PharmaSUG China 2024 - Paper DS-10042

# General Considerations for Integrating ADaM Data and Special Handling of One Complex Case in ISS

**Niandong Yang**, **Dizal Pharma**

## ABSTRACT

The Integrated Summary of Safety (ISS) is an essential part of a successful new drug application (NDA). However, integrating safety ADaM datasets can be tricky due to legacy datasets, outsourced to CRO, different study designs and various analysis purpose, so it requires a variety of considerations:

- Up version of Coding dictionaries and controlled terminology standardization.
- Harmonize variable attributes and formats.
- Harmonize Derivation logic and value-level metadata.

…

Building integrated ADaM datasets also needs conformance checks with CDISC standards, it's a challenging undertaking, especially in complex cases. This paper will introduce one complex case and its corresponding ADaM structure building strategy in one of our submissions. It's expected to shed some light on how handling the similar case:

- Six studies (DZ001, DZ002, DZ003, DZ004, DZ005, and DZ006) need to be integrated in ISS, DZ001 was a study where subject received monotherapy in a specific period and followed by combination therapy, but only monotherapy phase of the study need to be included in ISS.
- Three subjects who completed DZ002, subsequently enrolled in the study DZ006. How data to be integrated for such repeat participants so that integrated datasets satisfy analytical needs and conform with the CDSIC standards?

## INTRODUCTION

Let's start by outlining the scope of our Integrated Summary of Safety (ISS). This submission comprises six studies (refer to Table 1), two of which (DZ001 and DZ002) were outsourced to Contract Research Organizations (CROs) and concluded in 2019 and 2022, respectively. The remaining four studies were conducted internally.

| Study | Study Type | Indicator | Dose |
|---|---|---|---|
| DZ001 | Phase Ⅰ/Ⅱ | NSCLC | 75 mg QD |
| DZ002 | Phase Ⅰ/Ⅱ | r/r PTCL | 150 mg/ 250 mg QD |
| DZ003 (Pivotal) | Phase Ⅰ/Ⅱ | r/r PTCL | 150 mg QD |
| DZ004 | Phase Ⅱ | r/r CTCL | 150 mg QD |
| DZ005 | Phase Ⅱ | PTCL | 150 mg QD |
| DZ006 | Phase Ⅱ | r/r PTCL | 150 mg QD |

**Table 1. Studies involved in ISS**

**NSCLC: Non-small cell lung cancer; r/r PTCL: relapsed/refractory Peripheral T-Cell Lymphoma.**

## INTEGRATING DATASET USING INDIVIDUAL ADAM DATASETS

The process of building integrated analysis datasets can be approached in various ways. Given the potential for data inconsistencies between studies resulting from different study designs, outsourcing to CRO, and varying analysis needs, adhering to ADaM standards can provide the flexibility and

convenience necessary to handle these discrepancies. Ultimately, we have made the decision to construct our integrated datasets by utilizing individual ADaM datasets from each study. Throughout the ADaM stacking phase, any inconsistencies can be addressed by updating or adjusting individual datasets. Additionally, certain SDTM information is included in individual ADaM datasets within our ISS before they are stacked with other ADaM datasets. Moreover, some SDTM information is added to individual ADaM datasets in our ISS before stacking with other ADaM datasets.

## UP VERSION OF CODING DICTIONARIES AND CONTROLLED TERMINOLOGY STANDARDIZATION

Given that studies were conducted at different times and used different versions of coding dictionaries, there is a potential for data discrepancies. To ensure data consistency, it is crucial to code adverse events (AEs) and concomitant medications (CMs) using a single version of the respective dictionaries in the ISS. Additionally, the most recent version of the dictionaries should be used when summarizing study results. In our ISS, all coded variables will be upgraded to MedDRA 25.1/WHODrug-Global-B3 202209 during the integration of adverse events and concomitant medications, as per external documentation provided by the data management team. It is also important to store the original coding information in the specified variables (Table 2 and Table 3) to maintain traceability. The original coding information, the original coding information is stored in the following variables (Table 2 and Table 3).

| Variables | Label |
|---|---|
| DECDORG1 | PT in Original Dictionary 1 |
| BDSYORG1 | SOC in Original Dictionary 1 |
| HLTORG1 | HLT in Original Dictionary 1 |
| HLGTORG1 | HLGT in Original Dictionary 1 |
| LLTORG1 | LLT in Original Dictionary 1 |

