PharmaSUG 2018 – Paper SI14

Integrations Made Easier

Alyssa Wittle, Covance, Inc., King of Prussia, PA
Kirsty Lauderdale, Covance, Inc., Oakville, Ontario

ABSTRACT

Too often we are found to have an integration of a number of studies that potentially adds a few gray hairs for each of us. The documentation for data sources is missing, every study is a different standard, controlled terminology doesn't line up; the studies were so far in the past no one is quite sure *what* was done for the original, figuring out who actually was a roll-over subject, etc. The list can go on and on. Do not despair! There may be some tips that will help you to hold off on some of those grey hairs a bit longer!

This paper will present some tested and true tips for an integration with fewer headaches. Some of these tips consist of starting out with an integrated SDTM which combines study level SDTMs, particularly when dealing with lots of studies. If prepared well early, study level ADaMs could also be combined as another useful option. Controlled terminology is another hurdle, which has some easy to use methods to apply for your study and make studies you do over time cohesive. Site level analysis potential issues, roll-over subjects, and general mapping tips are among the topics which will also be presented.

Soon CDISC will release integration instructions, but until then, this paper offers a good starting point to make your integration quicker, more cost effective, and less stressful for everyone involved.

INTRODUCTION

Fire Chiefs create pre-planning documents to show how to attack a fire in their first-due response area. Teachers create lesson plans and syllabi to describe how the course objectives will be achieved. Coaches create play books describing the plays to use in order to win a game. Each of these areas plan ahead to accomplish a defined goal in an organized and thoughtful manner. Several scenarios are considered and one is chosen based on the requirements of the situation. When approaching the daunting task of a database integration for an Integrated Summary of Safety (ISS) or Integrated Summary of Efficacy (ISE), our strategy and approach should likewise be planned and handled in an organized approach which is clear to all participants. Read on for some tips to consider to attack an integration with ease rather than letting it become an inferno.

THE BEST PLACE TO START IS AT THE BEGINNING

The idea of starting at the beginning sounds logical enough, but it isn't always feasible. Sometimes studies are thrown into an integration without any thought ever have been given to combining it with another study. If given the chance to start from the very beginning, or perhaps having to create a plan somewhere in the middle of the game, here are some items to consider:

Study Data Standardization Plan (SDSP)

In an integration of any scale, planning is the first step to eliminate stress and chaos. The Study Data Standardization Plan (SDSP) should be the most frequently referenced document throughout the course of an integration and it should have an answer to virtually every question in regard to the studies involved in an integration and the steps which will be taken to transform them into a cohesive integrated database. Many of the tips referenced in this paper are worthwhile mentioning in the SDSP, or describe how that particular item will be done for the integration in question.

CDASH CRFs

The earlier in the process an integration can be considered, the better. If it is feasible to start planning for an integration before a study ever starts it will make life much easier. CDASH provides guidance for CRFs which allows them to be easily transformed into SDTM. Controlled Terminology is taken into account on the CRF itself. Databases have similar raw variables, similar values, visit values, page names all leading to similar SDTM mapping!

Starting an integration from similarities at this level is the dream and it does make for a less stressful experience, relatively.

Study Site Catalog

If any site level analysis is planned, or could eventually be planned, for an integration, it is important to maintain detailed information for each site that is used over the course of those studies. The earlier this document begins to be compiled, the easier it will be at the analysis level. For example, going back five years after the fact to determine if site 10001 which appears differently as "Seattle Grace Hospital" for Study A, "Seattle Grace Mercy West Hospital" for Study B, and in Study C as "Grey + Sloan Memorial Hospital" due to mergers, rebranding, etc. can be time consuming and tedious if the site numbers are not kept consistent across all studies for a single treatment or even for a single sponsor!

Study Level SDTMs

Whether or not the opportunity was available to gain a head start with CDASH CRFs, the next area with the potential to reduce the chaos of an integration is at the study level SDTMs. If there is the opportunity to start your consistent mapping strategy with each and every study at the SDTM level, it can aid an integration by leaps and bounds. Here are some items to keep in mind:

Consistent SDTM Implementation Guide (SDTM IG) version

Review whether you need to use the best and most recent IG, as it may not make the most sense for your study to do so. That is not to say, that if a new standardized domain which perfectly fits what you need for your study that you can't upgrade for a few studies. However, for an integration the pros and cons would need to be weighed to see if utilizing the new standard would assist or impede the effort.

