PsA subset to whom study treatment had been assigned. The PPS was a subset of the FAS and consisted of patients who completed the study up to week 16 and had with no major protocol deviations up to and including week 16. Discontinuations due to lack of efficacy before week 16 were included in the PPS provided they received at least four weeks (two doses) of treatment. Patients who completed treatment period 1 but failed to meet the entry criteria for treatment period 2 and therefore discontinued the study, were included in the PPS.

Endpoints from treatment period 2 were evaluated using the treatment period 2 and extension period per full analysis set (TP2+EP FAS) and the treatment period 2 and extension period per protocol analysis set (TP2+EP PPS). The TP2+EP FAS consisted of all patients who were re-randomized into treatment period 2 in this PsA subset. The TP2+EP PPS was a subset of patients of the TP2+EP FAS and consisted of all patients of the TP2+EP FAS who completed treatment period 1, treatment period 2, and the post-treatment follow-up period and had no major protocol deviations or additional exclusionary criteria (re-randomization criteria and treatment compliance criteria) during the study. Discontinuations due to lack of efficacy during treatment period 2 and the extension period were included in the TP2+EP PPS.

Summary statistics for both the mean actual and percent change from baseline for the total and eight category HAQ-DI scores were provided by visit. No imputation of missing data was performed in the analysis of HAQ-DI. While HAQ-DI is considered an exploratory endpoint that does not require hypothesis testing, to provide further information on the comparative clinical efficacy, we also calculated the treatment differences in mean actual change from baseline and corresponding 95% confidence intervals for visits in treatment period 1.

3.5 Evaluation of Efficacy in PsA Subjects

3.5.1 Patient Disposition, Demographic and Baseline Characteristics

A total of 98 subjects with a diagnosis of PsA were randomized in GP17-301. Approximately 85% of these subjects completed week 17 (treatment period 1) (Table 2). The percentages of discontinuations were relatively even between GP2017 and Humira, with the most frequent reason for discontinuation defined as “withdrawal by subject.” Two subjects discontinued due to adverse events on the Humira arm compared to zero subjects on the GP2017 arm. Of the patients re-randomized for treatment period 2 and the extension period, discontinuations were also evenly distributed across treatment arms (Appendix Table 1).

Table 2 Patient Disposition for Treatment Period 1, PsA Subset

	GP2017	Humira	Total
Randomized	52	46	98
Per Protocol Set	44 (84.6%)	40 (87%)	84 (85.7%)
Completed week 17	44 (84.6%)	39 (84.8%)	83 (84.7%)
Discontinued from study by week 17	8 (15.4%)	7 (15.2%)	15 (15.3%)
Adverse event	0 (0.0%)	2 (4.3%)	2 (2.0%)
Lack of efficacy	2 (3.8%)	1 (2.2%)	3 (3.1%)
Lost to follow-up	1 (1.9%)	0 (0.0%)	1 (1.0%)
Protocol deviation	1 (1.9%)	1 (2.2%)	2 (2.0%)
Withdrawal by subject	4 (7.7%)	3 (6.5%)	7 (7.1%)

Abbreviations: PsA= psoriatic arthritis
Source: Reviewer Program: Disposition.R

Demographic and baseline disease characteristics of the PsA patients were generally comparable between the randomized GP2017 and Humira arms (Table 3). There were balanced numbers of males and females, with slightly more females on the Humira arm (54%) and slightly less females on the GP2017 arm (48%). Most subjects were Caucasian (82%), with a mean age of 49.7 years old. Weight and BMI were largely similar between arms with a mean weight of 93.6 kg and mean BMI of 33.0 kg/m^2 across arms.

