Approaches to Missing Data in the Analysis of SpondyloArthritis International Society (ASAS 20) Response and the Creation of the Related CDISC Compliant Analysis Data Sets

Christine Joy C. Dureza, PPD Inc., Philippines

ABSTRACT

Randomized controlled trials (RCTs) is one of the preferred study designs for evaluating interventions. Unlike observational studies, the randomization of patients to interventions means that a direct causal association can be made between an intervention and its effect. In order to measure the treatment effect, there is a need to at least measure the patient's response at the end of the trial. In most instances, a series of responses will be measured at baseline and throughout follow-up. However, it is not always possible to collect all the intended data on each individual and in all given timepoints. Thus, analysis of the data that were collected without any further reflection generally leads to misleading conclusions. When the data is incomplete, conclusions between the intervention and response is compromised. In this paper, several approaches to data "missingness' is discussed in the analysis of ASAS 20 response. This response is used as an outcome measure, a composite measure, in the investigation of ankylosing sponydylitis (AS). This is a chronic disease characterized by ankylosis (stiffening and immobility) of the spine, and inflammation at the insertions of tendons. To be fulfilled, ASAS 20 response criteria requires ≥ 20% improvement or at least ≥ 1 unit reduction in a 0 to 10 unit scale in at least 3 of 4 domains, with no worsening in the fourth domain of ≥ 20% or ≥ 1 unit in a 0 to 10 unit scale. The 4 domains refer to patient global assessment, pain, physical function, and inflammation (morning stiffness), or duration of morning stiffness. The missing data approaches such as non-responder imputation (NRI), last observation carried forward (LOCF), and baseline observation carried forward (BOCF) are presented in this paper and considerations and recommentations on how to prepare a CDISC compliant analysis data sets that includes the abovementioned approaches are provided.

INTRODUCTION

Randomized controlled trials is one of the preferred study designs for evaluating interventions. Unlike observational studies, the randomization of patients to interventions means that a direct causal association can be made between an intervention and its effect. In order to measure the treatment effect, there is a need to at least measure the patient's response at the end of the trial. One of the potential sources of bias when analyzing clinical trials is missing data. When the proportion of missing values is substantial, oftentimes interpretation of the results of a trial is always problematic. In clinical trials, missing data is often attributed to patient refusal to continue in the study, patient withdrawals due to treatment failure, treatment success or adverse events, and patient movement from one area to another and only some of the reasons are related to the study drug (EMA, 2010). Hence, there are different scenarios of missing data: 1) measurements may be available only at baseline, 2) measurements may be missing at baseline, and 3) measurements may be missing for one, several or all follow-up assessments (EMA, 2010). Even if a patient completes the study, some data may remain simply unreported or uncollected.

This paper focuses on the discussion of missing data approaches in the analysis of ASAS Response. Only the data missingness methods that are usually used in the analysis of ASAS are discussed in this paper (Khan, 2008). The ASAS response is used to measure the efficacy of new products on the symptoms of AS. Also, this response includes the assessment of the 4 clinical domains and is a composite outcome measure that combines multiple outcomes that are cause-specific. Specifically, ASAS 20 response is considered here which is the acceptable primary efficacy end point to assess the efficacy to evaluate a new product from a new therapeutic class where major improvement is intended to be assessed (EMA, 2010). There are other related response such as ASAS 40 and ASAS 5/6 but are already beyond the scope of this paper. The last part of the paper presents some directions on how to create CDISC compliant data sets for each of the clinical domains considered in the response taking into consideration the missing data imputation and the data set for the ASAS 20 response.

