Statistical Analysis Plan

Real-World Outcome of Adalimumab on Rheumatoid Arthritis Patients in China (ROCKI Study)

Prepared for: AbbVie Pte Ltd

Protocol version used:
Version 4.0 – 9Jun2015

Case Report Form version used:
V9_0

Date Statistical Analysis Plan approved:

Preparation: ███████████

Approval:

Version No: 0.4

Date: [30May2019]

DOCUMENT HISTORY

Version #	Issue Date	Author/Editor	Sections Changed	Change Summary
V0.4	30 May,2019	▮	Section 2, 5	1. Add median (Q1, Q3) for describing continuous variables when normality is violated. 2. Use Q1,Q3 to replace 95% Confidence interval of mean, due to the violation of normality. 3. Added paired Wilcoxon signed- rank test as an alternative test for paired t-test when normality is violated.
V0.3	24 Dec, 2018	▮	Section 2, 5	1. Added paired Wilcoxon signed-rank test as an alternative test for HAQ DI and WPAI analyses, due to the violation of normality. 2. Changed ANOVA to ANCOVA for sensitivity analysis of HAQ DI change, because we used continuous DAS28. 3. Added Spearman correlation as an alternative method for association analyses, due to the violation of normality.
V0.2	2 Jan, 2018	▮	Section 2, 5	1. Patient's impression of change in RA at weeks 12 and 24 after the initiation of adalimumab has been added in the Study Objectives and Statistical Analysis Section according to protocol. 2. Disease status of RA (moderate, severe, missing) has been deleted in Table 2, since it's duplicated with DAS28 categorization. 3. Dosage of first adalimumab injection has been deleted in Table 2, since the DOSE should be 40MG for all adults.
V0.1	8 Aug, 2017	▮	All	Initial Version

TABLE OF CONTENTS

1.   INTRODUCTION

1.1   Study Rationale and Background

Disability has been defined as impairments, activity limitations and participation restrictions due to personal and environmental factors (1). The concept of disability is one where a physical health condition or disease is evaluated in terms of its impact, difficulties, or limitations on a range of tasks, activities, or roles that are considered typical of everyday life. Examples of affected activities include basic aspects of daily living such as eating, bathing, dressing, household chores and meal preparation, or participation in society, or participation in work.

For public health purposes disability is becoming increasingly important as an outcome measure. Despite this, there has been no data available, within our knowledge, on the effectiveness of adalimumab on health-related QoL and work productivity in patients with Rheumatoid Arthritis (RA) in China. Results from study of effect of adalimumab on Work Productivity and Activity Impairment (WPAI) scores and other Patient-Reported Outcomes (PROs) of work activity and well-being will be of interest to a variety of stakeholders in the healthcare system including patients, healthcare practitioners and payers in China.

The Chinese government has for some time been deeply concerned with an impending productivity decrease associated with ageing and chronic illnesses, including musculoskeletal disorders. These concerns have a large impact on their health financing and structuring policies. Therefore, having economic and health outcomes "real world evidence" (RWE) for the use of Humira (Adalimumab) in moderate RA patients from China will be valuable to support the reimbursement / pricing maintenance and expansion.

In China, Humira is not listed on either provincial reimbursed drug list (PRDL) or national reimbursed drug list (NRDL), which means that patients have to pay for the use of biologics including Humira. Given the high costs of biologics, the use of biologics, including Humira, in patients with RA is low in China. The average length of Humira use for patients with RA in China is only 3 months.

On August 30th, 2012, National Development and Reform Commission (NDRC), Ministry of Health (MOH) and other four ministries and commissions of China issued "Guidance about implementation of urban and rural residents' critical disease insurance (CDI) system". The CDI system aims to expand the coverage of the country's health care insurance system to include the treatment of critical illness, aiming to relieve urban and rural families of the heavy burden of catastrophic medical spending. By the end of August 2013, 94 regions across 23 provinces had piloted the critical illness cover system and seven provinces had fully implemented the system, benefiting 210 million people.

In Jan 2015, Humira was listed on the CDI in Qingdao. Patients with RA who are covered by CDI are reimbursed for 70 percent of the total costs. Given CDI is a new system in China, the Qingdao officials encourage companies to collect the RWE associated with the use of Humira to further demonstrate the value of Humira to patients. The result will assist decision makers to evaluate the CDI from a broader perspective and to determine the value of continuing the program.

In addition, generating RWE related with the use of Humira in patients with RA will be extremely valuable for application of Humira on the CDI list in other provinces in China. The successful applications in other provinces will ensure more patients with RA in China will be able to access the Humira.

The objective of this non-interventional, observational study is to assess the effect of adalimumab on health-related QoL, work productivity, and healthcare resource utilization (HCRU) in patients with RA in China.

2.   OBJECTIVES

The objective of this study is to assess the effect of adalimumab on health and disability outcomes in patients with the immune-mediated inflammatory diseases of rheumatoid arthritis using the real world data as observed. The effect of adalimumab on health and disability outcomes in these patients will be assessed by the primary outcome measure which is the change in Health Assessment Questionnaire Disability Index (HAQ DI) score at 24 weeks after the initiation of adalimumab. The HAQ DI is selected as the primary endpoint as it is commonly used to assess improvements in physical function in RA clinical trials and recommended by the US Food and Drug Administration (FDA) guidance on RA treatment development (3, 4). In addition, the HAQ-DI has been utilized as a predictor variable in investigations of productivity (5). The HAQ-DI has been demonstrated to be significantly correlated with work-related measures such as work capacity, household work performance, work task performance, and work disability (6-9). In addition, the effect of adalimumab will also be assessed by the secondary outcome measures which are changes to the WPAI, EuroQol 5 dimension (EQ-5D) score, and Short Form 36-Item Health Survey (SF-36) domain scores at 12 and 24 weeks after the initiation of adalimumab in RA.