Table 3 Baseline Characteristics for PsA Subset

		GP2017	Humira	Total
Randomized		52	46	98
Sex	F	25 (48.1%)	25 (54.3%)	50 (51%)
	M	27 (51.9%)	21 (45.7%)	48 (49%)
Race	Black	1 (1.9%)	3 (6.5%)	4 (4.1%)
	Caucasian	44 (84.6%)	36 (78.3%)	80 (81.6%)
	Native American	3 (5.8%)	3 (6.5%)	6 (6.1%)
	Other	3 (5.8%)	4 (8.7%)	7 (7.1%)
	Unknown	1 (1.9%)	0 (0.0%)	1 (1.0%)
Ethnicity	Hispanic or Latino	11 (21.2%)	10 (21.7%)	21 (21.4%)
	Mixed Ethnicity	8 (15.4%)	1 (2.2%)	9 (9.2%)
	South Asian	1 (1.9%)	0 (0.0%)	1 (1%)
	Southeast Asian	0 (0.0%)	1 (2.2%)	1 (1%)
	Unknown	5 (9.6%)	9 (19.6%)	14 (14.3%)
	Not Reported	5 (9.6%)	7 (15.2%)	12 (12.2%)
	Other	22 (42.3%)	18 (39.1%)	40 (40.8%)
Age		48.4 (13.9)	51.1 (14.4)	49.7 (14.1)
Weight (kg)		92.1 (27.4)	95.2 (26.9)	93.6 (27.1)
BMI kg/m^2		32.5 (8.7)	33.6 (8.7)	33.0 (8.7)

Abbreviations: PsA=psoriatic arthritis, BMI=Body Mass Index
Cell contents are means (standard deviation) for continuous characteristics or frequency (percentage) for categorical characteristics.
Source: Reviewer, Program: BaselineCharacteristics.R

3.5.2 Results and Conclusions

The mean change from baseline in HAQ-DI scores in the PPS during treatment period 1 was similar between the GP2017 and Humira arms, though at each time point, the mean change was slightly higher in the GP2017 arm (Table 4, Figure 2). At week 16, the mean change from

baseline was -0.33 for the GP2017-randomized subjects, compared to -0.13 in the Humira-randomized subjects (difference: -0.20, 95% CI: (-0.44, 0.03)). A similar trend was observed in the analysis using the FAS (Appendix Table 3, Appendix Figure 1). Furthermore, each of the eight categories followed a similar trajectory, with slightly larger changes observed in the GP2017 in each category except for eating (Appendix Figure 2).

Table 4 Mean Change from Baseline in HAQ-DI Scores in the Per Protocol Set During Treatment Period 1

Visit	Treatment	Change from Baseline in HAQ-DI Score		Difference (95% CI)
		n	Mean (sd)	
Week 11	GP2017	44	-0.30 (0.48)	-0.05 (-0.24, 0.14)
	Humira	38	-0.24 (0.39)	
Week 16	GP2017	42	-0.33 (0.53)	-0.20 (-0.44, 0.03)
	Humira	37	-0.13 (0.52)	
Week 17	GP2017	39	-0.31 (0.51)	-0.12 (-0.34, 0.10)
	Humira	37	-0.19 (0.44)	

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, sd=standard deviation, CI= confidence interval
n= number of subjects included in summary statistic
95% confidence interval calculated based on normal approximation
Source: Reviewer, Program: HAQDI.R