MISSING DATA IN CLINICAL TRIALS

The way missing data are handled can have an implication on the final results of a clinical trial and on the certainty with which conclusions can be drawn. Based on the European Medicines Agency (2010) guideline on missing data in confirmatory clinical trials, missingness can be defined as both the existence of missing data and the mechanisms

that explain the reason for the data being missing. It was also emphasized in the guideline that the extent to which missing values lead to biased conclusion about the magnitude of any treatment effect is influenced by many factors. Among these are the relationship between missingness, treatment assignment and outcome; the type of measure employed to quantify the treatment effect and the expected changes over time for the variables being measured. In addition, it should be noted that the strategy employed to handle missing values might in itself constitute a source of bias and that there is no universal best approach for all situation but nevertheless there are some rules which should be considered when handling missing data.

It is always important that before choosing an approach, the missingness mechanisms are understood. Table 1 shows a detailed description of each data missingness mechanism.

Table 1. Description and Diagram of Data Missingness Mechanisms

Mechanism	Description	Illustration
Missing Completely at Random (MCAR)	The observation is MCAR if the probability of an observation being missing does not depend on observed or unobserved measurement. A typical example is a patient moving to another city for non-health reasons. Patients who drop out of a study for this reason could be considered a random and representative sample from the total study population.	
Missing at Random (MAR)	If the probability of an observation being missing depends only on observed measurements then the observation is MAR. This assumption implies that the behaviour of the post dropout observations can be predicted from the observed variables, and therefore that response can be estimated without bias using exclusively the observed data. For example, when a patient drops out due to lack of efficacy reflected by a series of poor efficacy outcomes that have been observed, it would be appropriate to impute or model poor efficacy outcomes subsequently for this patient.	
Missing not at Random (MNAR)	When observations are neither MCAR nor MAR, they are classified as Missing Not At Random (MNAR). The probability of an observation being missing depends on unobserved measurements. In this scenario, the value of the unobserved responses depends on information not available for the analysis (i.e. not the values observed previously on the analysis variable or the covariates being used), and thus, future observations cannot be predicted without bias by the model. For example, it may happen that after a series of visits with good outcome, a patient drops out due to lack of efficacy. In this situation the analysis model based on the observed data, including relevant covariates, is likely to continue to predict a good outcome, but it is usually unreasonable to expect the patient to continue to derive benefit from treatment.	

Source: Carpenter & Kenward (2007).

COMMONLY USED IMPUTATION METHODS

For nearly a century, medical scientists have been dealing with missing data by deleting or arbitrarily filling in missing cases posthoc. This technique is prone to bias to the extent the study result may be deemed meaningless and may lead to inconclusive facts since all randomized patients are not fully utilized in the analysis. Over the past decades, great efforts have been made in the development of analytical techniques to estimate causal effects in the presence of missing data. Given that there is no universal method to analyze missing data, the National Research Council (2010) released a guideline on the Handling of Missing Data in Clinical Trials. It is focused on the following: 1) careful design and conduct to limit the amount of missing data and 2) analysis that makes full use of information on all randomized participants and is based on careful attention to the assumptions about the nature of the missing data underlying estimates of treatment effects. This paper is only limited to the methods commonly used in ASAS 20

response analysis, description of each method is presented in Table 2. Note that these imputations are post-baseline related imputations.

Table 2. Commonly Used Imputation Methods With Relevance to ASAS 20.

Imputation Methods	Description
Single Imputation	This approach creates a complete set of data for all randomized subjects by using a rule to set missing response to a value. This approach is not complicated as compared to the other approaches. In ASAS 20 response, a Non-responder imputation is used.
Last Observation Carried Forward (LOCF)	This is done by imputing the last measured value of the endpoint to all subsequent, scheduled, but missing, evaluations. In most of the papers, LOCF is considered a single imputation method. But in this paper, we will treat LOCF as a separate method to analyze ASAS 20 response.
Baseline Observation Carried Forward (BOCF)	A data imputation technique which populates missing values with the subject's nonmissing baseline observation.