2.1   Primary Objectives

- To describe the change in HAQ DI score at 24 weeks after the initiation of adalimumab (i.e. baseline), in those patients continuing on adalimumab (as observed population). This change will be summarized as the mean and 95% confidence interval or median (Q1,Q3) and will be tested with

a paired t-test or paired Wilcoxon signed-rank test without adjusting for baseline disease severity. For sensitivity analysis, the mean change in HAQ DI at Week 24 will be analyzed using ANCOVA adjusting for baseline disease severity.

2.2   Secondary Objectives

- To describe the change in HAQ DI score at 12 weeks after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population). This change will be summarized as the mean and 95% confidence interval or median (Q1,Q3) and will be tested with a paired t-test or a paired Wilcoxon signed-rank test without adjusting for baseline disease severity. For sensitivity analysis, the mean change in HAQ DI at Week 12 will be analyzed using ANCOVA adjusting for baseline disease severity.

- To describe the change in other patient reported outcomes (SF-36 domain scales, EQ-5D-3L index, WPAI total lost productivity) at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population). These changes will be summarized as the means and 95% confidence intervals or median (Q1,Q3) and will be tested with paired t-tests or paired Wilcoxon signed-rank tests.

- To summarize the number and percent of patients achieving a clinically meaningful improvement on the HAQ DI, as defined by a -0.22 point improvement or greater, at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population).

2.3   Exploratory Objectives

- To describe the change in HAQ DI score at 24 weeks after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population), compared with those patients not continuing on adalimumab (withdrawal population). The mean or median changes in these two groups will be compared using an independent t-test or paired Wilcoxon signed-rank tests.

- Patient's impression of change in RA at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population). Each response will be summarized as number and percentage.

- To summarize the healthcare resource utilization using number and percentage within each category (number of consultations, number of procedures received, number of surgeries, number of hospitalizations and length of stay, number and type of concomitant medications)

- To summarize the change in patient satisfaction questions at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population). These changes will be summarized as number and percentage of each response and compared with Wilcoxon signed rank test. Satisfaction will also be dichotomized and analyzed over time with Cochrane-Armitage test for trends.

- To analyze the associations between disease severity and PROs (HAQ DI, SF-36 domain scales, EQ-5D-3L index, WPAI total lost productivity)

- To analyze the associations between change in disease severity and change in PROs (HAQ DI, SF-36 domain scales, EQ-5D-3L index, WPAI total lost productivity)

- To explore the potential modifying effects of baseline measures on the changes in primary and secondary efficacy outcomes (i.e. PROs). These measures will include age, gender, baseline severity, and other diagnoses/co-morbidities.

3.   STUDY DESIGN

3.1   Overview of Study Design

This study is designed as a prospective, observational study to assess the effect of adalimumab on health-related QoL and work productivity in patients with RA in China in clinical practice.

This study is non-interventional. Patient therapy is not decided by the study protocol but falls within current medical practice, and the prescription of adalimumab is clearly separated from the decision to include the patient in this study. The subjects/investigator will follow the current clinical practice in each site and also the routine clinical follow up as determine by the treating physician.

RA patients will be recruited from within the clinical settings of each rheumatologist participating in the study. Approximately 55 patients diagnosed with RA that meet the inclusion and exclusion criteria will be enrolled at up to 8 sites in China.

To assess health and disability outcomes, the HAQ DI will be assessed at baseline, Week 12 and Week 24 after treatment initiation with adalimumab. In addition, other PROs of work activity and well-being, including the WPAI, EQ-5D-3L, and SF-36,

will also be collected. All above instruments have been validated for the Chinese version.

The patients identified by the recruiting investigators based on the study selection criteria will complete a set of patient questionnaires on QoL, functioning, work productivity, treatment satisfaction, impression of change and HCRU. Data will be captured at baseline (D0), Week 12 and Week 24.

The HCRU will also be collected. This includes surgical procedures, hospitalizations, bed days in hospital, physician consultations etc. Costs will be assigned based on the HCRU using standardized costs for each participating centers.

3.2   Study Population

Approximately 55 patients diagnosed with RA will be recruited from totally up to 8 sites. All eligible patients within the enrollment period will be included; inclusion and exclusion criteria for the population are as below.

3.2.1   Inclusion Criteria

Patients meeting all of the following inclusion criteria at baseline will be included:
- Subject has a diagnosis of RA as defined by the 1987 revised ACR classification criteria and/or the ACR/ the European League against Rheumatism (EULAR) 2010 classification criteria (any duration since diagnosis)
- Male or female subjects ≥18 years of age (local definition according to adalimumab label) who is in compliance with eligibility for adalimumab based on the local label
- Patients with moderate to severe RA defined as DAS28 (ESR) or DAS28 (CRP)>3.2
- Biologically treatment naïve and initiated adalimumab at baseline visit
- Availability of clinical data of the previous 12 weeks prior to baseline
- Ability to self-complete patient questionnaires
- Subject must be able and willing to provide written informed consent and comply with the requirements of this study protocol

Additional Inclusion Criteria
- Female subjects who are either not of childbearing potential, defined as postmenopausal for at least 1 year or surgically sterile (bilateral tubal ligation, bilateral oophorectomy and/or hysterectomy) or are of childbearing potential and are practicing an approved method of birth control throughout the study and for 150 days after the Week 24 dosing or the last dose of adalimumab.