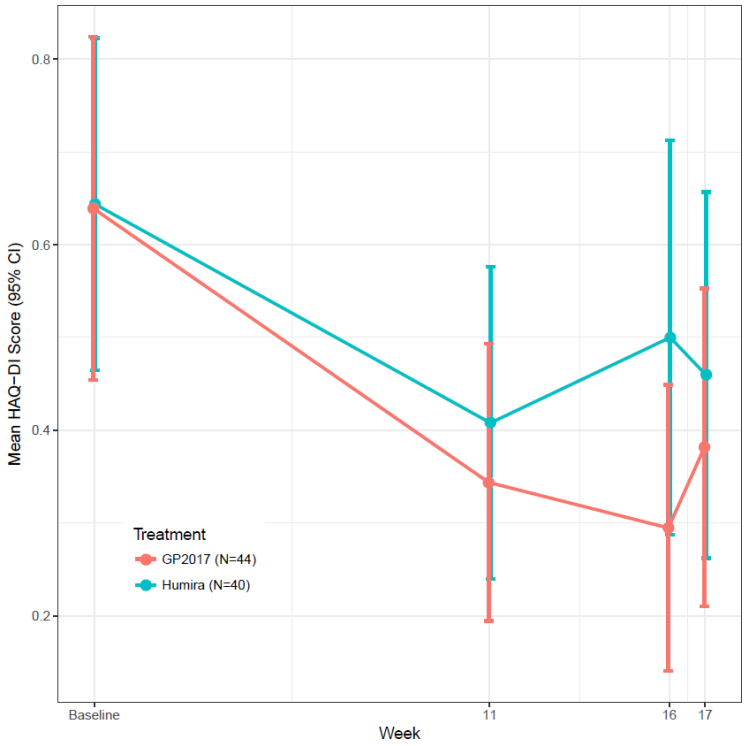


Figure 2 Mean HAQ-DI Scores During Treatment Period 1 (Per Protocol Set)
Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, N=number of randomized subjects in per protocol set with psoriatic arthritis, CI=confidence interval
Error bars correspond to 95% confidence intervals for mean HAQ-DI value at time point.
Source: Reviewer program: HAQDI.R

The TP2+EP PPS and TP2+EP FAS categorized re-randomized subjects as "continued GP2017", "GP2017 to Humira", "continued Humira" and "Humira to GP2017". While the number of subjects in each of these categories was small for the PsA population (9 to 20 subjects), the mean changes from baseline during treatment period 2 and the extension period (week 17 to 51) in TP2+EP PPS were similar across arms (Figure 3). A similar trend was observed in the analysis using TP2+EP FAS (Appendix Figure 3).

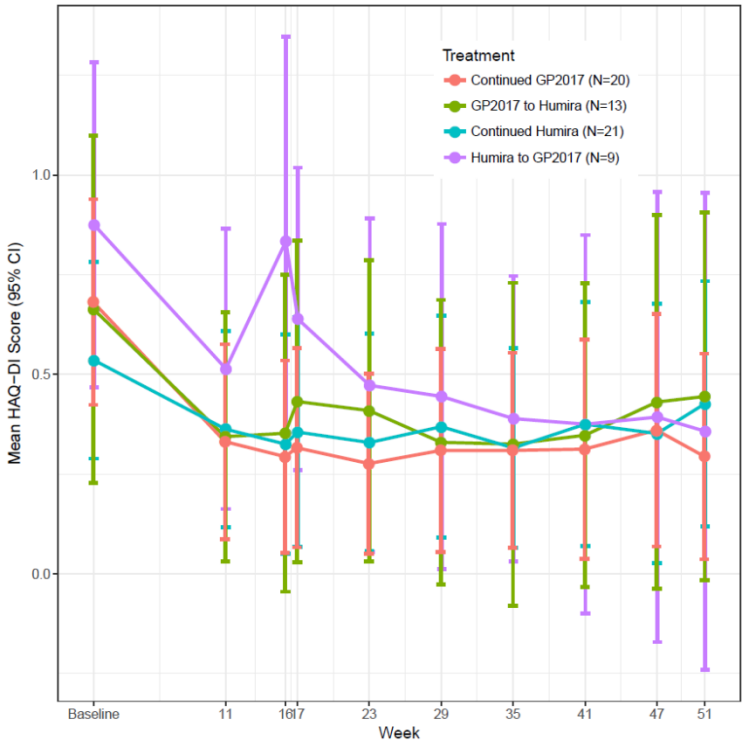


Figure 3 Mean HAQ-DI Scores Through Week 51 (TP2+EP PPS)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, TP2+EP PPS=treatment period 2 and extension period per protocol set, N=number of randomized subjects in TP2+EP PPS with psoriatic arthritis, CI=confidence interval
Error bars correspond to 95% confidence intervals for mean HAQ-DI value at time point.
Source: Reviewer program: HAQDI_All.R