AS AND ASAS 20

According to the Spondylitis Association of America, AS is a form of arthritis that primarily affects the spine, although other joints can become involved. It causes inflammation of the spinal joints (vertebrae) that can lead to severe, chronic pain and discomfort. In the most advanced cases, this inflammation can lead to new bone formation on the spine, causing the spine to fuse in a fixed, immobile position, sometimes creating a forward-stooped posture. AS can also cause inflammation, pain and stiffness in other areas of the body such as the shoulders, hips, ribs, heels and small joints of the hands and feet. Currently, there is no known cure for AS, but there are treatments and medications available to reduce symptoms and manage the pain. Recent studies show that the new biologic medications can potentially slow or halt the disease progression in some people. Meanwhile, the mean AS prevalence per 10,000 (from 36 eligible studies) was 23.8 in Europe, 16.7 in Asia, 31.9 in North America, 10.2 in Latin America and 7.4 in Africa (Dean, et al, 2013).

Asessment of patient status and response to treatment has been problematic. There were no defined measures of AS before to assess the fundamental aspects of disease course. To address the need for an effective clinical outcome measure optimal for studying new therapeutic agents, the ASAS Working Group was organized. This independent group of AS experts has been meeting for several years towards the establishment of AS assessment criteria. There were over 100 potential clinical outcome measures considered but ASAS identified a core set of 4 clinical domains considered essential for characterizing changes in AS: physical function, pain, patient global assessment, and inflammation. Using the data from some clinical trials in the treatment of AS, the ASAS Working Group developed the ASAS 20. The regulatory point of view was also put into consideration when this outcome was created that the claims of therapy should have the following: improvement of symptoms and signs such as pain and stiffness or enthesopathy, improvement of physical function, slowing or prevention of structural damage, and prevention of disability.
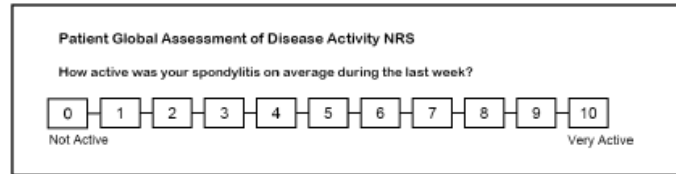
The ASAS 20 (Pentek, 2013) is defined as achieving:

- An improvement from baseline of $\geq$ 20% and $\geq$ 1 unit in at least 3 of the 4 ASAS domains on a scale of 0 to 10 units, and
- No worsening from baseline of $\geq$ 20% and $\geq$ 1 unit in the remaining ASAS domain on a scale of 0 to 10 units.

The four ASAS domains are the following (Pentek, 2013):

- Patient Global Assessment of Disease (0 to 10 unit Numerical Rating Scale [NRS]);
- Total Back Pain NRS;
- Function Bath Ankylosing Spondylitis Functional Index (BASFI) score NRS);
- Inflammation (mean of Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) NRS Question #5 and #6 for morning stiffness).

Patient Global Assessment of Disease

The Patient Global Assessment of Disease Activity is the subject's assessment of how active their spondylitis was on average during the last week. The subject will be asked to mark the box with an X on a 0 to 10 unit NRS on which the left-hand box of 0 represents "not active," and the right-hand box represents "very active".
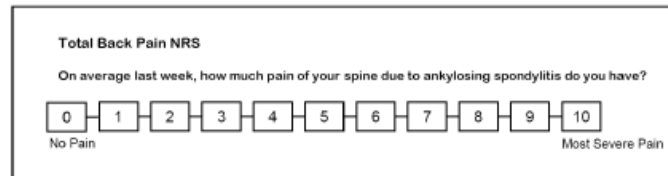

Illustration 1. Patient Global Assessment of Disease Questionnaire

Total Back Pain

The total back pain NRS is the subject's assessment of, on average last week, how much pain they have in their spine due to AS. The subject will be asked to mark the box with an X on a 0 to 10 unit NRS on which the left-hand box of 0 represents "no pain," and the right-hand box represents "most severe pain" .