- Subjects judged to be in good health as determined by the Principal Investigator based upon the results of medical history, laboratory profile, physical examination, chest x-ray (CXR), and a 12-lead electrocardiogram (ECG) performed no more than 3 months (90 days) have passed.
- Subjects have negative TB Screening records. If a subject has evidence of a latent TB infection, the subject must initiate and complete a minimum of 1 month (per China guideline) of an ongoing TB prophylaxis or have documented completion of a full course of TB prophylaxis, prior to Baseline

3.2.2   Exclusion Criteria

Patients meeting any of the following exclusion criteria at baseline will be excluded:
- Patients who are pregnant or breast feeding at enrolment or wish to become pregnant in the next 24 weeks
- Participation in any RA-related clinical trial at the time of enrolment, at baseline or at any point during the past 24 weeks prior to baseline
- Patients, who in the clinician's view, may not be able to accurately report their QoL or prior resource utilization
- Patients, who in the clinician's view, may not be able to adhere to adalimumab therapy over 24 weeks

Additional Exclusion Criteria
- Subject has been treated with any investigational drug of chemical or biologic nature within a minimum of 30 days or five half-lives (whichever is longer) of the drug prior to the Baseline Visit
- Infection(s) requiring treatment with intravenous (IV) anti-infectives within 30 days prior to the Baseline Visit or oral anti-infectives within 14 days prior to the Baseline Visit
- Prior exposure to biologics that have a potential or known association with PML (i.e., natalizumab (Tysabri®) or rituximab (Rituxan®))
- Known hypersensitivity to adalimumab or its excipients
- History of demyelinating disease (including myelitis) or neurologic symptoms suggestive of demyelinating disease
- History of invasive infection (e.g. listeriosis and histoplasmosis), human immunodeficiency virus (HIV)
- Subjects with any active viral infection that based on the investigator's clinical assessment makes the subject an unsuitable candidate for the study
- Hepatitis B: HBs Ag positive (+) or detected sensitivity on the HBV-DNA PCR qualitative test for HBc Ab/HBs Ab positive subjects
- Chronic recurring infections or active TB
- History of moderate to severe congestive heart failure (NYHA class III or IV), recent cerebrovascular accident and any other condition which would put the subject at risk by participation in the study

- Evidence of dysplasia or history of malignancy (including lymphoma and leukemia) other than a successfully treated non-metastatic cutaneous squamous cell or basal cell carcinoma or localized carcinoma in situ of the cervix
- Positive pregnancy test at Screening or Baseline
- History of clinically significant drug or alcohol abuse in the last 12 months
- Clinically significant abnormal screening laboratory results as evaluated by the Investigator

3.3   Data Sources/Data Collection Process

3.3.1   Data Sources

The data sources of this non-interventional observational study are Case Report Forms (CRFs) and patient questionnaires. Collection of data includes but not limited to subject demographics, clinical history, comorbidities, spontaneous adverse events, and concomitant medications. The following questionnaires will be utilized to collect data directly from participating subjects:

- HAQ DI
- SF-36
- EQ-5D-3L
- WPAI
- Patient Global Impression of Change (PGIC)
- Patient Treatment Satisfaction Questions
- HCRU

3.3.2   Data Collection

IMS will coordinate data collection with each site. At study initiation (Baseline, D0), patients will be asked to provide their written informed consent. After signing consent, the patients' questionnaires as recorded will be collected. At Week 12 and 24 time-points, the patients' questionnaires will be collected before them visiting their physicians. There will be a study coordinator or study nurse at each site to monitor the process. The study coordinator/nurse will collect the record of questionnaires directly. Physicians will not see the contents patients entered. The study coordinator/nurse will check to ensure all the forms are filled-in correctly.

In addition, physicians will also be asked to complete a clinical CRF in paper form based on the patient's medical records at baseline, 12 weeks, and 24 weeks. The CRF completed by physicians will be returned to the study coordinator/nurse at each site as well.

All data will be collected by the study coordinator/nurse at each site and returned directly to IMS. The data collection time is expected to be 10-13 months.

All patient and clinician data will be handled preserving confidentiality.

3.3.3   Ethics and Quality

Prior to any study-related data being collected, informed consent form will be reviewed, signed and dated by the patient and the person who administered the informed consent. A copy of the signed informed consent will be given to the patient and the original will be placed in the patient's medical record.

3.3.4   Quality Assurance

Prior to the initiation of the study, physician and site personnel will be trained on the study. Training will include a detailed discussion of the protocol, performance of study procedures, and completion of the CRFs and paper questionnaires.

All sites will be monitored during the course of study participation. One hundred percent (100%) source document review for safety will be performed.

All clinical data will be documented via the CRF. Study coordinators at each site will check the paper CRFs completed by the physicians and questionnaires completed by patients (e.g., all questions were answered). Data entry will be conducted by IMS. After entry of the data, computer logic checks will be run to check for inconsistent data. Any necessary corrections will be made to the database and documented via addenda, queries, and source data clarification forms.

Steps to be taken to ensure the accuracy and reliability of data include the selection of qualified investigators and appropriate study centers; review of protocol procedures with the investigator and associated personnel before the study and periodic monitoring visits by the sponsor. Written instructions will be provided for administration and collection of study questionnaires.

Guidelines for CRF completion will be provided and reviewed with study personnel before the start of the study. The sponsor will review CRFs and patient questionnaires for accuracy and completeness after transmission to the sponsor; any discrepancies will be resolved with the investigator or designee, as appropriate. After upload of the data

into the clinical study database they will be verified for accuracy and consistency with the data sources.

3.4 Data Management and Storage Process

3.4.1 Data Management

Data management and data quality check will be performed to remove errors and inconsistencies in order to assure the appropriateness of the study data set to assess the study objectives. Data entry screens will be developed and tested prior to initiating data collection to reduce data entry errors. If required, IMS will provide AbbVie with data for analysis.

Each site coordinator will be instructed to answer patients' queries which may arise in relation to the questionnaires and check the patients' input to make sure all questions are answered.

3.4.2 Storage Process

Following data quality checks, each dataset will be converted to SAS and merged for analysis. All information included in the CRF and the patients' questionnaires will be checked in order to detect possible queries to solve and will be extracted to a specifically designed database, where it will be validated by IMS personnel to ensure its quality. Finally, data analysis will be conducted and final results reported.