Consistency in Domain Mapping

Domain mapping is an important part of the process in an integration which allows the project team to know where to look for each study to find a piece of information. Some data we expect in every study, but occasionally there is an unusual value collected which may be indication, sponsor, or study specific which requires custom mapping. In order to make the integration easier, try to keep these custom domains or locations as consistent as possible. This effort can be aided, often, by the use of CDISC's Therapeutic Area User Guides (TAUGs) which have recommended locations for indication specific data.

Consistency in Controlled Terminology (CT) Version

New versions of Controlled Terminology are released quarterly which makes it very easy to have a different version of controlled terminology for every single study in an integration. Rather than up-versioning for every study for a specific treatment, pick the latest version at the start of the first study and stick with it. Similar to choosing versions of the SDTMIG to use, if a newer version of CT becomes available which has some terms which would be very useful in your study then the pros and cons would have to be discussed to determine if it is a worthwhile change with an integration in mind.

Consistency in SDTM Value Level Mapping

Being able to find data doesn't just stop at the domain name level. Particularly for large domains it is helpful to have a dictionary of values for which all of the studies in an integration should be mapped. The following are some examples of value level mapping to watch out for:

- **xxTEST/xxTESTCD:** While there may be many codelists available to map tests and testcodes, when adding custom values be sure to map consistently so that the exact same wording and abbreviations are used across studies in the naming of xxTEST and xxTESTCD.

- **QNAM/QLABEL:** These variables have no controlled terminology, but since in an integration these variables may be set together either in integrated SDTM or as variables in integrated ADaM, it is much easier to have logical and consistent naming for values in the study level SDTM. If SDTMs can't be completely consistent at this stage it is helpful in an integration to have meaningful QNAM and QLABEL values rather than the potentially lacking rawdata

variable name – 5 years down the road seeing QNAM=DATE1 will not be particularly helpful for someone trying to integrate!

- **Extensible Codelists:** An attempt should be made to keep any custom values added to an extensible codelist as consistent as possible over every study for a given treatment. While some of these might not be feasible, for example AE Relationship to Study Drug collected differently on the CRF, other values may be able to have some consistency or very intentional differences, such as EPOCH for example.

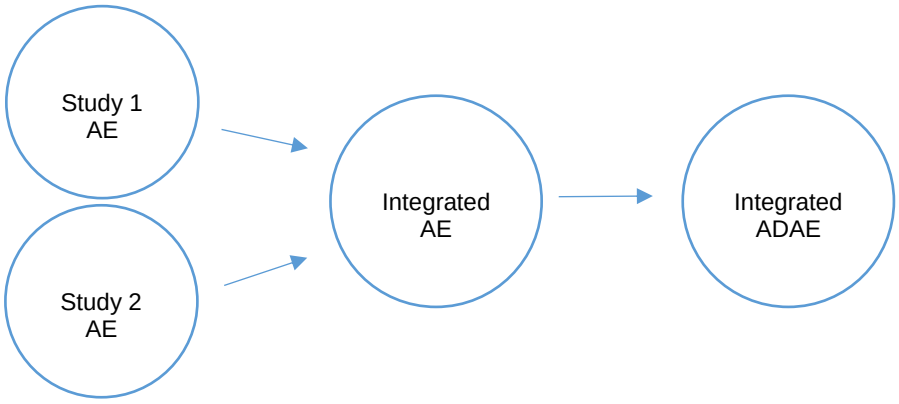
- **Categories/Subcategories:** Some categories and subcategories have codelists available (e.g. DSCAT), most do not. It is helpful to keep naming of categories consistent for the naming of the same data across studies.

- **Standardized Units:** In any study, standardizing units across different rawdata sources can be a pain, but throw in multiple studies *all* with several rawdata sources and it rapidly becomes a nightmare if it is not planned for and documented well. Using a sponsor level or treatment level collection of individual tests, which studies they are found in, and the standardized unit of choice for that test can not only help at an integration level, but it can also improve the speed and quality of mapping when all information of this type is found in the same document.

THREE ROADS DIVERGED IN A WOOD

Robert Frost's famous poem "The Road Not Taken" describes two roads diverging in a forest and picking the less-travelled road, "that has made all the difference". Determining the most effective way to create an integrated database can 'make all the difference' for a study, but rather than two roads diverging, integrations have three options which each have pros and cons. It is important to review the situation for every integration and determine the best route to take for that particular project.

