Paper PP17

From Missing to Meaningful: Tackling Partial Dates in Clinical Trials

Bhargav Koduru, Medidata Solutions, Boston, USA
Santosh Reddy Lekkala, Medidata Solutions, Morrisville, USA

ABSTRACT

Handling partial or missing dates is a recurring challenge in clinical trials, often necessitating imputation for key analyses like identifying treatment-emergent events. While many sponsors default to their Statistical Analysis Plan (SAP) for guidance, a universal approach is still lacking. Leveraging the existing guidance laid out in the PHUSE's 'SDTM ADaM Implementation FAQ,' we fine-tuned a more robust data imputation guideline using de-identified clinical data from approximately 50 clinical trials (eligible for Medidata analysis through Medidata's data collaboration program). Our paper will present these optimized guidelines, discussing specific scenarios, critical variables, and common pitfalls. This enhanced summary of scenarios will provide a customizable foundation for date imputation that can be further enhanced to fit your trial needs, bolstering the integrity of clinical trial data for regulatory submissions.

INTRODUCTION

Incomplete or missing data is a frequent challenge in clinical trials, with dates representing one of the most affected data elements. This is understandable, as individuals often struggle to recall precise dates for events in the distant past. Consequently, the start and stop dates for data related to Concomitant/Prior Medications (CM) and Adverse Events (AE) are frequently recorded with partial or missing information. To extract meaningful insights from this data, it is imperative to address these gaps by substituting the missing values with plausible values, a process called 'imputation.' This approach enables sponsors and researchers to draw reasonable conclusions while maintaining the integrity of the data as much as possible. Imputation is achieved at the Analysis Data Model (ADaM) dataset creation stage, usually relying on the imputation guidelines drafted by statisticians as a part of the study's Statistical Analysis Plan (SAP). Each study is unique, so there might be subtle differences in the imputation approach. In this paper, we have explored anonymized/de-identified clinical data from approximately 50 clinical trials (eligible for Medidata analysis through Medidata's data collaboration program), thus casting a wide net to identify the common variables that drive optimal imputation, as well as drafted some example cases to showcase the imputation strategies. This information can be used to draft sponsor-specific rules and possibly automate the date imputation process. We'll also briefly discuss some common data pitfalls post-imputation and how to mitigate them.

VARIABLES AND RULES THAT DRIVE THE IMPUTATION OF DATES

While conducting date imputation across various studies, we identified several key variable groups essential for the accurate imputation of dates. These include:

ANCHOR DATE

A fixed point in time is used as a reference for imputation. It provides a benchmark against which other dates are compared or adjusted. This is mostly the ADaM counterparts of RFXSTDTC (Date/Time of First Study Treatment) or RFXENDTC (Date/Time of Last Study Treatment), but it can be any date driving the analysis.

COUNTER DATE

This represents the complementary date relative to the date being imputed. For instance, the event start date is the counter date to the event stop date, and vice versa. The counter date's position, before or after the anchor date, is crucial in determining the imputation methodology.

For example, if the event start date only has the year value (Same year as the anchor date and event end date) and needs to be imputed, the event end date is available and is greater than the anchor date, which is the treatment start date. In this case, the start date can be imputed as the treatment start date to deem it a possible 'treatment-emergent event.' If the event end date were earlier than the treatment start date, the event start date would have been imputed as the first day of the year.

CUTOFF DATE

Establishing an upper boundary for imputed values is crucial to maintaining their relevance. For instance, when only a year is available for the event end date, it's often imputed as the end of the year. However, if the participant exited the trial before this date, such an imputed value could become irrelevant or speculative. Therefore, using the lesser of the

imputed value or the cutoff date for analysis is essential. The cutoff date might be the data cut date for analysis or the last contact date. The RFPENDTC (Date/Time of End of Participation) is often used as the cutoff date, ensuring that imputed dates remain meaningful and grounded in the context of the participant's actual trial participation.

TIMEPOINT VARIABLES

Including reference time points and their relative variables plays a crucial role in accurately determining imputed dates. These variables can offer insights into the appropriate handling of dates based on the event's status or timing in relation to treatment.

For example, if reference time point variables indicate that an event is "Ongoing," it would be inappropriate to derive an end date. Similarly, these timepoint variables may influence the upper limit for imputing dates. Consider a scenario where the event end date has just the year and month value, the same as the treatment start date (anchor date), year, and month, and the event start date precedes the treatment start date. In such instances, the event end date might be imputed to coincide with the treatment start date if timepoint variables like --ENRTPT equals "COINCIDENT" and --ENTPT is "FIRST DOSE DATE."

Establishing a standardized set of timepoint variables (e.g., "FIRST DOSE DATE," "END OF TREATMENT," "END OF STUDY") and associating them with relevant date variables can significantly enhance automation efforts. This approach allows for a more streamlined and efficient process of date imputation, ensuring that dates are imputed to reflect the clinical context and maintain the integrity of the trial's timeline.