The databases will be stored in IMS data servers. Data servers are submitted to daily backups in order to increase the security on all data managed in the collection process.

All paper based questionnaires will be stored in a secured, locked area for a period of 7 years, after which all data will be shred using an agency specialized in disposal of confidential documents.

3.5 Definitions of Study Variables

The key variables of interest in this study are demographic and clinical characteristics, patient-reported health and disability outcomes, and healthcare resource utilizations. Most variables will be measured at baseline (i.e. index), and at weeks 12 and 24 after

IMS CONFIDENTIAL AND PROPRIETARY INFORMATION

the initiation of Adalimumab; healthcare resource utilization will be summarized over the 6 month post-index period.

3.5.1   Outcome/Endpoint Variables - Scoring of PRO Questionnaires

Health Assessment Questionnaire Disability Index (HAQ DI)

In the HAQ DI, each category contains four response options ranging from 0-3, where 0 = without any difficulty and 3 = unable to do. In any event, a patient must have a score for at least six of the eight categories (i.e. Dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities). The HAQ DI's scoring conventions allow for the computation of two disability indices:

1. Standard HAQ DI (preferred and traditional scoring method; takes into account the use of aids/devices)
    There are three steps:
    1) Sum the 8 category scores by using the highest sub-category score from each category.
        ➢ For example, there are three sub-category items in "Eating". A patient responds with a 1, 2, and 0, respectively; thus, the category score is 2.
    2) Adjust for use of aids/devices and/or help from another person when indicated.
        ➢ Adjust the score for a category by increasing a 0 or 1 to a 2.
        ➢ If the patient's highest score is a 2, it would remain as 2.
    3) Divide the summed category scores by the number of categories answered to obtain a HAQ DI score of 0-3 (3=worst functioning).

2. Alternative HAQ DI Score (without aids/devices)
    There are two steps:
    1) Sum the scores for each category, ignoring scoring for aids/devices
    2) Divide by the number of categories answered (minimum 6 required). This yields a score of 0-3 (3=worst functioning).

Short Form (36) Health Survey (SF-36)

SF-36 scoring will be conducted by following these 7 steps in accordance to the developer's recommendations:

Step 1: Data entry
SF-36 scoring must ensure that the survey form is complete and the respondent's answers are unambiguous. If there are missing responses, using a scoring algorithm could derive scores across the eight SF-36 health domain scales for nearly all survey respondents. If there are multiple responses for a single item, the following rules are applied:

1. If a respondent marks two responses that are adjacent to each other, randomly pick one, and enter that number.
2. If a respondent marks two responses that are not adjacent to each other, consider that item missing.
3. If a respondent marks three or more responses, consider that item missing.

Alternatively, one can opt to treat all items with more than one response as missing.

Step 2: Recoding item response values

This is a process that derives the final item response values, or scores, to be used when calculating the raw scale scores for each health domain.

Step 3: Determining the health domain scale total raw scores

Following item recording, which includes resolving items with missing data, a total raw score is computed for each health domain scale. The total raw score is the simple algebraic sum of the final response values for all the values for all the items in a given scale. The simple scoring method is possible because all the items in a given scale have roughly equivalent relationships to the underlying health construct being measured and because no item is used on more than one scale. As a result, it is not necessary to standardize or weight items.

Step 4: Transforming health domain scale total raw scores to 0-100 scores

Health domain scale total raw scores are converted to 0-100 scores using the following formula:

$$\text{Transformed scale score} = \frac{(\text{Actual raw score} - \text{Lowest possible raw score})}{\text{Possible raw score range}} \times 100$$

Step 5: Transforming health domain scale 0-100 scores to T scores using health domain z scores

The advantages of standardizing the health domain scales and converting 0-100 scores to norm-based scores using a T-score transformation are that health domain scale results can be meaningfully compared with each other and that these scores have a direct interpretation in relation to the distribution of scores in the 2009 US general population.

The first step is to convert the 0-100 scores into a linear z-score. A z-score would ensure each health domain scale has a mean of 0 and a standard deviation of 1 in the 2009 US general population. A z-score is computed by subtracting each health domain scale's 2009 US general population mean from the 0-100 score for that scale, and dividing the difference

by the given scale's standard deviation. Thus, the following formula is applied to convert 0-100 scores into z-scores:

$$\text{z-score} = \frac{([0-100\,\text{score}]-[83.29094])}{(23.75883)}$$

Z-scores are then transformed into T-scores (mean = 50, SD = 10) using the following formula:

$$\text{T-Score} = 50 + (z - \text{score} \times 10)$$

Step 6: Scoring the physical and mental component summary measures using health domain z scores

The Physical Component Summary (PCS) and Mental Component Summary (MCS) measures are then scored using a three-step procedure, regardless of whether a standard or acute form was administered. First, the eight health domain scales are standardized using means and standard deviations from the 2009 U.S. general population. Second, these standardized scores are aggregated using weights (factor score coefficients) from the 1990 US general population. Computation of an aggregate physical component score consists of multiplying each health domain scale z-score by its respective physical factor score coefficient and then summing the eight products. Similarly, an aggregate mental component score is obtained by multiplying each health domain scale z-score by its respective mental factor score coefficient and summing the eight products. To illustrate, a portion of the formula for aggregating scales when estimating a standard form of aggregate mental component score is as follows:

$$\text{Aggregate mental component score} =$$

$$(\text{PF z} - \text{score} \times -0.22999)\ldots + (\text{MH z} - \text{score} \times 0.48581)$$

Third, aggregate PCS and MCS scores are standardized using a linear T-score transformation with a mean of 50 and a standard deviation of 10. The formulas for computing the norm-based T score for each component summary measure are:

$$\text{PCS T-score} = 50 + (\text{Aggregate physical component score} \times 10)$$

$$\text{MCS T-score} = 50 + (\text{Aggregate mental component score} \times 10)$$