**Table 2. Variables for original coding information in ADAE**

| Variables | Label |
|---|---|
| DECDORG1 | Standardized Med Name in Orig Dict 1 |
| DCD1ORG1 | Standardized Med Name 1 in Orig Dict 1 |
| AT1TORG1 | ATC Level 1 Text in Orig Dictionary 1 |
| AT1CORG1 | ATC Level 1 Code in Orig Dictionary 1 |
| AT2TORG1 | ATC Level 2 Text in Orig Dictionary 1 |
| AT2CORG1 | ATC Level 2 Code in Orig Dictionary 1 |
| … | |

**Table 3. Variables for original coding information in ADCM**

## HARMONIZE VARIABLE ATTRIBUTES AND FORMATS

It's crucial to note that variables with the same name in different studies may possess different attributes. For instance, in ADSL, the variable DCTREAS is 100 characters long in DZ002 but 160 characters long in DZ003. Additionally, a variable may have a character format in one study and a numeric format in another. Harmonizing variable attributes is essential to avoid programming warnings, errors, and data loss due to character truncation.

## HARMONIZE DERIVATION LOGIC AND VALUE-LEVEL METADATA

Due to the different analysis purposes within a single study, the logic of individual variables can also vary. It is important for programmers and statisticians to agree on consistent programming logic for deriving

these variables in ISS. For example, in DZ001, variable ANL01FL represented the worst result in ADEG; however, in other studies, ANL02FL was used to determine the worst result after treatment, while ANL01FL indicated the scheduled visit window. To standardize the derivation logic, we introduced a new variable, anl02fl, in DZ001 to replace the originalANL01FL. The derivation rule for ANL01FL was aligned with other studies. Reconciliation of value-level metadata is also essential for data analysis in ISS.

## ONE COMPLEX CASE IN ISS

Besides the above general considerations, in some cases, creating an integrated summary of safety requires careful consideration of complex situations. This is why it's not possible to produce integrated ADaM datasets by simply stacking them together. We encountered such complex case in this particular submission.

In our study, denoted by DZ001, subjects initially received monotherapy for a specific period and then received combination therapy (see Figure 1). However, based on the statistical analysis plan, only the data from the monotherapy phase should be included in the integrated summary of safety (ISS). This means that only the data from the monotherapy period should be included.

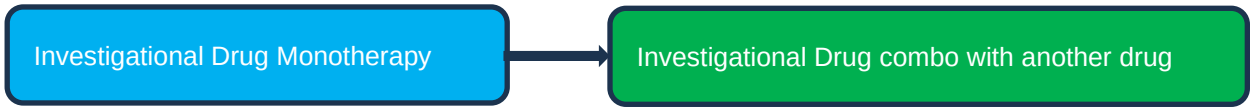
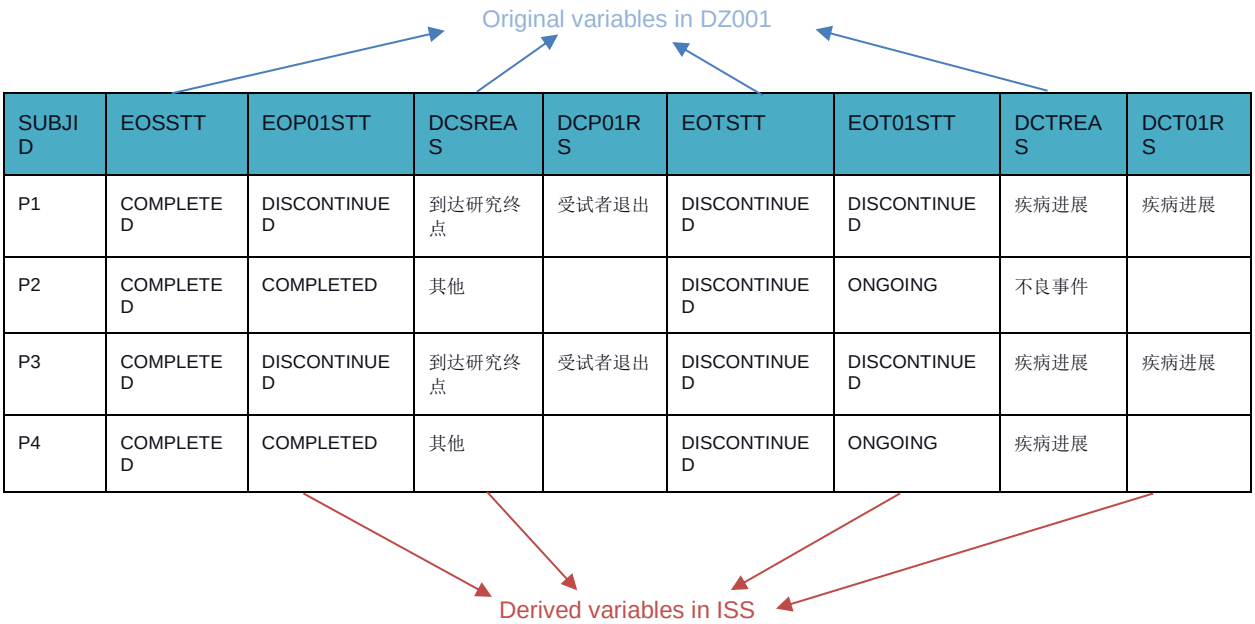