Option #1: Integrated SDTM from Study Level SDTM



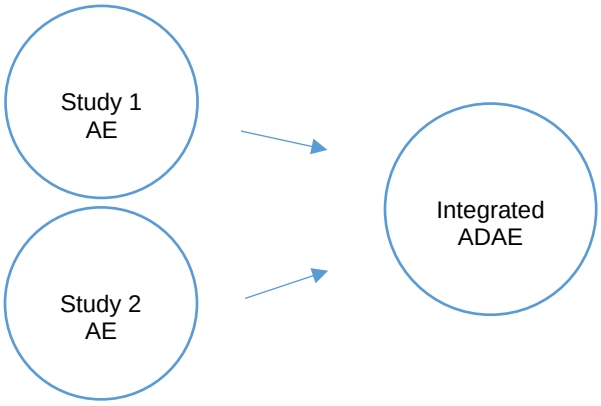
The first option for creating an integrated a database is to take study level SDTMs and create a set of integrated SDTMs which then create integrated ADaMs. This method is particularly useful when the studies which are being integrated have been done over time potentially with different versions of the SDTM IG, different versions of controlled terminology, and different mapping of test, testcode and supplemental qualifier variable values. If this situation arises then it is ideal to make a single integrated SDTM database in order to have everything under 1 IG version, and the same values mapped to a single place.

This method, with the type of database described, allows for much easier research when issues arise and greatly simplifies the programming of the integrated ADaMs compared to the alternative options below. Imagine having upwards of 50 studies programmed completely independently and trying to standardize the values plus programming ISS or ISE analyses all at the ADaM level – the programs would be enormous! Researching any issue that came up

in tables would take an extended amount of time just to find the right value to query! Instead, having a single set of SDTMs which standardizes everything is far less of a headache for everyone involved.

Using this approach eliminates the need for every study to have SDTM and/or ADaM which is often the situation when integrating historical studies.

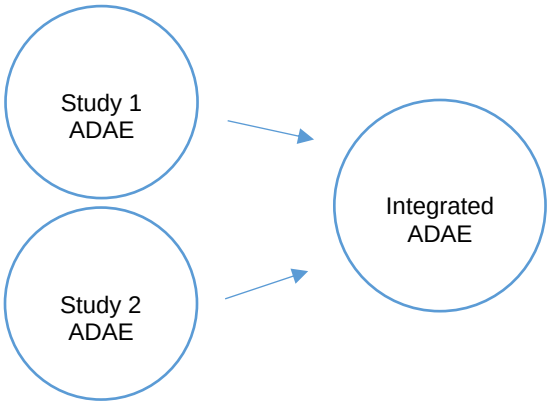
Option #2: Integrated ADaM from Study Level SDTM



The second option for integrating multiple studies into a single database is to use the study level SDTMs as the source for the set of integrated ADaMs. This option eliminates the use of the intermediate integrated SDTMs from Option #1. If the SDTMs are not made with the intent to eventually integrate, this can lead to very messy ADaM programming and review of the ISS/ISE.

With this approach, time saved by eliminating integrated SDTMs may be replaced with time spent QCing ADaMs and TFLs. This may be a good approach to use if SDTMs are created for every study in the integration (preferably using the tips in the first section of this paper), but ADaMs are not created for every study.

Option #3 Integrated ADaM from Study Level ADaM



The third and final option for creating an integrated database is to combine study level ADaM into a single set of integrated ADaMs. This approach is particularly useful if the individual studies were created with the end-goal of an integration in mind. When integrating study level ADaMs it is especially important to follow the tips in the first section of this paper so that test, testcd, parameter, visits, categories, formats, numbering, etc. are all created with the intent to integrate ensuring they are already consistent. When that is the situation, then integrating from the ADaM level consists mostly of setting together the various study level ADaMs and doing any new programming for integrated analyses which may not already be included in the study level ADaMs.

If a project team plans from the beginning of a project for a single treatment, this option quickly becomes not only most efficient integration, but also the easiest on the programmers. Of course, this approach does take large amounts of pre-planning and team coordination to accomplish effectively. It also requires the team to have study level SDTM and ADaMs available to integrate, preferably all of which would follow the tips in the first section of this paper.

