

## Associated Persons Domains – Who? What? Where? When? Why? How?

Alyssa Wittle, Chiltern, King of Prussia, PA

Michael Stackhouse, Chiltern, Cary, NC

### ABSTRACT

Many types of clinical studies collect information on people other than the person in the study. Family medical history – MH or a custom domain? What about organ donor data? Care-giver information? Until now we have there was no home in the CDISC SDTM structure to standardize this information. With the publication of the Associated Persons SDTM IG this standardization has arrived! However, since it is so new, many don't know that it exists and/or how to apply it.

This presentation will explore the Associated Persons SDTM Implementation Guide by exploring six investigational questions to make the application of the AP SDTM IG very logical and even easy. During the presentation we will discuss “What” the Associated Person IG is and for “Who” are these domains intended? “When” do these domains apply in a study and “Where” do I put all the information? One of the biggest questions that will be discussed is “How” to apply the information from the IG. Finally, we will discuss “Why” these domains are a necessary addition and why they should be applied to all studies wherever they are applicable.

In order to accomplish discussing all of these questions the presentation will include many applied examples. It will progress from study structure to CRFs and annotations to translating the information from the raw data into the Associated Persons domain structure. While the transition to using Associated Persons domains may seem complex for many, by the end of this presentation all attendees will have a thorough understanding of these domains and how to use them in their next clinical study.

### INTRODUCTION

Have you ever had the experience of seeing a new standards implementation guide (IG) from CDISC and not immediately understanding how it is applicable to your clinical trial? Or, perhaps you can see the applications, but are not fully understanding the purpose of each of the standards? The Associated Persons (AP) implementation guide is an excellent example of when these questions arise. What is an associated person? Why should you care about this standard when you already have SDTM? When do you use these new standards? Who fits into this category? Where does the data go? How do you make it happen? All of that and more will be explained over the next few pages!

### WHAT IS AN ASSOCIATED PERSON?

First, what is the SDTM Associated Persons Implementation Guide? Or perhaps a better question is: what is an Associated Person? The Associated Persons IG gives a standardized home for non-subject specific data collected during a study. Since this data is not actually about the study subject, it does not belong in a subject level SDTM. Therefore, CDISC created a place to store this data – within the Associated Persons domains.

### WHEN IS IT APPLICABLE?

When is this actually applicable to your study? Some common applications are family medical history and family substance abuse history. Associated persons domains are also used for accidental exposure data – a study treatment accidentally applied to someone who was not in the study (whether they have a relationship to the study subject or not) and any adverse events which are due to accidental exposure. There might also be AEs due to device malfunction which would impact a non-study subject. Donor information which is collected would be stored in an Associated Person domain as well as any information collected regarding a caretaker or a sexual partner. Environmental exposure information might be collected such as the smoking habits of someone who lives in the same household as the study subject. Finally, diagnostic samples may be collected when the topic of a study is an investigational device and the data would be regarding the tissue or blood sample used in a diagnostic test. These are only a few examples among many applications but hopefully this gives you a good idea to get you in the frame of mind for these domains.

### WHO IS AN ASSOCIATED PERSON?

So, *who* is an associated person? An associated person can be associated with an individual subject OR the study itself. It is not a requirement for an individual to have a relationship with a specific study subject in order to be

considered an associated person. For example, if the wrong patient in a hospital received the study medication, then they do not have a relationship to a particular study subject - they are just related to the study itself. The data in AP domains needs to be *about* the associated person. This means if a person *maintained* the data but the data isn't actually *about* that person, then that person is not an associated person. However, if the CRF collected specific information about someone other than the study subject, such as medical history, age, race, etc., then that data, and *only* that data, would be stored in an AP domain. Note also that an associated person is actually a *person*, that is, a *human*. This means data such as environmental exposure for a type of allergy would not be mapped in these domains.

## WHERE DOES THE DATA GO IN SDTM?

Next we want to determine where all of this information is kept in a standardized structure. Associated Persons models can be created based off of traditional SDTM domains. The datasets themselves get a prefix of "AP" added to the SDTM domain you are creating in order to identify these datasets as Associated Persons – yes, that means you will have dataset names that are four characters long, such as APDM. Also, one big departure from SDTM is that Associated Persons domains are only used as needed. What this means is that you only have the datasets for which you actually have data. If you don't have any demographics data collected for your study, then you will not have an APDM domain, which is acceptable. Finally, supplemental qualifiers for AP domains have slightly different naming characteristics than traditional SDTMs: you will have a prefix of "SQAP" rather than "SUPP".

## HOW TO MAKE IT HAPPEN

Now that we have a great basis and background for Associated Persons domains, let's finally discuss *how* to make it happen, starting with some general information regarding the characteristics of these domains! The domain labels for AP domains must start with "Associated Persons". Any kind of study drug exposure or protocol driven timing variables are generally not used in Associated Persons domains. This would include many DM variables, such as ARM, ARMCD, ACTARM, ACTARMCD, RFXSTDT, RFXENDTC, RFSTDT, RFENDTC, RFICDT, and RFPENDTC you would generally see in DM but not in APDM. Logically this makes sense since you would not expect associated persons to have any exposure or disposition event-type data for the study. Also, any regular SDTM variable will keep the xx prefix from APxx. For example, DMDTC is still DMDTC, not APDMDTC. However, the DOMAIN value should match the associated persons dataset name. Caution must be taken for the use of timing variables. Since associated persons are not directly associated with the timing structure defined in the protocol, timing variables, such as visit, need to provide context to the protocol for the study subject - not the associated person.

AP domains have a set structure of variables, with normal SDTM variables such as STUDYID, DOMAIN and --SEQ. In addition to these, AP specific core variables include: APID, RSUBJID, RDEVID, and SREL. Figure 1 has the spec for AP specific variables in any SDTM.