**Figure 1. Study flow chart in DZ001**

It is possible to differentiate between monotherapy and combination therapy based on specific time variables and other factors. In the ADSL, TR01SDT and TR01EDT display the start and end dates of each subject's treatment. Additionally, TRTP and TRTA in the ADaM datasets indicate the specific periods during which subjects were treated. The variables TREM01FL and TREM02FL are used to identify treatment-emergent records in ADAE for monotherapy and combination therapy, respectively.

However, it was observed that ADSL in DZ001 did not contain relevant variables regarding the reasons and status for the conclusion of the study or treatment during the monotherapy period. Consequently, using these variables directly to represent their values in monotherapy is not suitable. Therefore, new variables (EOP01STT, DCP01RS, EOT01STT, and DCT01RS) were added to the integrated ADaM datasets (Table 4).

Four subjects from DZ001 were listed in the integrated ADSL. Among them, P1 and P3 were patients who initially received monotherapy in DZ001 but later dropped out of the study. Their original treatment and study-related variables are displayed in Table 4, along with the derived values. For subjects who underwent combination therapy in DZ001, the assessment of relevant variables must account for the status of monotherapy in relation to the overall study.




| SUBJID | EOSSTT | EOP01STT | DCSREAS | DCP01RS | EOTSTT | EOT01STT | DCTREAS | DCT01RS |
|---|---|---|---|---|---|---|---|---|
| P1 | COMPLETED | DISCONTINUED | 到达研究终点 | 受试者退出 | DISCONTINUED | DISCONTINUED | 疾病进展 | 疾病进展 |
| P2 | COMPLETED | COMPLETED | 其他 | | DISCONTINUED | ONGOING | 不良事件 | |
| P3 | COMPLETED | DISCONTINUED | 到达研究终点 | 受试者退出 | DISCONTINUED | DISCONTINUED | 疾病进展 | 疾病进展 |
| P4 | COMPLETED | COMPLETED | 其他 | | DISCONTINUED | ONGOING | 疾病进展 | |

**Table 4. Variables in the integrated ADSL dataset**

Additionally, three individuals who finished DZ002 subsequently took part in study DZ006 (refer to Table 5). It is known that individuals who take part in multiple studies may have several SUBJID values; however, USUBJID should be unique for an individual across studies in the submission. In our ISS, the USUBJID for the three participants is unique across our compound. It is worth noting that the dose levels for two participants were 250 mg in DZ002 and then reduced to 150 mg in DZ006. All three subjects belong to the r/r PTCL category, as shown in the header of Table 6.

| USUBJID | Subjid in DZ002 | Subjid in DZ006 | Dose in DZ002 | Dose in DZ006 |
|---|---|---|---|---|
| DZ002-P1 | P1 | P4 | 250 mg | 150 mg |
| DZ002-P2 | P2 | P5 | 250 mg | 150 mg |
| DZ002-P3 | P3 | P6 | 150 mg | 150 mg |

**Table 5. The information of three subjects in DZ002 and DZ006 respectively**

| Pivotal | r/r PTCL | | Total | |
|---|---|---|---|---|
| 150 mg | 150 mg | ≥ 150 mg | 150 mg | ≥ 150 mg |

**Table 6. Reporting Header in ISS**

Integrated ADaM datasets need to conform to CDISC standards, which currently require ADSL to have one record per subject. This causes difficulties in integrating ADaM datasets for participants who are enrolled in more than one of studies in the submission. Although the "ADaM Data Structures for Integration," developed by CDISC, allows for multiple records per subject in IADSL, it has not been released yet.