Step 7: Scoring the response consistency index (RCI)

The Response Consistency Index (RCI) is optional, but is a simple and easy way to evaluate the consistency of responses to individual survey items. The RCI comprises of 15 pairs of items and assesses each pair for consistency. If a pair of responses is consistent, then the

RCI score for that pair would be 0. Conversely, a pair of inconsistent responses would earn a score of 1. For example, if a respondent indicates that he or she can "walk more than a mile" but, at the same time, cannot "walk 100 yards," then this item pair would be considered inconsistent and would earn 1 RCI point. For a given respondent, the final RCI score is the sum of the scores earned on the 15 consistency checks. Thus, the best (i.e., most consistent) RCI score is 0 and the worst (i.e., least consistent) score is 15. Note that it is not necessary for a respondent to have complete data for all 15 pairs to compute the RCI (pairs with missing or out-of-range data are not used in the final calculation). However, if all 15 pairs have missing data for one or both items, then the RCI for that respondent cannot be scored.

EuroQol 5 dimension, 3 level quality of life questionnaire (EQ-5D-3L)

The EQ-5D-3L questionnaire is made up for two components; health state description and evaluation.

As to the health state description, the EQ-5D-3L descriptive system contains 5 categories (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Each category contains three responses choices: top (having no problems), middle (having some or moderate problems), and bottom (being unable to do/having extreme problems). The top response choice is coded as "1", the middle response choice as "2", and the bottom response choice as "3"; while ambiguous values (e.g. multiple boxes ticked for one category) are treated as missing values, which are coded as "9". As a result, a person's health status can be defined by a 5-digit number, ranging from 11111 (having no problems in all dimensions) to 33333 (having extreme problems in all dimensions). For example, 12321 indicates having no problems in mobility and anxiety/depression, having slight problems in self-care and pain/discomfort, and having extreme problems in usual activities. Thus, there are potentially 243 ($=3^5$) different health states.

Visual analogue scale (VAS) is the second part of the questionnaire, asking to mark health status on the day of the interview on a 20 cm vertical scale with end points of 0 and 100. There are notes at the both ends of the scale that the bottom rate (0) corresponds to "the worst health you can imagine", and the highest rate (100) corresponds to "the best health you can imagine". The respondents have to draw a line from the box on the questionnaire to the scale indicates the health state of the interviewed day; while ambiguous values (e.g. the line crosses the VAS twice) should be treated as missing values, which are coded as "999". Eventually, the EQ-5D-3L results can be converted into a single index value; many different terms are in use for these

index values, such as preference weights, preference-based values, utilities, QALY weights, etc.

Work Productivity and Activity Impairment (WPAI)

WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity (i.e., worse outcomes) as follows:

Questions

Q1 = currently employed

Q2 = hours missed due to specified problem

Q3 = hours missed other reasons

Q4 = hours actually worked

Q5 = degree problem affected productivity while working

Q6 = degree problem affected regular activities

Scores

Multiply scores by 100 to express in percentages.

- Percent work time missed due to problem
$$Q2 / (Q2+Q4)$$

- Percent impairment while working due to problem
$$Q5 / 10$$

- Percent overall work impairment due to problem

$$Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4) \times (Q5/10)]$$
- Percent activity impairment due to problem
$$Q6 / 10$$

Patient Treatment Satisfaction Questions

The patient treatment satisfaction questions were developed de novo for this study and are not considered "validated" questions. There are totally 4 questions asking about the patients' satisfaction with the following 4 aspects respectively:

1. The way RA treatment improves the morning stiffness in and around the joints.

2. The way RA treatment improves the mobility.

3. The way RA treatment improves the ability to perform daily living requiring fine motor skills.

4. The overall RA treatment.

There are totally 5 responses for each of the question: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, and very dissatisfied.

3.5.2   Exposure/Independent Variables of Interest

- **Patient demographics (if data available)**
  - Date of birth
  - Age at baseline
  - Age groups
  - Gender at birth
  - Marital status (married, unmarried)
  - Medical plan type
    - ✓ The type of healthcare plan each patient is enrolled in at baseline will be determined, including National/Public insurance, Private insurance, Employer benefits, and Other

- **Clinical characteristics**
  - Initial Diagnosis of Rheumatoid Arthritis (RA)
    - ✓ Year of initial diagnose
  - Initiation of Adalimumab
    - ✓ First time administration (DD/MM/YYYY)
    - ✓ Dosage of first injection (in mg)
  - Disease status of RA
    - ✓ Moderate (defined as Disease Activity Score in 28 Joints (DAS28) between 3.2 and 5.1)
    - ✓ Severe (DAS28>5.1)
  - Erythrocyte sedimentation rate (ESR) (in mm/hour)
  - C reactive protein (CRP) (in mg/L)
  - Disease Activity Score in 28 joints (DAS28)

To calculate the DAS28, the physician or specialist nurse will:

1. Count the number of swollen joints (out of the 28)

2. Count the number of tender joints (out of the 28)

3. Take blood to measure the erythrocyte sedimentation rate (ESR) or C reactive protein (CRP); the latest measurement in the medical records may also be used if a blood draw is not standard for that visit

4. Ask the patient to make a 'global assessment of health' which will be indicated by marking a 10 cm line between very good and very bad

The physician or specialist nurse will then mark each component score on the CRF and the DAS28 score if calculated (an online scoring calculator for the DAS28 can be found at http://www.das-score.nl/das28/DAScalculators/dasculators.html).

