Corrona Statistical Analysis Plan: QD Analysis

Study A3921359 Compare Effectiveness of Tofacitinib 11 mg QD to Tofacitinib 5 mg BID

Prepared for: Pfizer

Prepared by: PPD , PhD

28 April 2017_rev03/20/2018

Statistical Analysis Plan

Title: Efficacy analysis of Tofacitinib 11 mg QD compared with Tofacitinib 5 mg BID initiators

Background:

Tofacitinib (tofa) 11 mg QD was approved in February 2016. It is hypothesized that this formulation will behave similarly to the Tofacitinib 5 mg BID dosage in terms of efficacy.

Overall Aim:

Compare efficacy of Tofacitinib 11 mg QD initiators to Tofacitinib 5 mg BID initiators

Overall Aim Population:

Initiators of tofa 11 mg QD
Initiators of tofa 5 mg BID

Primary Outcome for Overall Aim:

Minimally clinically important difference (MCID) improvement as defined by difference in CDAI from initiation to 6 month visit (dichotomous variable). MCID is dependent on baseline disease activity (at initiation); MCID≥2 if CDAI at initiation of Low CDAI $\leq$ 10, MCID≥ 6 if CDAI at initiation Moderate of 10<CDAI$\leq$22 and MCID≥11 if CDAI at initiation High of >22.

Secondary Outcomes for Overall Aim:

- Change from baseline to 6 month follow-up CDAI (as a continuous variable)
- Achievement of Remission (CDAI≤2.8) at 6 month follow-up (among all)
- Achievement of Remission (CDAI≤2.8) at 6 month follow-up (among those that were not in remission at initiation)
- Achievement of Low Disease Activity (LDA) (CDAI≤10) at 6 month follow-up (among all)
- Achievement of Low Disease Activity (LDA) (CDAI≤10) at 6 month follow-up (among those that were not in remission or LDA at initiation)
- Achievement of improvement from baseline to 6 month follow-up in modified HAQ (mHAQ) of 0.25 (as a dichotomous variable)
- Change from baseline to 6 month follow-up modified HAQ (mHAQ as a continuous variable)
- Achievement of improvement from baseline to 6 month follow-up in HAQ of 0.22 (as a dichotomous variable)
- Change from baseline to 6 month follow-up HAQ (HAQ as a continuous variable)
- Change from baseline to 6 month follow-up modified DAS (mDAS as a continuous variable)
- Change from baseline to 6 month follow-up DAS ESR (as a continuous variable)
- Change from baseline to 6 month follow-up DAS CRP (as a continuous variable)
- Change from baseline to 6 month follow-up patient pain (as a continuous variable)
- Change from baseline to 6 month follow-up patient fatigue (as a continuous variable)
- Change from baseline to 6 month follow-up EQ-5D index (as a continuous variable)

- Modified ACR20, ACR50, ACR70

Please see Appendix for more details on the secondary outcomes.

Patients that switched to an agent other than tofa during the 6 months of follow-up will be considered for the continuous outcomes at the time of switch (rather than at 6 months).

In **Tables 2 and 3**, for the dichotomous primary and secondary outcomes, patients that discontinued or switched to an agent other than tofa will be considered as non-responders (e.g. did not achieve remission for patients that discontinued or switched prior to their 6 month visit).

Analysis for Overall Aim:

Consider the demographic, patient and disease characteristics of the 11 mg QD initiators to the 5 mg BID initiators. Ideally, the two groups of patients would be from a similar timeframe, have similar disease activity, be of the same line of therapy and generally be similar aside from the fact that they have initiated tofa 11mg QD vs. tofa 5 mg BID. The goal is to create similar cohorts of patients with equal likelihood of response. The main objective behind this study design is to ensure that the two groups are similar prior to doing any comparisons of outcomes. Then, if differences in outcome are seen, we can be confident that they truly represent differences between the two groups and are not due to inherent dissimilarities between the two cohorts at baseline.

Patient and disease characteristics at initiation for the two groups will be reported in **Table 1** with standardized differences calculated. Standardized differences provide a measure of clinically important differences even if there is no statistically significant difference between the two groups. A standardized difference that is less than 0.1 is commonly taken to indicate a negligible difference between treatment groups. The most thorough approach to use if there is imbalance based on standardized differences is propensity score matching to account for these imbalances. If there are only a few differences, then accounting for these imbalances in the adjusted outcome analysis may be possible; however, propensity score (PS) matched would be a more robust approach and will be described here. For the initial look at the data, as number of patients is projected to be small, we would not recommend considering propensity score matching at that time. Rather, we would recommend adjusting for a small number of critical variables if needed. We expect to use PS matching in the final analysis and therefore it is described below.

CCI [redacted]

Based on factors in **Table 1** that differ between the two groups (e.g. standardized difference > 0.1) and clinical input on a priori factors to include, a propensity score will be estimated for propensity of being in the 11 mg QD group. A small number of factors will be considered in the propensity score model; these may include age, gender, disease activity, line of therapy, etc. The primary purpose of the propensity score approach is to determine the patients with "common support" – those patients falling in an overlapping distribution of the propensity score. Patients falling outside the common support would have no comparable patients in the other group and would not be able to be used for comparative analyses.

Here we describe the propensity score matching approach.

Propensity Score Matching Approach: To balance differences in baseline patient and disease characteristics among the 5 mg BID and 11 mg QD groups, we will use matched propensity scores.