Illustration 2. Total Back Pain Questionnaire

BASFI

The BASFI is a composite score based on a subject self-administered survey of ten questions using a 0 to 10 unit numerical rating scale (NRS) that assesses a subject's degree of mobility and functional ability . The questionnaire consists of eight questions regarding function in AS and the two last questions reflecting the subject's ability to cope with everyday life. The subject will be asked to mark the box with an X on a 0 to 10 unit NRS for each of the 10 questions, on which the left-hand box of 0 represents "easy," and the right-hand box represents "impossible." The resulting 0 to 100 score is divided by 10 to give a final 0 to 10 BASFI score. A higher BASFI score correlates to reduced functional ability.

*Bath Ankylosing Spondylitis Functional Index (BASFI)

Please indicate your level of ability with each of the following activities during **the last 7 days**. (i.e. ⊠)

(An aid is a piece of equipment that helps you perform an action or movement)

1. Putting on your socks or pantyhose without help or aids (a sock aid, for example).

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

2. Bending forward from the waist to pick up a pen from the floor without an aid.

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

3. Reaching up to a high shelf without help or aids (a helping hand, for example).

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

4. Getting up out of an armless dining room chair without using your hands or any other help.

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

5. Getting up off the floor without help from lying on your back.

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

6. Standing unsupported for 10 minutes without discomfort.

0	1	2	3	4	5	6	7	8	9	10

easy   impossible

Illustration 3. BASFI Questionnaire

BASDAI

The BASDAI is a composite score based on a subject self-administered survey of six questions using a 0 to 10 unit numerical rating scale (NRS) that assesses the subject's five major symptoms of AS: 1) fatigue; 2) spinal pain; 3) peripheral joint pain/swelling; 4) areas of localized tenderness; 5a) morning stiffness severity upon wakening; 5b) morning stiffness duration upon wakening. The subject will be asked to mark the box with an X on a 0 to 10 unit NRS for each of the 6 questions. To give each of the five symptoms equal weighting, the mean of the two scores relating to morning stiffness is taken. The resulting 0 to 50 score is divided by 5 to give a final 0 to 10 BASDAI score. A BASDAI score of 4 or greater is considered to be indicative of active AS disease.

*Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

Please check the box that represents your answer. (i.e. ☒)
All questions refer to **the last 7 days.**

1. How would you describe your overall level of fatigue/tiredness?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

none   very severe

2. How would you describe your overall level of **neck, back or hip** pain resulting from AS?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

none   very severe

3. How would you describe your overall level of pain/swelling in joints **other than** the neck, back or hips?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

none   very severe

4. How would you describe your overall level of discomfort from any areas tender to the touch or pressure?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

none   very severe

5. How would you describe your overall level of morning stiffness from the time you wake up?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

none   very severe

6. How long does your morning stiffness last from the time you wake up?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

0 hr   1 hr   2 or more hrs

Illustration 4. BASDAI Questionnaire

IMPUTATION METHODS FOR ASAS 20

For example, in a study participated by Patient A, B, and C, the efficacy assessment is done in the following timepoints: Visit 2, and Visit 3 where Visit 1 is the first dose administration day. Patient A completed all of the scheduled visits. Patient B discontinued after Visit 2. Patient C discontinued after Visit 1. In relation to the imputation methods presented in Table 2, the subsequent table present a more specific approach on how this will be applied in actual ASAS 20 response analysis.

Table 3. Imputation Methods for ASAS 20.

Condition*	Domain	Imputation
Subjects who prematurely discontinue the study drug for lack of efficacy before the first efficacy evaluation	Patient Global Assessment, Total Body Pain, BASFI, BASDAI	BOCF
	ASAS 20	Single imputation, non-responder.
If there are missing data in the response or in one or more of the component used to define the response.	Patient Global Assessment, Total Body Pain, BASFI, BASDAI	LOCF
	ASAS 20	LOCF

*AS is a chronic and life long disease. Hence, it is expected that the condition of the patient will not improve over time unless a clinical intervention is considered. In addition, according to the Spondylitis Association of America, severity has nothing to do with age and gender.