The DAS28 will be scored using the following formula:

$$DAS28 = 0.56 * sqrt(tender28) + 0.28 * sqrt(swollen28) + 0.70 * ln(ESR\ or\ CRP) + 0.014 * (Global\ assessment\ of\ Health)$$

DAS28 scores will be interpreted using the following categorization:
- ✓ Remission: DAS28 ≤ 2.6
- ✓ Low Disease Activity: 2.6 < DAS28 ≤ 3.2
- ✓ Moderate Disease Activity: 3.2 < DAS28 ≤ 5.1
- ✓ High Disease Activity: DAS28 >5.1

- comorbidities
  - ✓ Myocardial infarction
  - ✓ Congestive heart failure
  - ✓ Peripheral vascular disease
  - ✓ Cerebrovascular disease
  - ✓ Dementia
  - ✓ Chronic pulmonary disease
  - ✓ Connective tissue disease
  - ✓ Ulcer disease
  - ✓ Mild liver disease
  - ✓ Diabetes
  - ✓ Hemiplegia
  - ✓ Moderate or severe renal disease
  - ✓ Diabetes with end-organ damage
  - ✓ Any tumor
  - ✓ Leukemia
  - ✓ Lymphoma
  - ✓ Moderate or severe liver disease

- ✓ Metastatic solid tumor
- ✓ AIDS

3.5.3   Other Co-Variates/Control Variables

Baseline demographics and clinical characteristics will be used in statistical analysis models.

3.5.4   Derived Variables

- Number of consultations/visits to healthcare professionals for treatment of RA, categorized by the treatment department
  - ✓ Family Medicine
  - ✓ Rheumatologist
  - ✓ Gerontologist
  - ✓ Orthopedist
  - ✓ Internist
  - ✓ Emergency department
  - ✓ Traditional Medicine
  - ✓ Rehabilitation Medicine
  - ✓ Other
- Number of procedures received when visiting the healthcare professional(s) for treatment of RA
  - ✓ Chest X-Ray
  - ✓ Spine X-Ray
  - ✓ Neck X-Ray
  - ✓ Shoulder X-Ray
  - ✓ Hand X-Ray
  - ✓ Knee X-Ray
  - ✓ MRI
  - ✓ CT scan
  - ✓ Electrocardiogram
  - ✓ Blood sample taken
  - ✓ Urine test
  - ✓ Endoscopy
  - ✓ Bone scan
  - ✓ Liver function test
  - ✓ Tuberculin Skin Test

- ✓ Sputum tests
- ✓ Other

- Number of surgeries received when visiting the healthcare professional(s) for treatment of RA
  - ✓ Arthroscopy
  - ✓ Carpal tunnel release
  - ✓ Cervical spinal fusion
  - ✓ Total knee replacement
  - ✓ Total hip replacement
  - ✓ Knee arthrodesis (fusion)
  - ✓ Hip arthrodesis (fusion)
  - ✓ Synovectomy
  - ✓ Other

- Number of hospitalizations in relation to RA

- Length of stay: total days in hospital in relation to RA

- Number of concomitant medications taken for RA

- Disease Modifying Antirheumatic Drug (DMARD)
  - ✓ Auranofin (Ridaura, Myochrysine)
  - ✓ Azathioprine (Imuran)
  - ✓ Cyclosporine (Sandimmune and Neoral)
  - ✓ Hydroxychloroquine (Plaquenil)
  - ✓ Leflunomide (Arava)
  - ✓ Methotrexate (Rheumatrex, Folex)
  - ✓ Minocycline (Minocin, Dynacin and Vectrin)
  - ✓ Sulfasalazine (Azulfidine)
  - ✓ Others

- Anti-inflammatory Drug (NSAID)
  - ✓ Ibuprofen (Advil, Motrin, Nuprin)
  - ✓ Naproxen (Naprosyn, Anaprox, Aleve)
  - ✓ Celecoxib (Celebrex)
  - ✓ Aspirin (ASA)
  - ✓ Diclofenac (Voltaren)
  - ✓ Indomethacin (Indocid)
  - ✓ Others

- Others
  - ✓ Steroids

✓ Stomach or gastric medications

3.5.5   Conventions

Given the real world nature of the data, multiple imputation methods for missing data will introduce bias as missing data cannot be considered completely at random (MCAR) or at random (MAR). Due to the observational nature of this study, non-imputing process will be applied, except with date data where the exact day is missing; day "15" will be assumed. Efficacy measures are not assessed after a participant discontinues adalimumab. The exception to this will be the analysis of proportion of patients at 24 weeks who remain on adalimumab. Item level data on the PROs will be imputed according to the developers' recommendations. Therefore, the analysis will be based on available data. The number of patients with missing data for each variable will be summarized and reported.

4.   SAMPLE SIZE AND POWER

Approximately 100 patients diagnosed with RA will be recruited. The sample size was calculated by assuming an alpha of 0.05 and a power of 0.80 in a two-sided test, resulting in 52 subjects required to recognize a statistically significant improvement in HAQ DI (-0.21) (11). Given the importance of other secondary endpoints, an improvement of 9.9±25 hours of work time lost requires 58 subjects, assuming an alpha of 0.05 and a power of 0.80 in a two-sided test (12). However, since it is expected that approximately half the study population may no longer be working due to their age, approximately twice this number (100 patients) will be required to achieve the required number of respondents. We have also assumed a drop-out rate of 13% (assuming between 10%-15%).

5.   STATISTICAL ANALYSIS

5.1   General Considerations

The statistical analysis will be performed using SAS® 9.3 software or a later version (SAS Institute, North Carolina, USA) via SAS Enterprise Guide version 6.1. The statistical results will be displayed using tables, listings and/or graphs. Figures will be performed with SAS®.

Unless otherwise specified, results will be provided as descriptive statistics. For categorical measures, data will include the frequency (number of cases [N]) and percentage (%) of total study patients observed in each category. Continuous variables

will be described with number of patients with missing observations, mean, standard deviation (SD), median, minimum, and maximum.

The following rules will be followed to define the number of decimal positions to report in study results:

- For categorical variables:
  - 1 for the percentage.
- For most of the continuous variables:
  - 1 for the mean, SD, median, minimum, and maximum.
  - In continuous variables with low values, 2 decimals for all measures will be reported.

