PharmaSUG 2024 - Paper SS- 213

Is a Participation-Level ADaM Dataset a Solution for Submitting Integration Data to FDA?

Sandra Minjoe, ICON plc

ABSTRACT

In 2019, the CDISC ADaM team was prepared to publish a solution for datasets that support integrated analyses. This document allowed for an "integrated ADSL" dataset, structured as one record per subject per pool. However, the document was never published because it was determined to be challenging for FDA to perform a review when their tools depend on a one-record-per-subject ADSL. There is still no CDISC solution for Integration.

In 2022, the CDISC SDS team proposed a standard for handling DM content when a subject participates in a study more than once. It kept DM as one record per subject, allowing the sponsor to determine which of the multiple participations to put into DM, plus added proposed standard domain DC, with many of the same variables as DM, but structured as one record per subject per participation.

The proposed DM/DC solution prompted the CDISC ADaM team to consider a similar ADaM solution: keeping ADSL plus adding a dataset such as ADPL (Participation-Level Analysis Dataset) that contains the same variables as ADSL but structured as one record per subject per participation. A big takeaway is that this solution doesn't break any ADaM rules and can be used now.

Can ADPL or something similar be used for integrated analyses?

INTRODUCTION AND HISTORY

All the documents produced by the CDISC ADaM team are for study-level analysis. There is no official guidance from CDISC on how to handle integration.

Some of you might remember that the ADaM team put out a draft integration document for public review. It included an integrated ADSL (iADSL) that was structured as one record per subject per pool, to be used when subjects participated in multiple studies within the integration and would then end up in different pooled analyses. In 2019, the team had resolved all the public review comments and was ready to release the document as final, but at the last minute heard from a group at the FDA that they didn't want a dataset acting as ADSL that could have more than one record per subject. The ADaM team proposed using the "split" subdirectory, with a separate dataset to use as an ADSL for each pool, in addition to iADSL, but this was also not accepted. Since the ADaM team and FDA had reached an impasse, the entire Integration document was shelved by CDISC and remains unreleased.

PROPOSED SDTM SOLUTION FOR MULTIPLE PARTICIPATIONS

In 2022, the CDISC SDS team proposed a new standard domain, DC (Demographics as Collected), with similar variables as DM, but structured as one record per subject per participation (SUBJID). This would be used when a subject participates in a study more than once. This proposal keeps DM unchanged as one record per unique subject ID (USUBJID), and allows the sponsor to decide which one of the DC participations to consider "primary" and include in DM.

In addition to the new DC domain, the SDTM proposal included the ability to use variable SUBJID in all other SDTM domains. That means that not only can DM content be easily merged onto any other SDTM domain by USUBJID, DC content can also be easily merged onto any other SDTM domain by SUBJID.

Below are some example SDTM datasets for a single subject, where

- Table 1 shows the proposed SDTM DC content,

- Table 2 shows SDTM DM, which uses only one of the DC rows (in this case, the first, but it is up to the sponsor to make this choice), and
- Table 3 shows a couple tests for this subject in VS.

For space reasons, not all required variables in each dataset are shown.

USUBJID	SUBJID	RFXSTDTC	ARM
ABC-01	ABC-01	2023-05-02	A
ABC-01	ABC-25	2023-10-15	B

Table 1. Example SDTM DC for one subject

USUBJID	SUBJID	RFXSTDTC	ARM
ABC-01	ABC-01	2023-05-02	A

Table 2: Example SDTM DM for one subject

USUBJID	SUBJID	VSTESTCD	VSSTRESN	VSSTRESU	VSDTC
ABC-01	ABC-01	HEIGHT	167	cm	2023-05-01
ABC-01	ABC-01	WEIGHT	68	kg	2023-05-01
ABC-01	ABC-01	WEIGHT	65	kg	2023-05-22
ABC-01	ABC-25	HEIGHT	167	cm	2023-10-16
ABC-01	ABC-25	WEIGHT	66	kg	2023-10-16
ABC-01	ABC-25	WEIGHT	67	kg	2023-10-30

Table 3: Example SDTM VS for one subject

It's worth noting that the FDA seems to like this proposal. The FDA Study Data Technical Conformance Guide Section 4.1.1.2 specifically recommends including SUBJID in domains when needed to differentiate study participations, noting that doing so may cause validation errors.