ADaM DATA SET

According to CDISC ADaM Implementation Guide: "In basic data structure (BDS), subjects, analysis parameters, and analysis timepoints define rows and are identified in standard columns. Subject, parameter and timepoint in combination may not be enough to serve as natural keys (unique record identifiers). There may be mutiple rows within a given combination, depending on the number of observations collected or derived, baseline definition, etc." In BDS, the strong default is to add rows rather than columns. Hence, undue horizontalization should be avoided.

In order to implement the missing data imputation in the four clinical domains, we need to create a derived conceptual timepoint for the LOCF and BOCF observations. Such derivations should result in the creation of new derived records within the same parameter. In this case, a new row should be added, with a corresponding description in AVISIT, and the DTYPE (derivation type) column that contains a description on the row. For example, AVISIT = 'Visit 2 (LOCF)' and DTYPE = 'LOCF'.

Let's go back to the hypothetical study again.
- Patient A completed all the scheduled visits (Visit 1, 2 and 3).
- Patient B completed the first two scheduled visits (Visit 1 and 2).
- Patient C completed the baseline visit only (Visit 1).

There are 2 scenarios considered in the population of LOCF and BOCF records of a subject. The first scenario is when a subject has an actual observation for a visit and the second scenario is when the subject has a missing data for a visit. If the subject has an actual observation for that visit, the rule is to use that actual data as the value of the imputed LOCF and BOCF records. For example, Patient A in Illustrations 5 and 6 has an LOCF and BOCF values equal to the actual observations obtained for that particular visit. Patient A has a Patient Global Assessment Score of 8, 5 and 2 for Visits 1, 2, and 3, respectively. The Visit 2 (LOCF)/Visit 2 (BOCF) is equal to 5 and Visit 3 (LOCF)/Visit 3 (BOCF) is equal to 2.

If a patient has a missing post baseline observation, use the last non-missing record available for that patient. For Patient B in Illustrations 5 and 6, Visit 3 observation is missing. The Visit 2 actual value is used for the Visit 2 (LOCF) but Visit 1 actual value is used for the Visit 2 (BOCF) imputation. If ever a subject has a skipped visit and but has a later visit, carry over the last visit prior to the skipped visit.

If a patient does not have any other observation aside from the baseline observation, use that baseline value in the imputation of Visit 2 (LOCF), Visit 3 (LOCF), Visit 2 (BOCF) and Visit 3 (BOCF). In Illustration 5 for example, Patient C has a baseline Patient Global Assessment score of 8 which is used for all the imputations.

	subjid	visit	qstest	qsstresn	qsstresc	qsblfl
1	A	Visit 1	Patient Global Assessment	8	8	Y
2	A	Visit 2	Patient Global Assessment	5	5	
3	A	Visit 3	Patient Global Assessment	2	2	
4	B	Visit 1	Patient Global Assessment	4	4	Y
5	B	Visit 2	Patient Global Assessment	3	3	
6	C	Visit 1	Patient Global Assessment	8	8	Y

	subjid	visit	param	aval	dtype
1	A	Visit 1	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	
2	A	Visit 2	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	5	
3	A	Visit 3	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	2	
4	A	Visit 2 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	5	LOCF
5	A	Visit 3 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	2	LOCF
6	A	Visit 2 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	5	BOCF
7	A	Visit 3 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	2	BOCF
8	B	Visit 1	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	4	
9	B	Visit 2	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	3	
10	B	Visit 2 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	3	LOCF
11	B	Visit 3 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	3	LOCF
12	B	Visit 2 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	3	BOCF
13	B	Visit 3 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	4	BOCF
14	C	Visit 1	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	
16	C	Visit 2 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	LOCF
17	C	Visit 3 (LOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	LOCF
18	C	Visit 2 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	BOCF
19	C	Visit 3 (BOCF)	Patient Global Assessment of Disease Activity (0-10 Unit NRS)	8	BOCF

Illustration 5. Patient Global Assessment, Portion of SDTM and ADaM Data Sets.