A set of draft table shells (see the document, "Abbvie_ROCKI_Tableshell_V0.3_21Nov2018_China.xlsx") will be created in advance to describe the statistical analysis that will be done to fulfill the objectives of this study. Based on the analysis results the tables will be modified accordingly.

5.2   Analysis Plan for Demographic and Clinical Characteristics

- Descriptive statistics will be presented for patient demographics (Excel Table 1), and clinical characteristics (Excel Table 2) for all patients. The following measures will be assessed as of the patient's baseline:
  - Patient age (mean, SD, minimum, maximum, median; categories: 18-34, 35-44, 45-54, 55-64, 65-74, 75+, missing)
  - Gender (male, female, missing)
  - Marital status (married, unmarried, missing)
  - Medical insurance type (National/Public insurance, Private insurance, Employer benefits, other, no insurance, missing)
  - Disease Activity Score in 28 joints - DAS28 (mean, SD, minimum, maximum, median; categories: Remission, Low Disease Activity, Moderate Disease Activity, High Disease Activity, Missing)
  - Number of patients with each of the comorbidities at baseline visit (See 3.5.2), and number of patients with any of these comorbidities

5.3   Analysis Plan for Primary Objectives

The objective of the primary endpoint analysis will be to demonstrate that treatment with adalimumab improves functioning as measured by the HAQ DI score in subjects with RA following treatment initiation (i.e. baseline).

- The changes in HAQ DI score at 12 and 24 weeks after the initiation of adalimumab will be presented (Excel Table 3.1); these changes will be summarized as the mean, SD, minimum, maximum, median, Q1, Q3; as well as the number and percent of patients achieving a clinically meaningful improvement on the HAQ DI, as defined by a -0.22 point improvement or greater. In addition, the mean or median change will be tested with a paired t-test or paired Wilcoxon signed-rank test without adjusting for baseline disease severity; for sensitivity analysis, the mean change will be analyzed using ANCOVA adjusting for baseline disease severity.

- The change in HAQ DI score at 24 weeks after the initiation of adalimumab (mean, SD, minimum, maximum, median, Q1, Q3), in those patients continuing on adalimumab (observed population) will be compared with those patients not continuing on adalimumab (withdrawal population). The mean or median changes in these two groups will be compared using an independent t-test or paired Wilcoxon signed-rank test listed in the table (Excel Table 3.2).

5.4   Analysis Plan for Secondary Objectives

In addition to HAQ DI score, the effect of adalimumab will also be assessed by the secondary outcome measures which are changes to the Short Form 36-Item Health Survey (SF-36) domain scores, EuroQol 5 dimension 3 level (EQ-5D-3L) index, WPAI outcomes, and Patient Treatment Satisfaction Questions at 12 and 24 weeks after the initiation of adalimumab in RA.

- As to the Short Form 36-Item Health Survey (SF-36) domain scores, the Excel Table 4 summarizes the changes in PCS T-score and MCS T-score at 12 and 24 weeks after the initiation of adalimumab; these changes will be summarized as the mean, SD, minimum, maximum, median, Q1, Q3; and the mean or median change will be tested with paired t-tests or paired Wilcoxon signed-rank test .

  - As to the EuroQol 5 dimension 3 level quality of life questionnaire, the Excel Table 5 summarizes the changes in EQ-5D-3L index at 12 and 24 weeks after the initiation of adalimumab; these changes will be summarized as the mean, SD, minimum, maximum, median, Q1, Q3; and the mean or median change

will be tested with paired t-tests or paired Wilcoxon signed-rank test .

- As to the WPAI outcomes, the Excel Table 6 summarizes the changes in percent overall work impairment and percent activity impairment at 12 and 24 weeks after the initiation of adalimumab; these changes will be summarized as the mean, SD, minimum, maximum, median, Q1, Q3; and the mean or median change will be tested with paired t-tests or paired Wilcoxon signed-rank tests.

5.5   Analysis Plan for Exploratory Objectives

- Excel Table 7 describes the change in 4 patient satisfaction questions at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (observed population). These changes will be summarized as number and percentage of each response and compared with Wilcoxon signed rank test. Satisfaction will also be dichotomized and analyzed over time with Cochran-Armitage test for trends.

- Excel Table 8 summarizes the patients' global impression of change in RA at weeks 12 and 24 after the initiation of adalimumab, in those patients continuing on adalimumab (as observed population). Each response (very much better, much better, a little better, no change, a little worse, much worse, and very much worse) will be summarized as number and percentage.

- Descriptive statistics will be presented for post-index overall RA related direct healthcare resource utilization (Excel Table 9)
  - Consultations/visits to healthcare professionals for treatment of RA (number of patients with ≥ 1 visit; categorized by the treatment department listed in 3.5.4)
  - Procedures received when visiting the healthcare professional(s) for treatment of RA (number of patients with ≥ 1 procedure; categories listed in 3.5.4)
  - surgeries received when visiting the healthcare professional(s) for treatment of RA (number of patients with ≥ 1 surgery; categories listed in 3.5.4)
  - Hospitalizations in relation to RA (number of patients with ≥ 1 admission)
  - Length of stay in days (mean, SD, median)
  - Concomitant medications taken for RA (number of patients with ≥ 1 medication; categories listed in 3.5.4))

- Associations between disease severity (measured by DAS28) and PROs (HAQ DI score, SF-36 domain scales, EQ-5D-3L index, WPAI outcomes) will be analyzed using the bivariate linear regression model (Excel Table 10.1). Separate models will be estimated to identify the association between DAS28 and each of the PROs.If normality assumption for the residual of regression is violated, Spearman correlation coefficients and p-values will be estimated instead.

- Additionally, in Excel Table 10.2, separate bivariate liner regression models will be estimated to identify the association between change in DAS28 and change in each of the PROs at 12 and 24 weeks after baseline respectively. If normality assumption for the residual of regression is violated, Spearman correlation coefficients and p-values will be estimated instead.