PROPOSED ADAM SOLUTION FOR MULTIPLE PARTICIPATIONS

This SDTM solution has led the ADaM team to discuss a similar solution for multiple participations: keep ADSL as is (one record per USUBJID), plus add an ADPL (Participation-Level Analysis Dataset) to hold similar variables with one row per subject per participation (one record per SUBJID). Some things to note:

1. Like the DM and DC solution, the sponsor would be the one to choose which participation from ADPL to use in ADSL.
2. Because ADaM already allows SUBJID to be included on any other ADaM dataset, it would be trivial to merge on ADSL or ADPL data.
3. ADPL, structured as one record per subject per participation, doesn't conform to any of the current ADaM standard structures, but it could be used right now with the class ADAM OTHER.

The paper from the 2023 PharmaSUG Conference titled "ADaM Datasets with Multiple Participations per Subject" gets into a lot of detail on this topic, but here is a quick example, similar to what we just saw with SDTM DM and DC. Here,

- Table 4 shows an example ADSL, using the single record from SDTM DM in Table 2 as input,
- Table 5 shows an example ADPL, using both records from SDTM DC in Table 1 as input,
- Table 6 shows an example ADVS, merging ADSL content from Table 4 with VS content from Table 3, and

- Table 7 shows an example ADVS, merging ADPL content from Table 5 with VS content from Table 3.

For space reasons, not all required variables in each dataset are shown.

USUBJID	SUBJID	TR01SDT	TRT01P
ABC-01	ABC-01	02MAY2023	A

Table 4: Example ADaM ADSL for one subject

USUBJID	SUBJID	TR01SDT	TRT01P
ABC-01	ABC-01	02MAY2023	A
ABC-01	ABC-25	15OCT2023	B

Table 5: Example ADaM ADPL for one subject

USUBJID	SUBJID	PARAM	AVAL	ADT	TR01SDT	ADY	TRTP
ABC-01	ABC-01	Height (cm)	167	01MAY2023	02MAY2023	-1	A
ABC-01	ABC-01	Weight (kg)	68	01MAY2023	02MAY2023	-1	A
ABC-01	ABC-01	Weight (kg)	65	22MAY2023	02MAY2023	21	A
ABC-01	ABC-25	Height (cm)	167	16OCT2023	15OCT2023	168	A
ABC-01	ABC-25	Weight (kg)	66	16OCT2023	15OCT2023	168	A
ABC-01	ABC-25	Weight (kg)	67	30OCT2023	15OCT2023	182	A

Table 6: Example ADaM ADVS for one subject, using ADSL for reference values

USUBJID	SUBJID	PARAM	AVAL	ADT	TR01SDT	ADY	TRTP
ABC-01	ABC-01	Height (cm)	167	01MAY2023	02MAY2023	-1	A
ABC-01	ABC-01	Weight (kg)	68	01MAY2023	02MAY2023	-1	A
ABC-01	ABC-01	Weight (kg)	65	22MAY2023	02MAY2023	21	A
ABC-01	ABC-25	Height (cm)	167	16OCT2023	15OCT2023	-1	B
ABC-01	ABC-25	Weight (kg)	66	16OCT2023	15OCT2023	-1	B
ABC-01	ABC-25	Weight (kg)	67	30OCT2023	15OCT2023	16	B

Table 7: Example ADaM ADVS for one subject, using ADPL for reference values

Notice that in Table 6, when using ADSL for the single treatment start date, the values of ADY and TRTP are probably not what we want for the second participation. Table 7, which uses the treatment name and start date for each participation from ADPL, is more likely what would be needed for analysis.