Illustration 6. Total Body Pain, Portion of SDTM and ADaM Data Sets.

Same algorithm is followed in the imputation of missing records for BASFI and BASDAI, however, since these domains contain a series of questions, only the average score is considered in the imputation. For BASDAI, the mean score for question # 5 and #6 is used. See Illustrations 7 and 8 for the sample lay out of the data sets.

Illustration 7. BASFI, Portion of SDTM and ADaM Data Sets.

Illustration 8. BASDAI, Portion of SDTM and ADaM Data Sets.

Finally, a separate data set should be created for the ASAS 20 response. Table 4 presents a summary of scores from each domain as well as the change from baseline (CFB) and percent change from baseline (PCHG) scores. In the definition of ASAS 20, there should be an evident improvement of $\geq$ 20% from baseline and 1 unit or higher increase in at least there of the four domains. Use the PCHG value in looking for the $\geq$ 20% improvement and the CHG for the unit increase. As part of the ASAS 20 response criteria, there should also be no worsening from baseline of $\geq$ 20% and $\geq$ 1 unit in the remaining ASAS domain on a scale of 0 to 10 units.

Table 4. Scores in the Four Clinical Domains and ASAS 20 Response.

SUBJ	VISIT	METHOD	GPA			TBP			BASFI			BASDAI			ASAS 20
			A	C	P	A	C	P	A	C	P	A	C	P	
A	Visit 1	Observed	8			9			5.3			5			
	Visit 2	Observed	5	-3	-37.5	5	-4	-44.4	5.4	0.1	1.9	4.5	-0.5	-10	0
		LOCF	5	-3	-37.5	5	-4	-44.4	5.4	0.1	1.9	4.5	-0.5	-10	0
		BOCF	5	-3	-37.5	5	-4	-44.4	5.4	0.1	1.9	4.5	-0.5	-10	
		NRI													0
	Visit 3	Observed	2	-6	-75	2	-7	-77.7	1.5	-3.8	-71.7	2	-3	-60	1
		LOCF	2	-6	-75	2	-7	-77.7	1.5	-3.8	-71.7	2	-3	-60	1
		BOCF	2	-6	-75	2	-7	-77.7	1.5	-3.8	-71.7	2	-3	-60	
		NRI													1
B	Visit 1	Observed	4			4			2.6			4			
	Visit 2	Observed	3	-1	-25	3	-1	-25	0.2	-2.4	-92.3	2	-2	-50	1
		LOCF	3	-1	-25	3	-1	-25	0.2	-2.4	-92.3	2	-2	-50	1
		BOCF	3	-1	-25	3	-1	-25	0.2	-2.4	-92.3	2	-2	-50	
		NRI													1
	Visit 3	Observed													

	Visit	Method	A	C	P	A	C	P	A	C	P	A	C	P	
		LOCF	3	-1	-25	3	-1	-25	0.2	-2.4	-92.3	2	-2	-50	1
		BOCF	4	0	0	4	0	0	2.6	0	0	4	0	0	
		NRI													1
	Visit 1	Observed	8			9			7.9			9			
		Observed													
	Visit 2	LOCF	8	0	0	9	0	0	7.9	0	0	9	0	0	0
		BOCF	8	0	0	9	0	0	7.9	0	0	9	0	0	
C		NRI													0
		Observed													
	Visit 3	LOCF	8	0	0	9	0	0	7.9	0	0	9	0	0	0
		BOCF	8	0	0	9	0	0	7.9	0	0	9	0	0	
		NRI													0

GPA : Global Patient Assessment; TBP : Total Body Pain; BASFI : Bath Ankylosing Spondylitis Functional Index; BASDAI : Bath Ankylosing Spondylitis Disease Activity Index. Decrease in score means improvement in TBP, BASFI and BASDAI. Increase in score means improvement in PGA.
A : AVAL (Analysis Value) ; C : CHG (Change from baseline); P = PCHG (Percent change from baseline)