- In Excel Table 11, separate multivariate generalized linear models (GLM) will be estimated to test for the differential changes in primary and secondary efficacy outcomes after baseline, across the subgroups defined by the potential modifiers, including age, gender, baseline severity, and other diagnoses/comorbidities.

6.   LIST OF TABLES AND FIGURES IN EXCEL WORKBOOK

Results of the analysis for this study as outlined above will be summarized using MS Excel tables and will be presented as follows:

Excel Tables / Figures	Description
1	Demographic Characteristics at Baseline
2	Clinical Characteristics at Baseline
3.1	Change in HAQ DI at Weeks 12 and 24 after Initiation of Adalimumab
3.2	Change in HAQ DI at Weeks 24 after Initiation of Adalimumab – Observed VS Withdrawal Population
4	Change in SF-36 Domain Scales at Weeks 12 and 24 after Initiation of Adalimumab
5	Change in EQ-5D-3L Index at Weeks 12 and 24 after Initiation of Adalimumab
6	Change in WPAI at Weeks 12 and 24 after Initiation of Adalimumab
7	Change in Patient Satisfaction Questions at Weeks 12 and 24 after Initiation of Adalimumab
8	Patients' Global Impression of Change at Weeks 12 and 24 after Initiation of Adalimumab

9	Post-index Healthcare Resource Utilization – RA Related
10.1	Associations between Disease Severity and PROs
10.2	Associations between Change in Disease Severity and Change in PROs
11	GLM – Modifying Effects on Changes in PROs

Please see the document,
"Abbvie_ROCKI_Tableshell_V0.3_21Nov2018_China.xlsx" for the full set of table shells.

7.   ABBREVIATIONS

ACR	American College of Rheumatology
AE	Adverse Events
ANCOVA	Analysis of Covariance
CRF	Case Report Form
CRP	C-Reactive Protein
CTCAE	Common Terminology Criteria for Adverse Events
DAS28	Disease Activity Score in 28 Joints
EQ-5D-3L	EuroQol 5 dimension, 3 level quality of life questionnaire
ESR	Erythrocyte Sedimentation Rate
EULAR	The European League Against Rheumatism
GP	General Practitioner
HAQ DI	Health Assessment Questionnaire Disability Index
HCRU	Healthcare Resource Utilization
ICF	Informed Consent Form
ICH-GCP	International Conference on Harmonization of Good Clinical Practice
NRS	Numeric Rating Scale
PGIC	Patient Global Impression of Change
PHI	Protected Health Information
PRO	Patient-Reported Outcome
QC	Quality Check
QoL	Quality of Life
RA	Rheumatoid Arthritis
SAP	Statistical Analysis Plan
SF-36	Short Form 36-Item Health Survey
SOC	System Organ Class
SOP	Standard Operating Procedure
TNF	Tumor Necrosis Factor
VAS	Visual Analogue Scale
WPAI	Work Productivity and Activity Impairment

8.   REFERENCES

1 World Healht Organisation. Disability Fact Sheet.   [cited May 7, 2015]; Available from: http://www.who.int/disabilities/world_report/2011/factsheet.pdf

2 AbbVie Limited. Adalimumab Data Sheet. 27 November 2014(Version 32).

3 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), (CDRH) CfDaRH. Guidance for Industry Rheumatoid Arthritis: Developing Drug Products for Treatment. 2013.

4 Bruce B, Fries J. The health assessment questionnaire (HAQ). Clinical and experimental rheumatology. 2005;**23**(5):S14.

5 Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. The Journal of rheumatology. 2003;**30**(1):167-78.

6 Barrett EM, Scott DG, Wiles NJ, Symmons DP. The impact of rheumatoid arthritis on employment status in the early years of disease: a UK community-based study. Rheumatology (Oxford, England). 2000 Dec;**39**(12):1403-9.

7 Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. The Journal of rheumatology. 1998 Nov;**25**(11):2108-17.

8 Jantti J, Aho K, Kaarela K, Kautiainen H. Work disability in an inception cohort of patients with seropositive rheumatoid arthritis: a 20 year study. Rheumatology (Oxford, England). 1999 Nov;**38**(11):1138-41.

9 Fex E, Larsson BM, Nived K, Eberhardt K. Effect of rheumatoid arthritis on work status and social and leisure time activities in patients followed 8 years from onset. The Journal of rheumatology. 1998 Jan;**25**(1):44-50.

10 Dougados M, Schmidely N, Le Bars M, et al. Evaluation of different methods used to assess disease activity in rheumatoid arthritis: analyses of abatacept clinical trial data. Annals of the rheumatic diseases. 2009 Apr;**68**(4):484-9.

11 van der Bijl AE, Breedveld FC, Antoni CE, et al. An open-label pilot study of the effectiveness of adalimumab in patients with rheumatoid arthritis and previous infliximab treatment: relationship to reasons for failure and anti-infliximab antibody status. Clinical rheumatology. 2008 Aug;**27**(8):1021-8.

12 Bejarano V, Quinn M, Conaghan PG, et al. Effect of the early use of the anti-tumor necrosis factor adalimumab on the prevention of job loss in patients with early rheumatoid arthritis. Arthritis and rheumatism. 2008 Oct 15;**59**(10):1467-74.

13 Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis care & research. 2011 Nov;**63 Suppl 11**:S240-52.

14 The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health policy (Amsterdam, Netherlands). 1990 Dec;**16**(3):199-208.

15 Oemar M, Oppe M. EQ-5D-3L User Guide: Basic information on how to use the EQ-5D-3L instrument. The EuroQol Group; 2013.

16 Evans C, Crawford B. Patient self-reports in pharmacoeconomic studies. Their use and impact on study validity. PharmacoEconomics. 1999 Mar;**15**(3):241-56.