INTEGRATED ANALYSIS AND ADPL

Could this proposed ADPL be a solution for the integration impasse? It certainly has some advantages:

1. ADSL remains unchanged as one record per subject, so that seems to meet the FDA needs.
2. ADPL is useful for any analyses that need information about each participation, such as merging onto datasets that contain multiple participations.
3. SUBJID is already allowed on any ADaM dataset.

PARTICIPATION VS. POOL

One difference between iADSL and ADPL is that iADSL was structured as one record per subject per pool, and ADPL is structured as one record per subject per participation. Are participation and pool equivalent enough that we can interchange them? Let's look at an example where we need 2 different pools: double-blind and open-label.

First, consider the subject we've been looking at above. Let's say this subject's first participation was a double-blind study and second participation was an open-label study. In Figure 1, we see that one record per participation and one record per pool have the same net result.

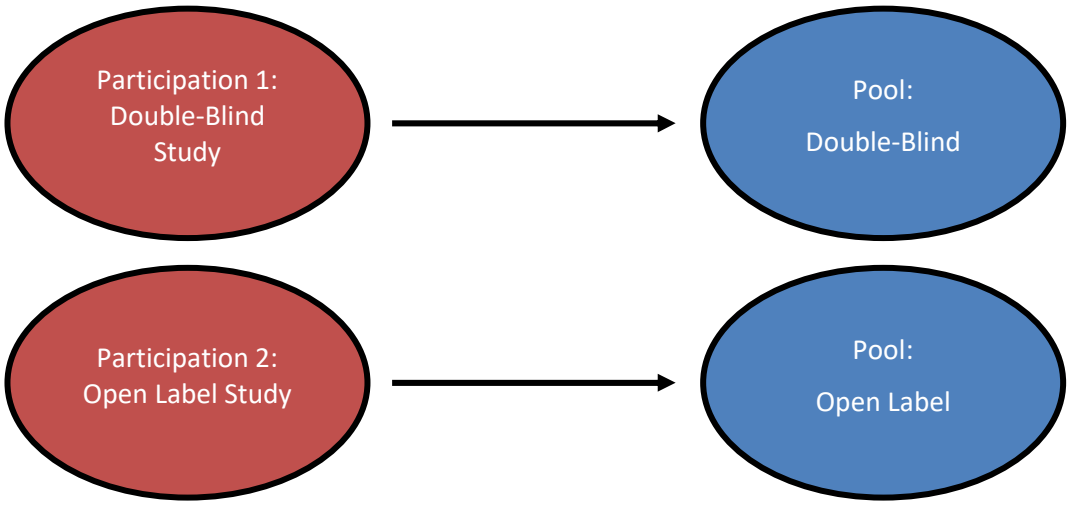

Figure 1: Each Participation Maps to a Separate Pool

Consider another subject who enrolled in a single study that included two arms: double-blind and open-label extension. In Figure 2: One Participation Maps to More than one PoolFigure 2 we see that there would be two records needed when talking of pooling (one for double-blind and one for open label) but only one record when talking of study participation.