> ***Logistic regression model approach to obtain propensity scores:*** We will use the logistic regression approach (outcome is 11 mg QD therapy) including variables such as the following covariates: age (continuous), gender (male is reference group), baseline CDAI (continuous), line of therapy (categorical), duration of disease (categorical), as appropriate. As mentioned above, we plan to examine all the Table 1 variables, but expect that only a select group will be considered in the propensity score model. The resulting logit model will be of the form:

$$\ln\left({}^{p}/_{1-p}\right) = m11mgQD$$
$$= \beta_0 + \beta_1 * age + \beta_2 * gender + \beta_3 * CDAI + \beta_4 * line\ of\ therapy + \cdots$$

where ***p*** is the probability of 11mg QD Tofacitinib therapy.

A patient's resulting propensity score from the model is the patients predicted 11 mg QD Tofacitinib (*p11mgQD*), given by

$$\boldsymbol{p11mgQD = \frac{e^{m11mgQD}}{1 + e^{m11mgQD}}}$$

After the propensity score model is fit, **Table 1** will again be constructed with this propensity score matched sample and presented in **Table 1a** to ensure that there are no longer any statistically significant differences between the two groups (11 mg QD and 5 mg BID therapy). For the propensity score matching, it is proposed that a caliper of 0.01 will be used, but the caliper may be increased if the resulting sample size available from the propensity score matching is prohibitive.

In addition to considering the standardized differences between the individual characteristics of the two groups, we will also perform additional diagnostics plot to assess whether the overall distributions of the two groups are well-matched and under common support. These would include considering kernel-density plots and box plots to compare the two distributions. If they were not found to be sufficiently overlapping, we would consider revisiting the propensity score model to include additional covariates that might not have been originally included.

Once the propensity score matched sample is identified, the primary outcome of MCID of CDAI will be considered in **Table 2**. Table 2a will present the distribution of the switching status of patients. Secondary outcomes will be considered in **Table 3**.

Population for Overall Aim:

For the primary analysis comparing QD to BID patients, we will consider those patients who initiated tofa 5 mg BID on or after February 2016 and those that initiated tofa 11 mg QD after it became available in February 2016. Although the sample size will be smaller than if we consider all tofa 5 mg BID patients, the addition of those patients may introduce noise since there is the possibility that they could differ from the more recent initiators. Therefore, we will consider the cleaner cohort of all patients initiating tofa on or after February 2016.

In order to adequately measure the outcome, the initiators will also be required to have a follow-up visit at 6 months (+/- 3 months) after tofacitinib initiation.

The data cut to be used will be the most recent available.

With regards to dosing, patients can be categorized as follows:
QD dosing: once-daily tofacitinib 11mg
BID dosing: 5mg twice-daily
Other: other dosages including 10 mg and will also include not specified

Those in the "other" category will not be considered in these analyses.

The first initial look at the data has indicated that the tofa 5 mg BID group differs on some characteristics from the tofa 11 mg QD group. To account for these differences, in a second look at the data, we propose to consider models adjusted for no more than 10 covariates. This will provide a robust analysis, while preserving sample size. The full propensity score analysis will be done in the final look at the data when the sample size is large enough (as per preliminary power calculations on page 23).

In this second look with adjusted models, we propose to consider covariates with standardized differences > 0.2 as well as age, gender, CDAI and duration of RA (e.g. gender, age, duration of RA, current statin use, prior cDMARD use, prior non-TNF use, current concomitant therapy, CDAI and EQ-5D) in adjusted models. When considering change in outcome, we will also include the baseline level of that covariate in the model (e.g. for change in CDAI, consider baseline CDAI). For the continuous outcomes, mean differences within subject adjusted for covariates will be assessed using linear regression models (least squares means and 95% CIs). For binary outcomes, we will consider multivariable logistic regression models (ORs and 95% CIs). For the primary outcome, the adjusted OR (95% CI) will give the odds of having MCID improvement in the 11 mg QD initiator group as compared to the odds in the 5 mg BID initiator group after adjusting for covariates; the 5 mg BID initiator group will be the reference group. The interpretation will be similar for the other binary secondary outcomes. Please refer to **Table 2_adj** for the primary outcome and **Table 3_adj** for the secondary outcomes.

Secondary Descriptive Aim:

Consider patients that switched from tofa 5 mg BID to tofa 11 mg QD ("switchers"). Patients will be considered to have switched from 5 mg BID to 11 mg QD if they initiated 5 mg BID and subsequently changed their dose to 11 mg QD. These patients would necessarily have to be considered after February 2016.

Secondary Descriptive Analysis:

Consider the patients that switched from tofa 5 mg BID to tofa 11 mg QD as a descriptive cohort and assess efficacy in these patients. Due to sample size limitations (about 50 switchers available through 4/1/2017 data), we would not perform a formal comparison to a matched cohort, but would generally compare the efficacy rates obtained in the group of switchers to what would be expected of tofa initiators overall based on historical data CCI [redacted].

Table 4 will present patient and disease characteristics of the "switchers". **Table 5** will present outcomes at 6 months after the time of switch from 5 mg BID to 11 mg QD. In addition to the outcomes presented in **Table 3**, an outcome measuring whether patients' CDAI values worsen by at least one category from time of switch to 6 months post-switch will be added to **Table 5**.

Statistical analyses will be performed using SAS Version 9.4 (SAS Institute, Cary, NC) and STATA Version 15 (StataCorp, LP, College Station, TX).

Overall Definitions:

Tofacitinib Initiation: A tofacitinib initiation is defined as the first ever use of tofacitinib (reported by the rheumatologist). At enrollment in Corrona a detailed medication history is captured for all RA-related medications.

Line of therapy: For the purposes of these analyses, line of therapy will be aligned with the definition provided for previous Pfizer reporting.