4 SUMMARY AND CONCLUSIONS

4.1 Statistical Issues

During the review of the PsA subset in this application and the HAQ-DI endpoint, there were no major statistical issues. One minor issue was in the applicant's presentation of mean actual and percent change from baseline in HAQ-DI. In these analyses, the applicant excluded subjects who had a score of 0 at baseline. While this may be necessary for the calculation of percent change from baseline, which would require dividing by zero, the actual change should be calculated for all subjects in the analysis set, regardless of baseline value. We calculated the summary statistics for this broader population for mean actual change from baseline, and therefore, there were small discrepancies compared to the applicant's presented results. This change did not impact the overall conclusions about the HAQ-DI endpoint.

4.2 Conclusions and Recommendations

The comparison of change from baseline in HAQ-DI in the subset of subjects with a diagnosis of PsA randomized to GP2017 and Humira was included as a secondary objective of GP17-301. Though sample sizes were small within this subset and, due to the exploratory nature of the endpoint, assessments were limited to descriptive statistics, the effects of GP2017 and Humira on physical function as measured by HAQ-DI appear to be similar. This review did not raise any concerns in terms of the comparison of functional ability by HAQ-DI in patients with PsA treated with GP2017 versus Humira.

5 APPENDIX

Appendix Table 1 Patient Disposition for Treatment Period 2 and Extension Period, PsA Subset

	Continued GP2017	GP2017 to Humira	Continued Humira	Humira to GP2017	Total
Re-Randomized	24	15	25	13	77
Per Protocol Set	20 (83.3%)	13 (86.7%)	21 (84%)	9 (69.2%)	63 (81.8%)
Completed week 35	20 (83.3%)	13 (86.7%)	21 (84%)	13 (100%)	67 (87%)
Discontinued from study by week 35	4 (16.7%)	2 (13.3%)	4 (16%)	0 (0%)	10 (13%)
Adverse event	1 (4.2%)	0 (0%)	3 (12%)	0 (0%)	4 (5.2%)
Death	1 (4.2%)	0 (0%)	0 (0%)	0 (0%)	1 (1.3%)
Lack of efficacy	1 (4.2%)	2 (13.3%)	1 (4%)	0 (0%)	4 (5.2%)
Withdrawal by subject	1 (4.2%)	0 (0%)	0 (0%)	0 (0%)	1 (1.3%)
Entered Extension Period	18	12	20	12	62
Per Protocol Set	17 (94.4%)	10 (83.3%)	19 (95%)	8 (66.7%)	54 (87.1%)
Completed week 51	17 (94.4%)	11 (91.7%)	17 (85%)	10 (83.3%)	55 (88.7%)
Discontinued from study by week 35	1 (5.6%)	1 (8.3%)	3 (15%)	2 (16.7%)	7 (11.3%)
Adverse event	0 (0%)	0 (0%)	0 (0%)	1 (8.3%)	1 (1.6%)
Lack of efficacy	0 (0%)	0 (0%)	2 (10%)	1 (8.3%)	3 (4.8%)
New therapy for study indication	1 (5.6%)	1 (8.3%)	0 (0%)	0 (0%)	2 (3.2%)
Withdrawal by subject	0 (0%)	0 (0%)	1 (5%)	0 (0%)	1 (1.6%)

Abbreviations: PsA= psoriatic arthritis
Source: Reviewer, Program: Disposition_Extension.R

Appendix Table 2 Descriptive Statistics of HAQ-DI Scores in the Per Protocol Set During Treatment Period 1