RULES

In addition to the variables mentioned, general guidelines for date imputation can be standardized and implemented as macro-options for automating the process. These guidelines include:

- Absolute vs. Average Approach: In clinical trials, a conservative approach to imputation is often adopted. For example, if an end date is missing the day, it's typically imputed as the last day of the month, whereas a start date missing the day would be imputed as the first of the month. However, instead of automatically opting for the month's start or end, an alternative could be to impute the date as the month's mid-point. This method represents an average scenario more closely, providing a statistically balanced estimate.

- Drug Half-life Driven Anchor Date: In cases where the treatment's end date serves as an anchor for imputing an event's date, particularly when the end date is unknown or incomplete, and the event is identified as 'treatment-emergent', the imputation strategy may be influenced by the drug's half-life. For instance, the event date might be imputed as the 'Treatment end date' or 'Treatment end date plus XY days', where XY days represent the drug's half-life. This approach ensures that the imputation reflects pharmacological considerations, enhancing the relevance and accuracy of the analysis.

VALIDATING YOUR IMPUTATIONS

To ensure the accuracy and integrity of date imputations, the following edit checks are proposed:

CONSISTENCY BETWEEN START AND END DATES

Verify that the imputed start date is always earlier than or equal to the end date for all events and vice versa.

ALIGNMENT WITH ANCHOR DATES

Check that imputed dates (both start and end) fall within the range defined by relevant anchor dates, such as treatment start and end dates.

VALIDATION AGAINST CUTOFF DATES

Ensure that imputed end dates do not exceed the cutoff date, such as the date of the end of participation or the last contact date.

CROSS_CHECK WITH TIMEPOINT VARIABLES

Verify that the imputed dates are consistent for imputations based on specific timepoint variables (e.g., FIRST DOSE DATE, END OF TREATMENT).

LOGICAL SEQUENCE OF EVENTS

Ensure that the sequence of imputed dates across related events (e.g., the start of an adverse event to its end) makes logical sense within the context of the study timeline.

COMPARISON WITH ORIGINAL DATA

Where possible, compare imputed dates against any available partial data to ensure that the imputation respects the original data's constraints (e.g., matching year and month).

EXPLORING DATE IMPUTATION THROUGH EXAMPLES

Having outlined the variables and principles that inform date imputation, let's delve into the practical application of these concepts through specific examples:

END DATE IMPUTATION

The event end date is often imputed first, before the start date, as some components of start date imputation rely on the event end date. In the examples that follow, we operate under certain assumptions: end dates are imputed to the last day of the given month or year (foregoing the average approach), with anchor dates encompassing the treatment start date and the end date (excluding considerations of drug half-life), and the cutoff date set as RFPENDTC.

Furthermore, variables such as --ENTRPT and --ENTPT are utilized alongside the event start date when available. The prefix "XX" denotes the two-letter SDTM Domain value, illustrating how these imputation strategies are applied across different data scenarios.

End date imputation examples:

Category	Example Cases	XXSTDTC	XXENDTC	RFXSTDTC/ RFSTDTC	RFXENDTC/ RFENDTC	IMPUTED ENDT	XXENRTPT	XXENTPT	RFPENDTC
Day value of the end date is missing	ENDTC has MONTH and YEAR, the same as the treatment start date year and month, and the event start date < treatment start date. In this case, the event end date is imputed as the treatment start date, based on XXENTPT and XXENTRTPT values.	2016-01-15	2016-08	2016-08-10	2016-08-20	2016-08-10	COINCIDENT	FIRST DOSE DATE	2016-09-12
	Same scenario as above but XXENTPT and XXENTRTPT values are missing. In this case, the event end date is imputed as the end of the month.	2016-01-15	2016-08	2016-08-10	2016-08-20	2016-08-31			2016-09-12
	ENDTC is partial and is after Treatment end date but Imputing it to end of the month would make the imputed date greater than RFPENDTC. Thus RFPENDTC is considered as the Imputed date.	2016-01-15	2016-09	2016-08-10	2016-08-20	2016-09-12			2016-09-12
Month value of the end date is missing	ENDTC has only a Year value, the same as the treatment start date year, and the event start date < treatment start date. In this case, the event end date is imputed as the treatment end date, based on XXENTPT and XXENTRTPT values.	2016-01-15	2016	2016-08-10	2016-08-20	2016-08-20	COINCIDENT	END OF TREATMENT	2016-09-12
	Same scenario as above but XXENTPT and XXENTRTPT values are missing. In this case, the event end date is imputed as a minimum of the following values: End of the year, RFPENDTC.	2016-01-15	2016	2016-08-10	2016-08-20	2016-12-31			2017-09-12
	ENDTC has only a Year value, the same as the treatment start date year and the event start date > treatment end date. In this case, the event end date is imputed as a minimum of the following values: End of the year, RFPENDTC.	2016-08-25	2016	2016-08-10	2016-08-20	2016-09-12			2016-09-12
	If Year and Day are present with month missing disregard the day value populated and apply the same rules as defined for when only year is available.	2016-01-15	2017—25	2016-08-10	2016-08-20	2017-12-31			2018-09-12
End date is completely missing, or Year is missing	XXENRTPT states as 'ONGOING', and thus, the end date will NOT be IMPUTED.	2016-01-15		2016-08-10	2016-08-20		ONGOING	END OF STUDY	2016-09-12
	The event start date < treatment start date, and the end date is missing. In this case, the event end date is imputed as the treatment end date, based on XXENTPT and XXENTRTPT values.	2016-01-15		2016-08-10	2016-08-20	2016-08-20	COINCIDENT	END OF TREATMENT	2016-09-12
	The event start date > treatment end date, then the event end date is imputed as minimum of the following values: End of the year, RFPENDTC.	2016-08-25		2016-08-10	2016-08-20	2016-12-31	COINCIDENT	END OF STUDY	2017-09-12