1st line: Naïve to conventional synthetic DMARD-IR (csDMARDs) and all Biologic DMARDs
2nd line: Conventional synthetic DMARD-IR, but biologic naïve
3rd line: Biologic DMARD-IR
4th + line: ≥2 Biologic DMARD-IR

Race and Ethnicity: patient self-reported race and ethnicity from the Corrona patient enrollment form.

History of Malignancy: all except non melanoma skin cancer.

History of CVD: Includes history of hypertension (HTN), coronary artery disease, cardiac revascularization procedure (CABG, stent, angioplasty), ventricular arrhythmia, cardiac arrest, myocardial infarction (MI), acute coronary syndrome, congestive heart failure (CHF), unstable angina, stroke, TIA, other CV events, deep vein thrombosis, peripheral artery disease, pulmonary embolism, peripheral arterial thrombosis, urgent peripheral revascularization and peripheral ischemia/gangrene, hyperlipidemia or carotid artery disease (CAD).

Shell Tables

Figure 1. Flowchart of study population.

N=XX aged 18+ patients with RA
N=XX Tofacitinib initiations
N=XX 11 mg QD initiations
N=XX 5 mg BID initiations
N=XX switchers from 5 mg BID to 11 mg QD

Table 1. Patient Demographic and Clinical Characteristics for Tofacitinib Initiators by Dose (11 mg QD and 5 mg BID)

At time of Tofacitinib initiation	11 mg QD N=	5 mg BID N=	Standardized difference
Patient Characteristics			
Female: n (%)			
Age: Mean ± SD			
Duration of RA: Mean ± SD			
Race: n (%)			
White			
Black			
Asian			
Other			
Race: n(%)			
White			
Non-white			
Weight: Mean ± SD			
BMI			
Normal/underweight: n %)			
Overweight: n (%)			
Obese: n (%)			
Comorbid conditions: n (%)			
Hx of Hypertension			
Hx of Diabetes			
Hx of Malignancy*			
Hx of CV disease**			
Current use of statins			
RA Treatment History			
Prior number of cDMARDs: Median (IQR)			
Prior cDMARD use			
0			
1			
2			
3+			

Table 1 continued

At time of Tofacitinib initiation	11 mg QD N=	5 mg BID N=	Standardized difference
Prior TNF Use			
0			
1			
2+			
Prior non-TNF Use			
0			
1			
2+			
Prior Biologic Use			
0			
1			
2			
3+			
Current concomitant therapy			
Monotherapy			
Combo w/MTX alone			
Combo w/other cDMARD (not MTX)			
Combo w/MTX + other cDMARD			
Current medication at the time of initiation			
Prednisone use: n (%)			
Prednisone dose among users			
≤10 mg			
>10 mg			
Disease Activity: Mean ± SD			
Tender Joint Count (28)			
Swollen Joint Count (28)			
Physician Global Assessment (0-100)			
Patient Global Assessment (0-100)			
CDAI			
Disease activity category			
Remission (CDAI <= 2.8)			
Low (2.8 < CDAI <= 10)			
Moderate (10 < CDAI <= 22)			
Severe (CDAI > 22)			
Patient Pain (0-100)			
Patient reported fatigue (0-100)*			
EQ-5D (0-1)			
mDAS: Mean ± SD			
DAS ESR: Mean ± SD			
DAS CRP: Mean ± SD			

Table 1 continued

At time of Tofacitinib initiation	11 mg QD N=	5 mg BID N=	Standardized difference
Line of therapy			
1st			
2nd			
3rd			
4th			

SD = Standard Deviation
IQR = Interquartile Range: 25th percentile, 75th percentile

Standardized differences rather than p-values can be more informative regarding the actual size of the difference between the two groups; hence, it is suggested that they be used rather than p-values to inform which variables are to be included in the propensity score model.

*The scale ranges from 0-100 with 0 being fatigue is no problem to 100 fatigue is a major problem.

Table 1a. Patient Demographic and Clinical Characteristics for Tofacitinib Initiators by Dose (5 mg BID and 11 mg QD) after Propensity Score Matching.

Repeat Table 1 in the Propensity Score Matched cohort. *Table only generated for the final analysis.*

Table 2: Primary outcome of 11 mg QD vs. 5 mg BID Initiators
Table generated "after propensity score matching" only for the final analysis.

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value*
Primary Outcome	**Response Rate**	**Response Rate**	
MCID Improvement in difference in CDAI from initiation to 6 month visit			

*The p-value will be reported from the t test of the mean difference between the 11 mg QD and 5 mg BID initiators according to the primary outcome.

Table 2a: Other outcomes of 11 mg QD vs. 5 mg BID Initiators
Table generated "after propensity score matching" only for the final analysis.

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value*
Other outcomes	**Response Rate**	**Response Rate**	
Switching status**			
% Remained on drug at 6 month visit			
% Discontinued initiated drug and did not start another biologic at/before 6 month visit			
% Switched initiated drug to another biologic at/before 6 month visit			

*P-value will be from a 2 degree of freedom chi-squared test investigating the relationship between the two treatment arms and the 3 switching status categories.
**The "discontinued initiated drug" category will be defined as discontinuation of tofacitinib without a new biologic started. The "switched initiated drug" category was defined as discontinuation of tofacitinib with a new biologic started.

Table 2b: Additional information on MCID improvement

Table generated "after propensity score matching" only for the final analysis.