Visit	Treatment	HAQ-DI Score		Percent Change from Baseline in HAQ-DI Score	
		n	Mean (sd)	n	Mean (sd)
Baseline	GP2017	44	0.64 (0.61)	--	--
	Humira	40	0.64 (0.56)	--	--
Week 11	GP2017	44	0.34 (0.49)	32	-49.72 (51.73)
	Humira	38	0.41 (0.51)	31	-41.81 (61.01)
Week 16	GP2017	42	0.29 (0.49)	30	-57.89 (51.60)
	Humira	37	0.50 (0.64)	29	-29.50 (101.48)
Week 17	GP2017	39	0.38 (0.53)	29	-47.02 (52.05)
	Humira	37	0.46 (0.59)	30	-31.14 (98.61)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, sd=standard deviation
n= number of subjects included in summary statistic
Subject's with score of 0 at baseline excluded from percent change analysis.
Source: Reviewer, Program: HAQDI.R, corresponds to Clinical Study Report Table 14.2-8.1-II

Appendix Table 3 Mean Change from Baseline in HAQ-DI Scores in the Full Analysis Set During Treatment Period 1

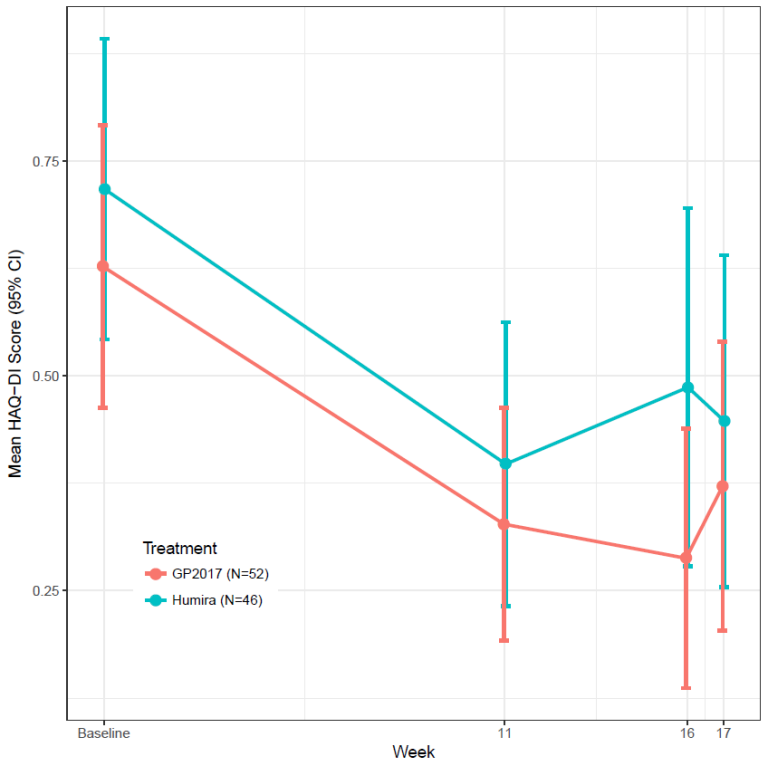
Visit	Treatment	Change from Baseline in HAQ-DI Score		Difference (95% CI)
		n	Mean (sd)	
Week 11	GP2017	50	-0.28 (0.48)	-0.02 (-0.21, 0.16)
	Humira	39	-0.26 (0.40)	
Week 16	GP2017	43	-0.33 (0.53)	-0.17 (-0.41, 0.06)
	Humira	38	-0.15 (0.53)	
Week 17	GP2017	40	-0.30 (0.51)	-0.10 (-0.31, 0.12)
	Humira	38	-0.20 (0.45)	

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, sd=standard deviation, CI= confidence interval
n= number of subjects included in summary statistic
95% CI calculated based on normal approximation
Source: Reviewer, Program: HAQDI.R

Appendix Table 4 Descriptive Statistics of HAQ-DI Scores in the Full Analysis Set During Treatment Period 1