EXPECT THE UNEXPECTED

Roll-Over Subjects

One of the primary rules of CDISC is that each unique subject maintains the same USUBJID value throughout every study of which they participate. That means, if a subject participates in Study ABC as subject ABC-1001-001 and they also participate in Study XYZ, their USUBJID must remain ABC-1001-001 across both studies. This becomes important for integrations so that, to the best of the sponsor's ability, the path of a subject across all studies, can be traced from beginning to end. But then, you ask, a demographics domain must have exactly 1 record per subject, correct? For a single study – yes, absolutely. However, the SDTM and ADaM teams at CDISC have yet to publish final guidance on how to handle integrations. An option in this situation, if an integration opts to have a set of integrated SDTMs, then uniqueness would be pushed to unique per STUDYID and USUBJID to accurately display each study in which a subject participated, while maintaining the same USUBJID across the various STUDYID values. At the ADaM level, similarly one record per study would be maintained in ADSL and then perhaps an overall analysis record would be created which would have their overall time spent on treatment in the analysis variables.

Unless the inclusion/exclusion criteria for every study specifies that a potential subject cannot have participated in a study for the investigational treatment for a non-extension study, it is worthwhile to do a check of the data to see if there are any subjects to be considered further to determine if they are indeed a roll-over subject. Basic demographics information can be used to get a general list such as sex, race, ethnicity, site, date of birth, etc. If there are any suspicious subjects, then the sponsor can investigate further into each study to determine if there are any roll-over subjects.

Compliance Rules

Just as SDTM and ADaM teams have not released guidance for integrated studies, the related CDISC teams who release compliance rules also have not released guidance for checking compliance of integrated studies. Though Pinnacle 21 may be helpful in checking some general compliance items at a value level and domain level, any issues which come down to subjects and study levels may not be checked accurately with Pinnacle 21 – especially when roll-over subjects are involved. This means that you will likely need to go back to making your own checks in some cases by taking the list of rules published by CDISC and tweaking them for your needs on your study. This may be as simple as adding STUDYID anytime USUBJID is present in a check or be more complicated. Having a clean Pinnacle 21 report is great, but not always possible for integrations – and that is okay!

Consistent dictionaries

One aspect of an integration which can be done at any stage in the process is dictionary versioning. MedDRA and WHO Drug versions should be re-coded in an integration so that every study has consistent versions for table summaries. This can be done at the integrated SDTM level, while keeping the original study level coding values in supplemental qualifiers, if so desired, as a way to have the information easily accessible when comparing integrated TLFs to study level TLFs. It also can be done in a similar fashion at the ADaM level using the variables created for exactly this purpose in Table 3.2.10.1 and Table 3.2.10.2 in the ADaM Occurrence Data Structure (OCCDS) version 1.0 document.

Traceability

In any integration, one of the most important part is to be able to know exactly where any given data point came from originally. Maintaining this traceability is simple when using CDISC standards. ADaM specifies that any SDTM variable, with the exception of dictionary coding variables described above, maintain the rule same variable – same value. This means that something like DM.AGE pulled into ADSL.AGE must be an exact copy of the SDTMs for not only the variable name, but also the variable value and label. If a new analysis value is needed for the integration

which differs for any study from the SDTM then ADSL.AAGE would need to be created to hold this value. This situation is the same for any SDTM value which may not match analysis needs. In addition, an easy way to provide a quick reference back to SDTM level is by maintaining the SDTM sequence number either by name: AESEQ or by using the ADaM variables of SRCDOM and SRCSEQ. It is also essential to have easy to follow specifications which allow programmers and reviewers to easily trace exactly what transformations occurred to any value along the course of a study. Finally, the SDSP will be the most important document to aid traceability in the study and should show the data flow for each study as it gets integrated.

CONCLUSION

Database integrations can be daunting tasks and have the potential to be very messy. However, with careful planning and picking the best route for success, the end goal of a successful ISS and/or ISE can be easily within reach. Be pro-active – consider an eventual integration as far in advance as possible. Document and be consistent so that details and decisions can be easily referred back to as team members change. Finally, choose the route which makes the most sense for your integration. Always expect the unexpected twists and turns in the road, no trip is without it's bumps and sometimes you are forced to take a detour. No two integrations are the same, but all can be successful. After all, *three* roads diverged in a wood and I – I took the one less travelled by, and that has made all the difference.

CONTACT INFORMATION

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

Name: Alyssa M Wittle
Enterprise: Covance, Inc.
Address: 1016 West Ninth Avenue
City, State ZIP: King of Prussia, PA 19406
Work Phone: 717-449-9765
E-mail: Alyssa.Wittle@Chiltern.com

Name: Kirsty Lauderdale
Enterprise: Covance, Inc.
Address: 1155 North Service Road West, Unit 11
City, State ZIP: Oakville, ON, ON L6M 3E3, Canada
Work Phone: 778-819-6517
E-mail: Kirsty.Lauderdale@Chiltern.com