	11 mg QD Initiators N=	5 mg BID Initiators N=
Measure	**Response Rate**	**Response Rate**
Percent achieving MCID ≥ 2 if CDAI at initiation is low (CDAI ≤ 10)		
Percent achieving MCID ≥ 6 if CDAI at initiation is moderate (10 < CDAI ≤ 22)		
Percent achieving MCID ≥ 11 if CDAI at initiation is high (> 22)		

Table 3: Secondary outcomes of 11 mg QD vs. 5 mg BID
Table generated "after propensity score matching" only for the final analysis.

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value*
Continuous outcomes Change from baseline CDAI to 6 month CDAI Change from baseline mHAQ to 6 month mHAQ Change from baseline HAQ to 6 month HAQ Change from baseline mDAS to 6 month mDAS Change from baseline DAS ESR to 6 month DAS ESR Change from baseline DAS CRP to 6 month DAS CRP Change from baseline pain VAS (0-100) to 6 month pain VAS Change from baseline fatigue (0 to 100) to 6 month fatigue Change from baseline EQ-5D (0-1) to 6 month EQ-5D	**Mean ± SE**	**Mean ± SE**	
Binary outcomes measured at 6 months Achievement of LDA† Achievement of LDA‡ Achievement of Remission§ Achievement of Remission‖ Improvement in mHAQ≥0.25¶ Improvement in HAQ≥0.22¶	**Response Rate**	**Response Rate**	
mACR20**			

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value*
mACR50[**¶]			
mACR70[**]			

*The p-value will be reported from the t test of the mean difference between the 11 mg QD and 5 mg BID initiators according to each continuous secondary outcome. Binary outcome p-values will be calculated from chi-squared tests comparing the two groups.

[†] Low Disease Activity defined as 6-month CDAI≤10 among those with baseline CDAI>10

[‡]Low Disease Activity defined as 6-month CDAI≤10 for all patients

[§]Remission defined as 6-month CDAI≤2.8 among those with baseline CDAI>2.8

[‖]Remission defined as 6-month CDAI≤2.8 for all patients

[¶]Improvement in mHAQ will be calculated only for patients with baseline mHAQ≥0.25, and improvement in HAQ will be calculated only for patients with baseline HAQ≥0.22

[**]modified ACR: based on 2 out of 4 measures (not using ESR or CRP)

CCI [redacted]

Table 2_adj: Primary outcome of 11 mg QD vs. 5 mg BID Initiators

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value**
Primary Outcome MCID Improvement in difference in CDAI from initiation to 6 month visit	**OR (95% CI)***	**Ref**	

*The adjusted OR (95% CI) will give the odds of having MCID improvement in the 11 mg QD initiator group as compared to the odds in the 5 mg BID initiator group after adjusting for covariates; the 5 mg BID initiator group will be the reference group.
**The p-value will be reported from a logistic regression model considering the 11 mg QD vs. 5 mg BID group adjusting for covariates that differ between the 11 mg QD and 5 mg BID initiators according to the primary outcome.

Table 3_adj: Secondary outcomes of 11 mg QD vs. 5 mg BID[*]

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value[†]
Continuous outcomes	**Least squares mean (95% CI)***	**Least squares mean (95% CI)***	
Change from baseline CDAI to 6 month CDAI			
Change from baseline mHAQ to 6 month mHAQ			
Change from baseline HAQ to 6 month HAQ			
Change from baseline mDAS to 6 month mDAS			
Change from baseline DAS ESR to 6 month DAS ESR			
Change from baseline DAS CRP to 6 month DAS CRP			

	11 mg QD Initiators N=	5 mg BID Initiators N=	p-value[†]
Change from baseline pain VAS (0-100) to 6 month pain VAS Change from baseline fatigue (0 to 100) to 6 month fatigue Change from baseline EQ-5D (0-1) to 6 month EQ-5D			
Binary outcomes measured at 6 months Achievement of LDA[‡] Achievement of LDA[§] Achievement of Remission[‖] Achievement of Remission[¶] Improvement in mHAQ≥0.25[**] Improvement in HAQ≥0.22[**] mACR20[††] mACR50[††] mACR70[††]	**OR (95% CI)****	**1 Ref**	

[*]For continuous outcomes, the least squares means (95% CI) will be reported from a linear regression model adjusted for covariates considering the 11 mg QD and 5 mg BID initiator groups. For binary outcomes, the adjusted OR (95% CI) will give the odds of achievement of the outcome in the 11 mg QD initiator group as compared to the odds in the 5 mg BID initiator group after adjusting for covariates; the 5 mg BID initiator group will be the reference group.

[†]For continuous outcomes, the p-value will be reported from a linear regression model for the outcome adjusted for covariates considering the mean difference between the 11 mg QD and 5 mg BID initiators. For binary outcomes, the p-value will be calculated from a logistic regression model considering the 11 mg QD vs. 5 mg BID groups adjusting for covariates.

[‡]Low Disease Activity defined as 6-month CDAI≤10 among those with baseline CDAI>10

[§]Low Disease Activity defined as 6-month CDAI≤10 for all patients

[‖]Remission defined as 6-month CDAI≤2.8 among those with baseline CDAI>2.8

[¶]Remission defined as 6-month CDAI≤2.8 for all patients

[**]Improvement in mHAQ will be calculated only for patients with baseline mHAQ≥0.25 and improvement in HAQ will be calculated only for patients with baseline HAQ≥0.22

[††]modified ACR: based on 2 out of 4 measures (not using ESR or CRP)

Table 2_adj will present the primary outcome and Table 3_adj will present the secondary outcomes for the second look incorporating the adjusted analysis.