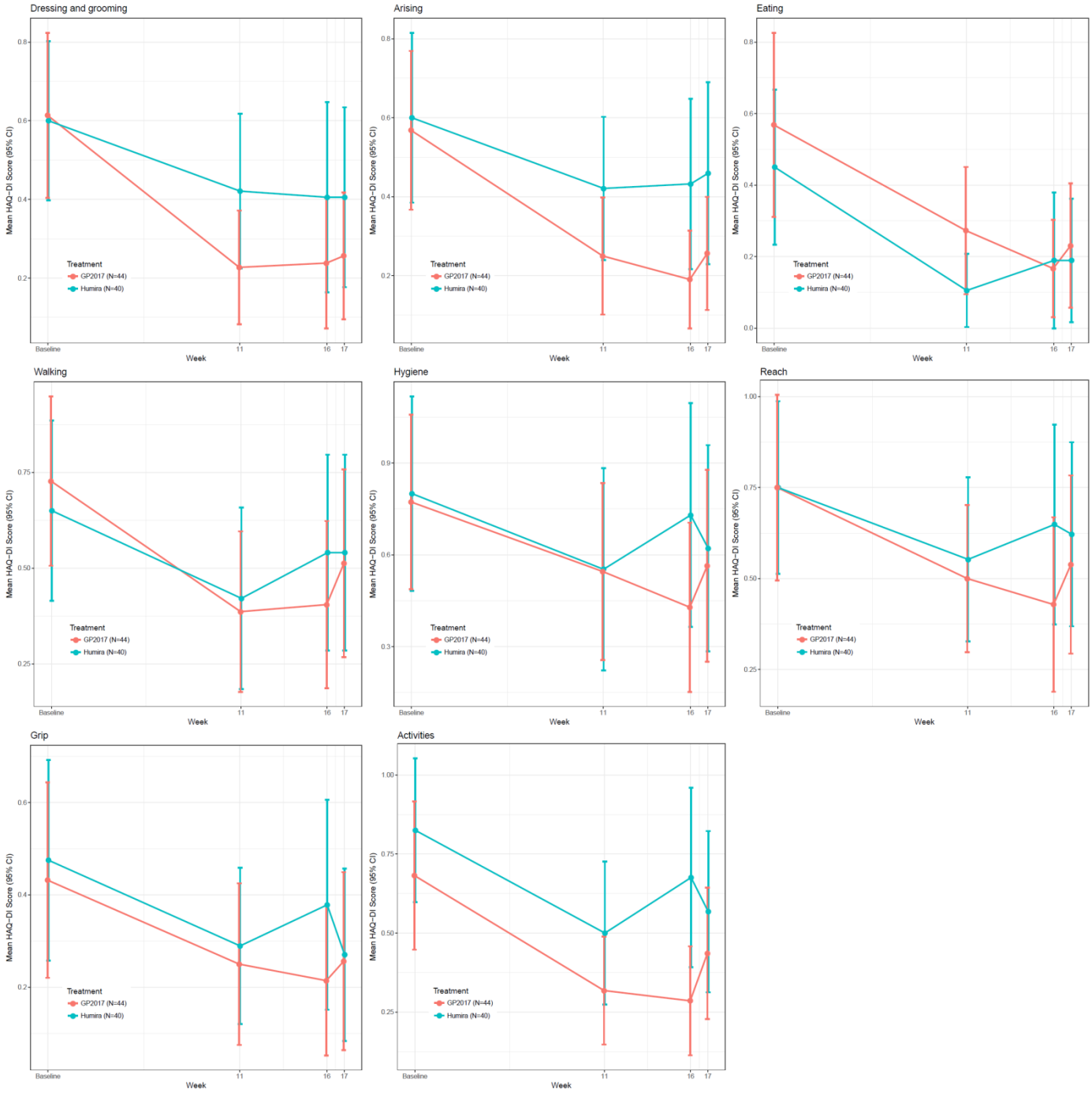
Visit	Treatment	HAQ-DI Score		Percent Change from Baseline in HAQ-DI Score	
		n	Mean (sd)	n	Mean (sd)
Baseline	GP2017	52	0.63 (0.59)	--	--
	Humira	46	0.72 (0.59)	--	--
Week 11	GP2017	50	0.33 (0.48)	36	-52.53 (51.47)
	Humira	39	0.40 (0.51)	32	-43.63 (60.90)
Week 16	GP2017	43	0.29 (0.49)	30	-57.89 (51.60)
	Humira	38	0.49 (0.63)	30	-31.85 (100.54)
Week 17	GP2017	40	0.37 (0.53)	29	-47.02 (52.05)
	Humira	38	0.45 (0.59)	31	-33.36 (97.73)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, sd=standard deviation
n= number of subjects included in summary statistic
Subject's with score of 0 at baseline excluded from percent change analysis.
Source: Reviewer, Program: HAQDI.R, corresponds to Clinical Study Report Table 14.2-8.2-II