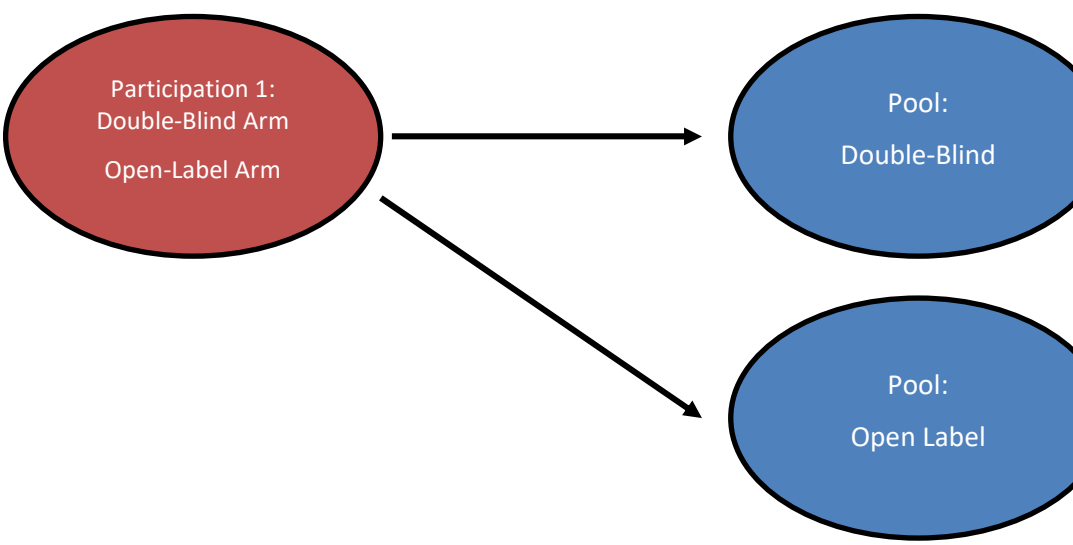

Figure 2: One Participation Maps to More than one Pool

So, it doesn't seem like ADPL will work for all integrations, but is that the end of the story?

PROPOSAL FOR AN INTEGRATION POOL-LEVEL ANALYSIS DATASET

While ADPL might not work directly for integration, the same concepts could be used. Instead of setting up a dataset in addition to ADSL that is one record per participation, why not create one that is one record per integration pool? I'm going to call this ADIPL (Integration Pool-Level Analysis Dataset), rather than ADPL (Pool-Level Analysis Dataset) to distinguish it from the ADPL described above. I'm also making use of variable POOLID, which was used in the draft ADaM Integration document.

Let's look at some example datasets for three subjects:

- Table 8 shows the subject who was only in a double-blind study,
- Table 9 shows the subject who was in a two-arm study: first double-blind and then open label, and
- The next 2 tables show the subject who was first in a double-blind study (Table 10) then later in an open label study (Table 11):

STUDYID	USUBJID	SUBJID	TR01SDT	TRT01P
AAA	AAA-01	AAA-01	02FEB2023	A

Table 8: ADSL for Subject 1 in a Double-Blind Study

STUDYID	USUBJID	SUBJID	TR01SDT	TRT01P	TR02SDT	TRT02P
BBB	BBB-02	BBB-02	02APR2023	A	15JUN2023	B

Table 9: ADSL for Subject 2 in a Two Arm Study

STUDYID	USUBJID	SUBJID	TR01SDT	TRT01P
AAA	AAA-03	AAA-03	10MAY2023	B

Table 10: ADSL for Subject 3 in a Double-Blind Study

STUDYID	USUBJID	SUBJID	TR01SDT	TRT01P
CCC	AAA-03	CCC-05	02SEP2023	A

Table 11: ADSL for Subject 3 in an Open Label Study

Table 12 combines the 4 prior tables into a single ADIPL dataset:

STUDYID	USUBJID	SUBJID	POOLID	TR01SDT	TRT01P
AAA	AAA-01	AAA-01	DB	02FEB2023	A
BBB	BBB-02	BBB-02	DB	02APR2023	A
BBB	BBB-02	BBB-02	OL	15JUN2023	B
AAA	AAA-03	AAA-03	DB	10MAY2023	B
CCC	AAA-03	CCC-05	OL	02SEP2023	A

Table 12: ADaM ADIPL for three subjects

Note that in Table 12, the POOLID value of "DB" was used for double-blind and "OL" for open label.