Table 3_switchers. Number of initiators of 11 mg QD (from the cohort of 11 mg QD initiations considered in Tables 1-3) that switch* to 5 mg BID by their 6 month follow-up visit.

*Switch defined as discontinued 11 mg QD and initiated or restarted 5 mg BID.

Table 4. Patient Demographic and Clinical Characteristics for the cohort of patients that switched from Tofacitinib 5 mg BID to Tofacitinib 11 mg QD

At time of Tofacitinib 5 mg initiation	**Patients that switched from Tofacitinib 5 mg BID to Tofacitinib 11 mg QD N=**
Patient Characteristics	
Female: n (%)	
Age: Mean ± SD	
Duration of RA: Mean ± SD	
Race: n (%)	
White	
Black	
Asian	
Other	
Missing	
Race: N (%)	
White	
Non-white	
Weight: Mean ± SD	
BMI	
Normal/underweight: n %)	
Overweight: n (%)	
Obese: n (%)	
Missing	
Comorbid conditions: n (%)	
Hx of Hypertension	
Hx of Diabetes	
Hx of Malignancy*	
Hx of CV disease**	
Current use of statins	
RA Treatment History	
Prior number of cDMARDs: Median (IQR)	
Prior cDMARD Use: N (%)	
0	
1	
2	
3 +	
Prior TNF Use	
0	
1	
2+	

Table 4 continued

At time of Tofacitinib 5 mg initiation	**Patients that switched from Tofacitinib 5 mg BID to Tofacitinib 11 mg QD N=**
Prior non-TNF Use	
0	
1	
2+	
Prior Biologic Use	
0	
1	
2	
3+	
Current concomitant therapy	
Monotherapy	
Combo w/MTX alone	
Combo w/other cDMARD (not MTX)	
Combo w/MTX + other cDMARD	
Current medication at the time of initiation	
Prednisone use: n (%)	
Prednisone dose among users	
≤10 mg	
>10 mg	
Disease Activity: Mean ± SD	
Tender Joint Count (28)	
Swollen Joint Count (28)	
Physician Global Assessment (0-100)	
Patient Global Assessment (0-100)	
CDAI (0-76): Mean ± SD	
Disease Activity Category: N (%)	
Remission (CDAI ≤ 2.8)	
Low (2.8 < CDAI ≤ 10)	
Moderate (10 < CDAI ≤ 22)	
Severe (22 < CDAI)	
Patient Pain (0-100)	
Patient reported fatigue (0-100)	
EQ-5D (0-1)	
mDAS: Mean ± SD	
DAS ESR: Mean ± SD	
DAS CRP: Mean ± SD	

Table 4 continued

At time of Tofacitinib 5 mg initiation	**Patients that switched from Tofacitinib 5 mg BID to Tofacitinib 11 mg QD N=**
Line of Therapy [d]: N (%)	
1st	
2nd	
3rd	
4th	

SD = Standard Deviation
IQR = Interquartile Range: 25th percentile, 75th percentile

Table 5. Outcomes at 6 months from switch of the cohort of patients that switched from Tofacitinib 5 mg BID to Tofacitinib 11 mg QD

	Switchers
Total Patients	N
Continuous Outcomes: Change from Switch to 6 Months [a]	Mean ± SD
CDAI (N_{miss}=)	
mHAQ (N_{miss}=)	
HAQ (N_{miss}=)	
mDAS (N_{miss}=)	
DAS ESR (N_{miss}=)	
DAS CRP (N_{miss}=)	
Patient Pain (N_{miss}=)	
Patient Fatigue [b] (N_{miss}=)	
EQ-5D (N_{miss}=)	
Binary Outcomes: Measured at 6 Months	N (%)
Worsening CDAI [c]	
Achievement of Low Disease Activity [d]	
Achievement of Low Disease Activity [e]	
Achievement of Remission [f]	
Achievement of Remission [g]	
Improvement in mHAQ≥0.25 [h]	
Improvement in HAQ≥0.22 [h]	
mACR20 [i]	
mACR50 [i]	
mACR70 [i]	

N_{miss} = number of observations with missing values for the specified variable; SD = Standard deviation

[a] Change defined as value at 6-month post-switch visit minus value at switch

[b] Patient fatigue scale defined from 0='Fatigue is no problem' to 100='Fatigue is a major problem'

[c] Worsening defined for patients in remission, low, or moderate disease activity at switch and worsening by at least one category at 6-months post-switch

[d] Low Disease Activity defined as 6-month post-switch CDAI≤10 among those with CDAI>10 at switch

[e] Low Disease Activity defined as 6-month post-switch CDAI≤10 for all patients

[f] Remission defined as 6-month post-switch CDAI≤2.8 among those with CDAI>2.8 at switch

[g] Remission defined as 6-month post-switch CDAI≤2.8 for all patients

[h] Improvement in mHAQ calculated only for patients with mHAQ at switch ≥0.25, and improvement in HAQ calculated only for patients with HAQ at switch ≥0.22

[i] modified ACR: based on 2 out of 4 measures (not using ESR or CRP)

Table 5_switchers. Number of switchers from 5 mg BID to 11 mg QD (from the cohort considered in Tables 4 and 5) that switch* to 5 mg BID by their 6 month follow-up visit.

*Switch defined as discontinued 11 mg QD and initiated or restarted 5 mg BID.

Appendix:

Pfizer QD Power Calculations:

Since the approval of Tofacitinib 11 mg QD in 2/2016, we have been assessing whether there are enough patients to adequately compare these patients to patients on 5 mg BID. Power calculations below indicate that to detect a reasonable difference of 15% in the percentage meeting MCID in CDAI, about 300-500 patients per group would be needed.