Referring to the imputation methods in Table 3, ASAS 20 is imputed using the NRI and LOCF. Since response is relative to the baseline value and is only derived for post baseline visits, the BOCF method is no longer applicable here, thus, the use of NRI. In Illustration 9, an AVAL=1 is considered as a ASAS 20 response and AVAL=0 is considered as a non-response. For Patient A, since the only available observation is the baseline record, this patient is automatically considered as a non-responder.

	subjid	avisit	param	aval
1	A	Week 4	ASAS 20 Response (Observed)	0
2	A	Week 12	ASAS 20 Response (Observed)	1
3	A	Week 4	ASAS 20 Response (LOCF)	0
4	A	Week 12	ASAS 20 Response (LOCF)	1
5	A	Week 4	ASAS 20 Response (NRI)	0
6	A	Week 12	ASAS 20 Response (NRI)	1
7	B	Week 4	ASAS 20 Response (Observed)	1
8	B	Week 4	ASAS 20 Response (LOCF)	1
9	B	Week 12	ASAS 20 Response (LOCF)	1
10	B	Week 4	ASAS 20 Response (NRI)	1
12	C	Week 4	ASAS 20 Response (LOCF)	0
13	C	Week 12	ASAS 20 Response (LOCF)	0
14	C	Week 4	ASAS 20 Response (NRI)	0
15	C	Week 12	ASAS 20 Response (NRI)	0

Illustration 9. ASAS 20, Portion of ADaM Data Set.

Finally, it is recommended that a sensitivity analysis is done to assess the sensitivity of the results obtained to the assumptions made or the methods applied to handle the missing data as part of the clinical tiral reporting.

CONCLUSION

The importance of CDISC Data Standardization has been increasingly regarded since FDA recommends that all clinical data be in standardized format for review and approval. The problem of data missingness has always been part of a clinical trial and can never be neglected. Thus, it is important that records created from missing data methods follow the CDISC standards. This is very important when the outcome of the trial is a composite measure, like the ASAS 20, such that implementation of the CDISC standards is observed from the preparation of the component domains up to the derivation of the final outcome measure.

REFERENCE

Analysis Data Model (ADaM) ImplementationGuide v2.1. CDISC 2009. Available at
http://www.cdisc.org/system/files/all/standard/application/pdf/analysis_data_model_v2.1.pdf

Carpenter, James & Kenward, Michael. 2007. Missing data in randomised controlled trials – a practical guide. Pg 49-68. Medical Statistics Unit, London School of Hygiene and Tropical Medicine, UK.

Dean LE, Jones GT, MacDonald AG, Downham C, Sturrock RD, Macfarlane GJ. APR 2013. Global prevalence of ankylosing spondylitis. Rheumatology (Oxford).
European Medicines Agency (EMA). Guideline on Missing Data in Confirmatory Clinical Trials. July 2010. Available at
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf

European Medicines Agency (EMA). Guideline on Clinical Investigation of Medicinal Products for the Treatment of Ankylosing Spondylitis. June 2005. Available at
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003424.pdf

Khan, Muhammad Asim. Ankylosing Spondylitis. 2008. Oxford American Rheumatology Library. Oxford University Press. Available at
https://books.google.com.sg/books?id=mHhMCAAAQBAJ&dq=missing+data+imputation+ankylosing+spondylitis&ots=Qq2g8KsalP&q=locf#v=snippet&q=locf&f=false

National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. 2010. Available at
http://www.nap.edu/catalog.php?record_id=12955

Pentek, Marta. Systematic review and analysis of evidences on clinical efficacy and cost-effectiveness of biological drugs for the treatment of Ankylosing Spondylitis. 2013. Budapesti Corvinus Egyetem Egészségügyi Közgazdaságtan Tanszék, Budapest. Available at http://unipub.lib.uni-corvinus.hu/1391/1/Pentek_Biologicals.in.Ankylosing.Spondylitis_978_963_503_561_8_16.Dec.2013.pdf