In addition to ADIPL, an ADSL would be needed. Perhaps something like Table 13:

USUBJID	TRTSEQP	TR01SDT	TRT01P
AAA-01	A	02FEB2023	A
BBB-02	A-B	02APR2023	A
AAA-03	B-A	10MAY2023	B

Table 13: ADaM ADSP for three subjects

How to populate ADSL required variables STUDYID and SUBJID would need to be determined. Since ADSL was written for use with individual studies, perhaps they don't make sense to include in a one record per subject integration ADSL? That discussion is beyond the scope of what I want to cover in this paper, but possibly something to consider in a future ADaM Integration document.

ADIPL, POOLID AND COMPLIANCE

There is nothing preventing the use of an ADIPL dataset, as described above, right now. Specifically,

1. Because ADIPL would be used in addition to ADSL, it doesn't break the rule to include an ADSL dataset.
2. There is no standard ADaM class for ADIPL, but it could be called ADAM OTHER. Perhaps in the future ADaM will add a class for a one record per subject per pool dataset.
3. Variable POOLID can be added to ADaM datasets without issue. This will allow for merging between ADIPL and any other ADaM dataset used for integration.

In BDS datasets you'll need to watch out for cases where a different baseline is needed for each pool. Using existing ADaM BDS variable BASETYPE, with a different value for each pool, will avoid this issue. For example, looking at just subject 2 from Table 12, we can see in Table 14 that there are two baselines within SUBJID and PARAM, one for each pool, so this dataset uses a BASETYPE that is just a copy of POOLID.

SUBJID	PARAM	POOLID	TRTP	AVAL	ADT	ADY	ABLFL	BASETYPE
BBB-02	Weight (kg)	DB	A	72	01APR2023	-1	Y	DB
BBB-02	Weight (kg)	DB	A	70	22MAY2023	51		DB
BBB-02	Weight (kg)	OL	B	71	13JUN2023	-2	Y	OL
BBB-02	Weight (kg)	OL	B	73	30OCT2023	46		OL

Table 14: ADVS for one Subject in Multiple Pools

Note that this use of BASETYPE is similar to how you might handle a second baseline in a crossover study, where the first treatment period uses an initial baseline, and the second treatment period uses a new baseline.

CONCLUSION

Based on FDA feedback, there seems to be a need for an integration solution that maintains a one record per unique subject (ADSL) dataset but also allows for capturing content used for analysis of multiple pools. Building on the ADaM proposed ADPL (Participation-Level Analysis Dataset), which built on the SDTM proposed DC (Demographics as Collected), a dataset structured as one record per unique subject per pool (which I'm calling ADIPL, or Integration Pool-Level Analysis Dataset) could be used in addition to ADSL. ADIPL and any dataset to be analyzed by pool would need to contain a variable like POOLID to show which records belong to which pool.

Dataset ADIPL and variable POOLID can be used now, without breaking any ADaM rules. For submission, ADIPL would be of class ADAM OTHER, since there is currently no standard structure that is

one record per subject per pool, though I hope in future there would be an ADaM class created to meet this need.

In datasets where there is a different baseline record needed for each pool within a subject, make use of variable BASETYPE to show which baseline applies to each record.

ACKNOWLEDGMENTS

I would like to thank Kent Letourneau, who, at our 2023 presentation on ADPL, asked the question on whether that dataset structure could be used for integration. It got me thinking!

REFERENCES

CDISC ADaM documents available at https://www.cdisc.org/standards/foundational/adam.

FDA Study Data Technical Conformance Guide available at https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources.

Dennis, E. et al. “ADaM Datasets with Multiple Participations per Subject”. *PharmaSUG Conference*. May, 2023. Available at https://www.pharmasug.org/2023-proceedings.html.

Zhong, W. et al. “ADaM Structures for Integration: A Preview”. *PharmaSUG Conference*. May, 2018. Available at https://www.pharmasug.org/2018-proceedings.html.

CONTACT INFORMATION

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

Sandra Minjoe
Senior Principal Clinical Data Standards Consultant
Clinical Data Science IOD
ICON plc
Sandra.Minjoe@iconplc.com