According to the table header in ISS's table shell (Table 6), three participants should be counted in the r/r PTCL group and only once in all the tables. Since ISS analyzed safety data for the overall study, data for three participants should be treated as a whole. However, this also brings up several questions for our analysis:

- Since P1 and P2's dose levels decrease from 250 mg to 150 mg, which treatment group should be counted for them in our ISS? 150 mg or ≥ 150 mg?
- How should we deal with the data for these three subjects participating in both DZ002 and DZ006 when pooling the ADaM datasets? Although they had their baselines in each single study, we must treat one subject's data collected from different studies as a wholeduring the data pooling.

- According to the CDISC standard, one subject should have only one record in ADSL, and only one study can be stored in STUDYID. How should we restructure the pooling ADSL for three patients? How should we pick the appropriate STUDYID and the information for treatment from DZ002 and DZ006 to represent the subject record in pooling ADSL?
- We must structure the analysis dataset conforming to the CDISC standard and fit the SAP analysis requirements.

In the following, I will demonstrate how we address these questions in our ISS-integrated datasets. This can provide insight into how to approach a similar scenario.

## ADSL

The ADSL dataset (Table 7), or "Subject-Level Analysis Dataset," contains only one record per subject. For three participants who completed DZ002 and subsequently enrolled in the study DZ006, we should follow the rule of keeping only three records for these three subjects.

To choose the appropriate information to represent in the ADSL, we need to decide which treatment group should be assigned to participants who had dose changes. After consulting with biostatisticians and medical experts, we agreed to use the treatment information in DZ002 as the treatment group for pooling into the ADSL. The STUDYID also corresponds to "DZ002," and baseline information came from participants' records in DZ002 because it was the project that subjects initiated the first dose.

We used the information in DZ002 at the start date of treatment, and end of treatment information was obtained from DZ006, such as TR02EDT, DCTREAS, etc. To maintain data traceability, we used variables with names containing 01xx to record the information for three participants in DZ002, and the information in DZ006 was recorded in variables named 02xx (as shown in Table 7).

While it may not be ideal to use TR01SDT and TR02SDT to represent the records in the two studies, this method helps the pooled ADSL meet the CDISC standard and be suitable for data analysis.

| STUDYID | USUBJID | TRT01P | TRT01A | TRT02P | TRT02A | TRT01SDT | TRT01EDT | TRT02SDT | TRT02EDT |
|---|---|---|---|---|---|---|---|---|---|
| DZ002 | DZ002-P1 | 250 mg | 250 mg | 150 mg | 150 mg | 2020-05-21 | 2022-05-31 | 2022-06-01 | 2023-10-12 |
| DZ002 | DZ002-P2 | 250 mg | 250 mg | 150 mg | 150 mg | 2020-07-10 | 2022-06-02 | 2022-06-03 | 2022-08-01 |
| DZ002 | DZ002-P3 | 150 mg | 150 mg | 150 mg | 150 mg | 2020-09-21 | 2022-06-17 | 2022-11-21 | 2023-10-12 |

**Table 7. ADSL dataset in ISS**

## ADAE (OCCDS)

ADAE is the Occurrence Data Structure (OCCDS) in CDISC. It is based on restructuring ADSL. We reassigned the value of STUDYID, and variables related to treatment information, such as the start date/end date of treatment, the reason for the end of treatment/end of study, etc. Since the start date and end date of the pooling dataset had been changed for these three participants who enrolled in both DZ002 and DZ006, the original ASTDY/AENDY also needs to be re-derived. To maintain traceability between the pooling and single datasets, we added a new variable named "ASTUDYID" to identify the original record from which study. TRTP/TRTA represented the planned and actual treatment for this record. For example, if one record is from P1' ADAE in DZ002, then the TRTP/TRTA would show 250 mg in this record in pooling ADAE (Table 8).