START DATE IMPUTATION

The start date is more influenced by the anchor date and the event end date. The examples in the following page showcase some of these scenarios in detail.

Start Date imputation examples:

Category	Imputation Rule	Example Cases	XXSTDTC	XXENDTC	RFXSTDTC/RFSTDTC	RFXENDTC/RFENDTC	IMPUTED STDT	RFPENDTC
Day value of the start date is missing	Impute the DAY as the Day value of the treatment start date for the following scenarios: 1. The month and year of the event are the same as the month and year of the treatment start date, and the event end date ≥ treatment start date. 2. The month and year are the same as the month and year of the treatment start date, and the event end date is missing. If NOT, impute as the 1st of the Month.	The start date's month value < treatment start date, and the end date < treatment starts. Thus, the imputed start date is the 1st of the month.	2016-01	7/31/2016	8/10/2016	8/20/2016	1/1/2016	9/12/2016
		The event start date's month/year is the same as the treatment start date's month/year and the event end date > treatment start date. Thus, the start date will be imputed as the treatment start date.	2016-08	8/15/2016	8/10/2016	8/20/2016	8/10/2016	9/12/2016
		The event start date's month/year is the same as the treatment start date's month/year and the event end date is missing thus the start date will be imputed as the treatment start date.	2016-08		8/10/2016	8/20/2016	8/10/2016	9/12/2016
Month value of the start date is	When MONTH and DAY are both missing, then impute the MONTH and DAY as the MONTH and DAY value of the treatment start date for the following scenarios: 1. The year of the event is the same as the year of the treatment start date, and the event end date ≥ treatment start date. 2. The year of the event is the same as the year of the treatment start date, and the event end date is missing. If NOT, impute as the 1st of January. *When DAY is available, along with YEAR, then disregard the Day value and use the same logic steps as above.	Event end date < treatment start date and hence, the start date is imputed as 1st January.	2016	7/30/2016	8/10/2016	8/20/2016	1/1/2016	9/12/2016
		Event start date's year value is the same as the treatment start date's year value, and the end date > the treatment start date. Thus, it is imputed as the treatment start date.	2016	8/15/2016	8/10/2016	8/20/2016	8/10/2016	9/12/2016
		If present, we disregard the 'DAY' value and apply the same rules as when 'DAY' and 'MONTH' are missing.	2016---20	9/15/2016	8/10/2016	8/20/2016	8/10/2016	9/12/2016
		If the year is the same as the treatment start date and the end date is missing, then impute the start date as the treatment start date.	2016		8/10/2016	8/20/2016	8/10/2016	9/12/2016
Start date is completely missing or,	Impute as treatment start date if end date ≥ treatment start date Impute as 1st January of the same year as end date, if end date < treatment start date	End date > treatment start date and thus event start date is imputed as the treatment start date.		8/15/2016	8/10/2016	8/20/2016	8/10/2016	9/12/2016
		Event end < treatment start thus event start date is imputed as 1st January of the same year as the event end date.		7/15/2015	8/10/2016	8/20/2016	1/1/2015	9/12/2016

CONCLUSION

In this paper, we've laid out a comprehensive analysis of the key variables crucial for date imputation and outlined a series of verification steps to ensure the accuracy of these imputed values. By providing detailed examples, we guide users through the common assumptions and considerations for crafting their date imputation algorithms. Additionally, we delve into strategies for handling various data irregularities, suggest methods for integrating external data points to refine imputation accuracy and emphasize the importance of aligning imputed dates with clinical trial objectives and regulatory requirements. This holistic approach not only aids users in developing robust imputation models but also underscores the significance of thorough validation processes to uphold data integrity and support reliable clinical trial outcomes.

CONTACT INFORMATION

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

Bhargav Koduru
Medidata Solutions
bkoduru@mdsol.com

Santosh Reddy Lekkala
Medidata Solutions
slekkala@mdsol.com