With the following assumptions,

- Consider the dichotomous outcome of meeting MCID in CDAI
- We would like to have 90% power to detect a difference between the two groups in percentage meeting MCID in CDAI
- We would like to have a low chance (alpha of 5%) of declaring that there is a difference in the outcomes between the two groups when in actuality there is not (type I error)

The following table gives the number needed per group and total number needed for different scenarios of meeting MCID for CDAI and percent of patients who are available under common support (e.g. percent of patients in the control group that are similar enough to the treatment group so that we are comfortable that the two cohorts are similar). If we assume that the percentage meeting MCID in CDAI is 50%, the following sample sizes are required.

Difference between the two groups in percentage meeting MCID in CDAI that can be detected	**Analyzable Data Set, N per group**	**% under common support**	**N per group required to recruit**	**Total N required to recruit**
10%	535	50%	1070	2140
10%	535	75%	713	1426
15%	**237**	**50%**	**474**	**948**
15%	**237**	**75%**	**316**	**632**
20%	131	50%	262	524
20%	131	75%	175	350
25%	84	50%	168	336
25%	84	75%	112	224

Preliminary Projections:

The next step was to consider when an adequate number of patients would be reached. For this, we completed preliminary projections.

Under different scenarios for projections of the total number of Tofacitinib 11 mg QD initiators over time, we find that there will be:

- About 125-135 Tofacitinib 11 mg QD initiators with 6 month follow-up visits by the end of March 2018 (e.g. report ready for Day 80 due 6/18/2018)
- About 150-170 Tofacitinib 11 mg QD initiators with 6 month follow-up visits by the end of June 2018 (e.g. report ready for Day 120 due 9/14/2018)
- About 180-220 Tofacitinib 11 mg QD initiators with 6 month follow-up visits by the end of 2018
- About 260-320 Tofacitinib 11 mg QD initiators with 6 month follow-up visits by the end of 2019

Note that there is the need for patients that have a follow-up visit at 6 months (+/- 3 months) in order to assess efficacy at 6 months. As of February 2017, about 60% of the total number of initiators had at least one follow-up visit.

Interim Look:

Oftentimes when designing a study, investigators would like to get an initial look at the data in order to inform future decisions. While it is not possible to continually re-analyze the data, it is possible to do pre-planned interim looks at the data prior to when there are sufficient patients as per the power calculation. There are however drawbacks to looking at the data prematurely. There is the possibility that incorrect inferences are drawn due to the limited information. This is particularly an issue if the patients accrued at the beginning of the study are somehow different from later patients. Also, when computing the final outcome of the study, it is necessary to adjust for the fact that you have already taken a look at the data. This means that if you are considering that a p-value less than 0.05 would indicate a difference between the two groups, then if you have already taken a look, the p-value required to say that there is a difference between the two groups would need to be less than 0.05.

We are now proposing to consider two interim looks at the data – one at 33% accrual and another at 50% accrual. If we assume that we wish to detect a change in 15% from the proposed MCID of CDAI of 50% with 75% under common support, we would need 237/group analyzable and the p-value for the first interim look would need to be less than 0.001 to be considered statistically significant, a p-value for the second interim look would need to be less than 0.003 to be considered statistically significant and a p-value for the final look would need to be less than 0.049 to be considered statistically significant.

Supplemental Analyses

CCI [redacted]

[redacted].

CCI [redacted]

[redacted]

Background Information on Outcomes

HAQ and mHAQ:
The Health Assessment Questionnaire Score is calculated as a total score divided by the number of non-missing measures. The score is a composite measure across 20 questions of the following 8 realms: dressing & grooming, arising, eating, walking, hygiene, reach, grip and activities. Please see the subject questionnaire for details on the questions asked.

From the online documentation, twenty specific activities are assessed on a 4-point Likert scale where 0 = without difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do. The 20 activities are grouped into 8 functional categories with each category given a single score equal to the maximum value of their component activities (0, 1, 2, or 3). Total score is between 0-3.0 in 0.125 increments with increasing scores indicating worse functioning with 0 indicating no functional impairment and 3 indicating complete impairment.

The modified HAQ considers 8 specific questions that are representative of the realms listed above. The questions considered are "dress yourself, including tying shoelaces and doing buttons", "get in and out of bed", "life a full cup or glass to your mouth", "walk outdoors on flat ground", "wash and dry your body", "bed down to pick up clothing from the floor", "turn faucets on and off", and "get in and out of a car". The scale is considered missing if the number of non-missing components is 6 or less.

From the online documentation, eight activities are rated on a 4-point Likert scale where 0 = without any difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do. The mHAQ scale ranges from 0-3 with higher scores indicating worse functioning.

The following references give details on the coding of the HAQ.
Redelmeier et al. Arch Int Med 1993;153:1337-42
Wells et al J. Rheumatol 1993;20:557-60
Guzman et al. Arth Rheum 1996 39:5208

The following references give details on the coding of the mHAQ.
https://www.niehs.nih.gov/research/resources/assets/docs/haq_instructions.pdf
(Uhlig, et al, Rheumatology, 2006)
This reference states that "A change in MHAQ of 0.25 has been suggested as clinically meaningful" and references the Wolfe et al article.
http://onlinelibrary.wiley.com/doi/10.1002/acr.20620/full
Wolfe F, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. Arthritis Rheum 1999; 42: 1797–808.