| STUDYID | ASTUDYID | USUBJID | TRTP | TRTA | TRTSDT | TRTEDT | ASTDY | AENDY |
|---|---|---|---|---|---|---|---|---|
| DZ002 | DZ002 | DZ002-P1 | 250 mg | 250 mg | 2020-05-21 | 2023-10-12 | | |
| DZ002 | DZ006 | DZ002-P1 | 150 mg | 150 mg | 2020-05-21 | 2023-10-12 | Re-derive | Re-derive |
| DZ002 | DZ006 | DZ002-P1 | 150 mg | 150 mg | 2020-05-21 | 2023-10-12 | Re-derive | Re-derive |

**Table 8. ADAE dataset in ISS**

### ADLB (BDS)

Similarly, we retrieved the STUDYID and treatment-related data from the pooling ADSL for the ADLB dataset. To maintain traceability among the datasets, the original STUDYID in a single ADLB dataset was stored in ASTUDYID. When pooling ADLB, the planned treatment group and the actual treatment group in the original dataset were assigned TRTP/TRTA. It's important to note that for three participants, their baseline information such as ABLFL and BASE should reflect their baseline value in DZ002 (the first study in the compound) rather than their original baseline value in ADLB records in DZ006. Consequently, related variables derived from baseline variables should be re-calculated due to the change in baseline value, such as CHG and PCHG (see Table 9).

| STUDYID | ASTUDYID | USUBJID | AVISIT | PARAMCD | AVAL | BASE | CHG |
|---|---|---|---|---|---|---|---|
| DZ002 | DZ002 | DZ002-P1 | 基线 | CA | 2.34 | 2.34 | . |
| DZ002 | DZ002 | DZ002-P1 | 第 1 周期第 8 天 | CA | 2.3 | 2.34 | -0.04 |
| DZ002 | DZ002 | DZ002-P1 | 第 2 周期第 1 天 | CA | 2.27 | 2.34 | -0.07 |
| DZ002 | DZ006 | DZ002-P1 | 长期随访-筛选期 | CA | 2.29 | 2.34 | -0.05 |
| DZ002 | DZ006 | DZ002-P1 | 长期随访-第 2 周期第 1 天 | CA | 2.3 | 2.34 | -0.04 |

**Table 9 ADLB dataset in ISS**

### ADEXSUM

ADEXSUM has a BDS structure where each parameter contains summary statistics per patient per analysis interval. These three participants already had their respective parameter analyses in a single study. However, according to the SAP, we must combine their records in DZ002 and DZ006 and derive the overall parameter analysis in ISS. Specific formulas must be specified in the SAP and annotated in the footnote.

## CONCLUSION

The appropriate integration of safety ADaM datasets is essential but also challenging. The information we have shared should offer some potential solutions for projects facing similar issues. The existing CDISC guidelines can be interpreted in different ways, allowing for the use of alternative methods for dealing with complex cases.

Various approaches in previous published papers might work well as long as there clear and detailed specification either in define.xml or analysis data reviewer guide (adrg).

The integration guidelines developed by CDISC seem to be making progress in helping to make these decisions as more data-pooling efforts are carried out. The CDISC conformance rules can be released at the time of publication, allowing Pinnacle to make necessary adjustments.

## REFERENCES

Bharath Donthi, Lingjiao Qi, 2019 “Best Practices for ISS/ISE Dataset Development” *PharmaSUG 2019*

Laura Phelan, Magalie Gallet, 2019 “Choosing the right path to follow when integrating ADaM” *PhUSE Connect 2019*

Wayne Zhong, Accretion Softworks; Kimberly Minkalis, 2018 “ADaM Structures for Integration: A Preview” *PharmaSUG 2018*

Navneet Agnihotri, Sumit Pradhan, 2021 “Integrated Summary of Safety and Efficacy Production Strategies” *PhUSE Connect 2021*

## ACKNOWLEDGMENTS

The author thanks Weibin Cai and other team members at Statistical and Programming department for their thoughtful review of this manuscript.

## CONTACT INFORMATION

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

Niandong Yang
Dizal Pharma
Niandong.Yang@dizalpharma.com