Spondylitis Association of America. Available at http://www.spondylitis.org/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Christine Joy C. Dureza
Enterprise: PPD, Inc.
Address: 9th Floor Sun Life Centre, 5th Ave. corner Rizal Drive, Bonifacio Global City
City, State ZIP: Taguig City, Philippines 1634
Work Phone: (02) 464 7534
E-mail: ChristineJoy.Dureza@ppdi.com

```
*SAMPLE CODES for LOCF and BOCF;
*DATA PREPARATION;
DATA <dataset>;
  set <dataset>;
  where <conditions>;
run;

*ONLY POPULATE LOCF IF AT LEAST ONE
POST BASELINE RECORD;
PROC SORT data=<dataset> out=LOCFPOP
(keep=USUBJID PARAMCD)nodupkey;
  where BASE^=. and VISIT><basevisit>
  and AVAL^=.;
  by USUBJID PARAMCD;
run;

*ONLY POPULATE BOCF IF AT LEAST ONE
BASELINE RECORD;
PROC SORT data=<dataset> out=BOCFPOP
(keep=USUBJID PARAMCD)nodupkey;
  where BASE^=. ;
  by USUBJID PARAMCD;
run;

DATA <dataset1>;
  merge <dataset> LOCFPOP(in=a)
   BOCFPOP(in=b) ;
  by USUBJID PARAMCD;
  if a then LOCF_FLAG='Y';
  if b then BOCF_FLAG='Y';
run;

*CREATE LOCF BOCF RECORD IF AVISITN
RECORD IS AVAILABLE FOR THAT VISIT;
*THESE SUBJECT HAVE A RECORD FOR
AVISITN;
PROC SORT data=<dataset1>
out=CHECK1(keep=USUBJID PARAM)
nodupkey;
  where AVISITN=<visit> and AVAL^=.;
  by USUBJID PARAMCD ;
run;

PROC SORT data=<dataset1> out=LOCF1;
  where AVISITN=<visit> and AVAL^=.;
  by USUBJID PARAMCD AVISITN ADT;
run;

DATA LOCF_REC1;
  length DTYPE $20.;
  set LOCF1;
  AVISITN=<Visit_LOCF>;
  DTYPE='LOCF';
  if LOCF_FLAG='Y' then output;
  AVISITN=<Visit_BOCF>;
  DTYPE='BOCF';
  if BOCF_FLAG='Y' then output;
run;

*THESE SUBJECTS DO NOT HAVE A POST
BASELINE RECORD CREATE LOCF RECORD AS
LAST NON-MISSING POST BASELINE;
DATA LOCF2;
  merge <dataset1> (in=a) CHECK1(in=b);
  by usubjid paramcd;
  if a and not b;
  if AVAL^=. then output;
  PROC SORT; by usubjid paramcd avisitn
adt;
run;

DATA LOCF_REC2;
  length DTYPE $20.;
  set LOCF2;
  by usubjid paramcd avisitn adt;
  if last.paramcd then do;
    AVISITN=<Visit_LOCF>
      DTYPE='LOCF';
  if LOCF_FLAG='Y' then output;
  end;
run;

*THESE SUBJECTS DO NOT HAVE A RECORD
FOR AVISITN CREATE BOCF
BY SETTING TO BASELINE VALUE;
DATA BOCF1;
  length DTYPE $20.;
  merge <dataset> (in=a) CHECK1(in=b);
  by usubjid paramcd ;
  if a and not b;
  if ABLFL='Y' then do;
    AVISITN=<Visit_BOCF>;
      DTYPE='BOCF';
    if BOCF_FLAG='Y' then output;
end;
run;

*SET ALL RECORDS;
DATA <final>;
  set LOCF_REC1 LOCF_REC2 BOCF1;
run;
```