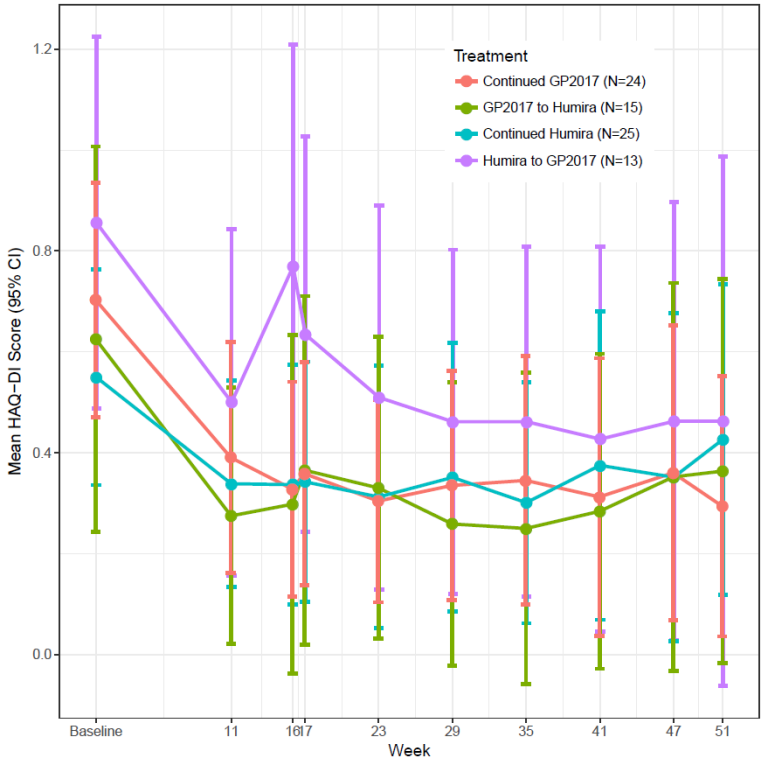
Appendix Figure 1 Mean HAQ-DI Scores During Treatment Period 1 (Full Analysis Set)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, N=number of randomized subjects in full analysis set with psoriatic arthritis, CI=confidence intervals
Error bars correspond to 95% confidence intervals for mean HAQ-DI value at time point.
Source: Reviewer, Program: HAQDI.R



Appendix Figure 2 HAQ-DI Scores by Category During Treatment Period 1 (Per Protocol Set)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, N=number of randomized subjects in full analysis set with psoriatic arthritis, CI=confidence intervals
Error bars correspond to 95% confidence intervals for mean HAQ-DI value at time point.
Source: Reviewer, Program: HAQDIComponents.R



Appendix Figure 3 Mean HAQ-DI Scores Through Week 51 (TP2+EP FAS)

Abbreviations: HAQ-DI= Health Assessment Questionnaire Disability Index, TP2+EP FAS=treatment period 2 and extension period full analysis set, N=number of randomized subjects in TP2+EP FAS with psoriatic arthritis, CI=confidence interval
Error bars correspond to 95% confidence intervals for mean HAQ-DI value at time point.
Source: reviewer program: HAQDI_All.R

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

/s/

REBECCA S ROTHWELL
06/27/2018

GREGORY P LEVIN
06/27/2018