DAS28 and mDAS28:
DAS28 is an indicator of disease activity that is calculated using a weighted combination of four measures: tender joint counts (0-28), swollen joint counts (0-28), erythrocyte sedimentation rate (ESR) and patient reported global health assessment (0-100). The DAS28 range was 0 to 9.4. Laboratory values were not required in the Corrona registry, so this information is missing for many of our patients. Therefore, we suggest that a modified DAS28 measure be used instead. This measures is also a weighted combination of measures: tender joint counts (0-28), swollen joint counts (0-28), modified health

assessment questionnaire (mHAQ) measure (0-3), patient reported pain (0-100), physician reported global health assessment (0-100) and patient reported global health assessment (0-100). As Jeff mentioned, we have published a manuscript validating the mDAS for use in place of the DAS measure (Bentley, Greenberg and Reed, The Journal of Rheumatology, 2010).

The mDAS28 ranges from 0-9.4 and higher scores indicate higher disease activity.

ACR20/50/70 and mACR20/50/70:
ACR20 is defined as a composite endpoint; if a patient has a response, then he/she shows at least a:

- 20% improvement in tender joint count and
- 20% improvement in swollen joint count and
- At least a 20% improvement in 3 out of 5 of the following endpoints:
    - Patient pain assessment (0-100)
    - Patient global assessment (0-100)
    - Physician global assessment (0-100)
    - Patient self-addressed disability (0-3) - mHAQ
    - Acute phase reactant (erythrocyte sedimentation rate - ESR or C reactive protein - CRP)

ACR50 and ACR70 are defined in a similar fashion.
Because the measures include laboratory values that were not collected for all patients as part of routine clinical practice, e.g. ESR and CRP, a modified ACR20/50/70 is suggested. The mACR20 is also defined as a composite endpoint; if a patient has a response, then he/she shows at least a:

- 20% improvement in tender joint count and
- 20% improvement in swollen joint count and
- At least a 20% improvement in 2 out of 4 of the following endpoints:
    - Patient pain assessment (0-100)
    - Patient global assessment (0-100)
    - Physician global assessment (0-100)
    - Patient self-addressed disability (0-3) - mHAQ

mACR50 and mACR70 are defined in a similar fashion.
The Ranganath et al article (ACR remission criteria and response criteria, Clinical and Experimental Rheumatology, 2006) provides justification for the ACR measure.

Patient pain
The question asked is: “How much pain have you had because of your arthritis IN THE PAST WEEK?”. The scale ranges from 0-100 with 0 being no pain and 100 being pain as bad as it could be.

Patient fatigue
The question asked is: “How much of a problem has unusual fatigue of tiredness been for you IN THE PAST WEEK?” The scale ranges from 0-100 with 0 being fatigue is no problem to 100 fatigue is a major problem.

EQ-5D Index

The EQ-5D descriptive system comprises the following 5 dimensions: mobility (1-3), self-care (1-3), usual activities (1-3), pain/discomfort (1-3) and anxiety/depression (1-3). Each dimension has 3 levels: no problems (1), some problems (2), severe problems (3). The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the 5 dimensions.

The EQ-5D index provides an overall measure ranging from 0 to 1 where 0 is death and 1 is perfect health. The scoring algorithm was derived from: Shaw JW, Johnson JA, Coons SJ.
U.S. Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model. Medical Care. Submitted 2004.

[redacted]

Statistical Analysis Plan (SAP) - Pfizer 054c

Title: QD Efficacy Analysis – Addendum

As a complement to the primary analysis, an exploratory analysis will be conducted using the full pre- matched analysis population (N = 450). In addition to the results, we will provide a summary describing new analyses, methods used including differences from prior methods, and interpretation of results.

Objective 1: Repeat the primary analyses (Tables 1 – 3_adj) using the full analysis population (N = 450).

New analyses will be incorporated into the existing report and tables will be named so as to avoid confusion with pre-existing analyses. Along with each newly analyzed section, we will describe any changes from the previous analyses and we will provide interpretation of all results.

Analysis: We will update the analyses in the following ways:
Table 1: Will not be altered. This table shows baseline characteristics of the full analysis population.
Table 1a: Will not be re-generated. Table 1a showed baseline characteristics among the PS matched population and will be excluded from this analysis.
Table 2: Will be generated using the full analysis population.
Table 2_adj: Will be generated using the full analysis population. The unadjusted estimate will remain unadjusted, and the adjusted estimate will be generated using a model which considers imbalanced covariates (as defined by the standardized differences from Table 1). We will perform model selection to ensure that we optimize the combination of a high level of explanatory power and a concise number of degrees of freedom.
Table 2a: Will be generated using the full analysis population. Table 2b: Will be generated using the full analysis population. Table 3: Will be generated using the full analysis population.
Table 3_adj: Will be generated using the full analysis population. Adjusted estimates will be calculated similar to those in Table 2_adj.

Objective 2: Provide detailed information on exclusion criteria used when selecting the 297 eligible QD initiators (out of the original 611 initiators) as well as the 153 BID initiators (out of the original 258 initiators).

Analysis: We will update the sample selection figure to reflect the sample derived from the end of September 2018 version of the registry. Reasons for exclusion will be detailed and will include: initiated treatment too late (i.e. after 3/31/2018) and so had not reached time for 6 month visit, no 6 month visit despite having initiated treatment

longer than 6 months ago, no valid CDAI at initiation, and no valid CDAI at 6 months despite attending visit.

	5 mg BID (N/%)	**11 mg QD (N/%)**
Initiated after 3/31/2018 (not reached 6 months		
No 6 month visit		
No valid CDAI at initiation		
No valid CDAI at attended 6 month visit		

In addition to the analyses, we will provide an interpretation of the results, including thoughts on the reasons for the listed difference. We will also assess the potential biases that may be introduced into the analyses due to the exclusion criteria.

Objective 3: Repeat the primary analyses (Tables 1 – 3_adj) using a propensity score trimmed population.

To ensure the adjusted analyses do not extrapolate due to the use of dissimilar patients, the propensity score which was previously used to match patients can be used to trim the population of the initiations for which there is no similar initiation in the other group – carrying out the multivariable adjustment for patients on "common support" across the covariates of interest. The propensity score distribution will be examined in both populations and each population will be trimmed for extreme values: patients in one group with propensity scores higher or lower than scores in the other group will be "trimmed" (i.e., excluded from analyses). Only patients with scores overlapping both distributions will be included in the analysis.

Analyses: The analyses will be updated in the same way as in objective 1, using the propensity score trimmed population instead of the full analysis population. When considering covariates to include in adjusted analyses, we will adjust for age, gender, duration of RA, and baseline CDAI (as in the primary analyses), in addition to any imbalanced covariates (those with standardized difference > 0.2). New analyses will be incorporated into the existing report and tables will be named so as to avoid confusion with pre-existing analyses. We will update Table 2 to include the results of a chisquare test in addition to the original t-test results and we will provide an interpretation of both. Along with each newly analyzed section, we will describe any changes from the previous analyses and we will provide interpretation of all results.

CCI

[redacted]

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

CCI [redacted]

C CI [redacted]

[redacted]

CCI [redacted]

CCI [redacted]

References:

1. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP) Guideline on the Choice of the Non-Inferiority Margin. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf. Accessed 15 July 2018

2. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP) Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-Clinical and Clinical Issues. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/01/WC500180219.pdf. Accessed 15 July 2018

Corrona Statistical Analysis Plan (SAP): QD Efficacy Analysis – 2nd Addendum

Prepared for: PPD, PhD; Pfizer

Prepared by: PPD, MS and PPD, PhD

12 September 2019

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

Proposed Analyses:

1. *Repeat all analyses (Tables 1 – NI) using more recent data*

Of the previously defined cohort of 869 tofacitinib initiators (611 initiators of 11 mg QD, and 258 initiators of 5 mg BID) who initiated on or after 2016/02/01, 306 patients (245 initiators of 11 mg QD, and 61 initiators of 5 mg BID) had no available follow-up visits post-initiation using the original 2018/09/30 version of the RA registry, and so were excluded from analysis. In order to remedy the high proportion of initiators with no follow-up, we will consult the most recent data available (RA registry version 2019/08/31) in order to provide the most complete understanding of the availability of follow-up for these patients.

Preliminary investigations indicate that 117 of these 306 initiators have an eligible six-month follow-up visit available and can be included in further analyses (Appendix Table). With the addition of these patients, we would proceed to further analyses with 401 (66%) out of the 611 initiators of 11 mg QD, and 172 (67%) out of the 258 initiators of 5 mg BID (Table 2). After updating the eligible sample and replicating all analyses, the report previously delivered on CCI [redacted] will be re-created. (Note: final analyses will also require valid CDAI measurements at initiation and six-month follow-up, which will exclude a small number of patients from these cohorts)

Table 2. Tofacitinib initiators' eligibility status categories by initiation dose, using 2019/08/31 version of RA database

	All Initiators	11 mg QD	5 mg BID
After 1/31/2016	869	611	258
No follow-up	113 (13%)	79 (13%)	34 (13%)
First follow-up not during first year	67 (8%)	51 (8%)	16 (6%)
Follow-up within first year BUT no six-month visit	116 (13%)	80 (13%)	36 (14%)
Eligible six-month visit	573 (66%)	401 (66%)	172 (67%)

2. *Perform sensitivity analysis of primary outcome*

Using the updated eligible cohort detailed in Table 2 above, 34% of initiators will remain ineligible for analysis. We will present an additional table similar to Addendum Full Table 2 CCI [redacted] which will include all 869 initiators of 11 mg QD and 5 mg BID, treating those who are not eligible due to lack of six-month follow-up visit as not achieving MCID in CDAI.

[redacted]

Justification for proposed analyses:

The original analysis excluded 48% of initiators (51% of 11 mg QD initiators and 40% of 5 mg BID initiators) due to lack of eligible follow-up. When incorporating the most updated follow-up information, the exclusion rate will be reduced to 33%. In addition to the raw reduction in exclusion rate, the rates will be balanced between initiation dose groups (34% of 11 mg QD initiators excluded compared to 33% of 5 mg BID initiators excluded).

Additionally, presenting the rate of achievement of MCID in CDAI while assuming that all ineligible initiators do not achieve MCID will attempt to determine the potential effects of a theoretical worst-case scenario.

CCI [redacted]

[redacted]

CCI [redacted]

Where currently we only consider visits between 3 and 9 months as eligible to be the six-month follow-up visit, we will widen the window of time defining a six-month follow-up visit, considering any visit within the first year after initiation as eligible to be the six-month follow-up visit. This altered definition will create a larger analysis cohort (Table 3); therefore, we will replicate all analyses performed CCI [redacted] using this new sample of 689 patients.

CCI [redacted]

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

Proposed analyses:

1. *Compare eligible initiators to non-eligible initiators across several key characteristics at initiation*

   We will create a baseline table comparing eligible (the green highlighted row in Table 2 above) and non-eligible (the white, red, and orange highlighted rows in Table 2 above) initiators. The table will include key characteristics: sex, age, duration or RA, line of therapy, and CDAI, and comparisons will be made using standardized differences.

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

CCI [redacted]

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

CCI

CCI

[redacted]

[redacted]