Variable Name	Variable Label	Type	Controlled Terms, Codelist, or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char		Identifier	Four-character abbreviation for the domain, AP--.	Req
APID	Associated Persons Identifier	Char		Identifier <i>Except in APDM where it is the Topic variable</i>	Identifier for a single associated person, a group of associated persons, or a pool of associated persons. If APID identifies a pool, POOLDEF records must exist for each associated person.	Req
--SEQ	Sequence Number	Num		Identifier	Sequence number to ensure uniqueness of records within a domain for associated persons. Unique within APID within the domain.	Req
RSUBJID	Related Subject	Char		Record Qualifier	Identifier for a related study subject or pool of study subjects. The subject(s) may be human or animal. RSUBJID will be populated with the USUBJID of the related subject or the POOLID of the related pool. RSUBJID will be null for data about associated persons who are related to the study but not to any of the study subjects.	Exp
RDEVID	Related Device	Char		Record Qualifier	Identifier for a related device. RDEVID will be populated with the SPDEVID of the related device	Perm
SREL	Subject, Device, or Study Relationship	Char	(RELSUB)	Record Qualifier	If RSUBJID is populated, describes the relationship of the associated person(s) identified in APID to the subject or pool identified in RSUBJID. If RDEVID is populated, describes the relationship of the associated person(s) identified in APID to the device identified in RDEVID. If RSUBJID and RDEVID are null, SREL describes the relationship of the associated person(s) identified in APID to the study identified in STUDYID.	Req

Figure 1. Spec for AP specific variables in any SDTM

APID is similar to a USUBJID with some key differences. First, APID does not need to be unique outside of a study. When there is no relationship or the relationship is unknown, you still need to have distinct APIDs across datasets. Ensure that if you have a subject in APEX and APAE that you retain the same APID across datasets, even if the subject is only related to the study. A final important item to note with APID is that it defines a single unit. A unit can be defined as a single person OR as a group of people.

When an APID is a group of people we consider this a Pooled APID. If you have a pooled APID, you are required to have a POOLDEF dataset. This dataset identifies the individual persons defined within a pooled APID. In Figure 3, there is a combination of Pooled and individual APIDs defined in the AP dataset on the left. These pooled values translate as shown to the POOLDEF dataset on the right which has a very simple structure. POOLDEF lists out which APIDs belong to each POOLID. Take notice also in this dataset that it is possible that a particular APID can have individual APIDs and relationships to a study subject *and* be part of a Pooled ID, such as APIDs AP001 through AP005.

Row	STUDYID	DOMAIN	APID	LBSEQ	RSUBJID	RDEVID	SREL
1	ABC	APLB	POOL1	1		GRNMO13	DONOR, SAMPLE
2	ABC	APLB	POOL2	1		GRNMO13	DONOR, SAMPLE
3	ABC	APLB	AP001	1		CK001	DONOR, SAMPLE
4	ABC	APLB	AP002	2		CK001	DONOR, SAMPLE
5	ABC	APLB	AP003	3		CK001	DONOR, SAMPLE
6	ABC	APLB	AP004	4		CK001	DONOR, SAMPLE
7	ABC	APLB	AP005	5		CK001	DONOR, SAMPLE

  

Row	STUDYID	POOLID	APID
1	ABC	POOL1	AP001
2	ABC	POOL1	AP002
3	ABC	POOL1	AP003
4	ABC	POOL1	AP004
5	ABC	POOL1	AP005
6	ABC	POOL2	AP006
7	ABC	POOL2	AP007
8	ABC	POOL2	AP008
9	ABC	POOL2	AP009
10	ABC	POOL2	AP010

Figure 2. On left: APLB with pooled APIDs. On right: POOLDEF dataset

Relationships in Associated Persons domains are handled in SREL which contains the relationship, if there is any, to a study subject. You can pull these values from the RELSUB Controlled Terminology codelist. If there is no relationship the subject, or just an accidental relationship to the study, you would use "ACCIDENTAL ASSOCIATION" as the value for SREL. You will find that there are similar values used between SREL and a typical EVAL variable. SREL can also account for multiple relationships, in which case the value of SREL is "MULTIPLE" and an APRELSUB table is required. APRELSUB contains the definitions for each of the individual relationships. One associated person can have multiple relationships to a single subject, OR a single associated person could be related to multiple subjects. You should determine the SREL classifications of relationships using the controlled terminology, and this will often make APRELSUB necessary. Be sure to make your APID in such a way that it simplifies the APRELSUB process. Figure 3 is an example of an APRELSUB table. This looks very similar to the general structure we have for AP domains with only variables that have to do with relationships between Associated Persons. Note that domain and SEQ variables have dropped from this dataset model. If the study is a device study then it is possible for RDEVID to also be included in this table following RSUBJID.

There are many logical ways for a subject to have multiple relationships. In Figure 3 within the yellow box there are five different associated persons who are related to the first subject. Next, in the red box we have a biological aunt who is also a caregiver to subject AMR\_001\_02, and APID A006 is also subject AMR\_001\_02's mother. But, as we can see in the green box, APID A006 is related to the next subject as well. Though they are the mother to subject AMR\_001\_02, they are also the biological paternal cousin to subject 3! In the purple box, Subject 7 has two associated persons, one who is a caregiver but also a biological aunt *and* a step-mother, and another associated person who is their dizygotic twin brother. Clearly, it is quite common to have multiple relationships in AP domains and that they are simple to handle by the use of controlled terminology.

Rounding out details regarding specific variables are some quick points on RSUBJID and RDEVID. First, both of these variables are null if the associated person is associated with the study rather than a subject or device. For RSUBJID, as I mentioned before, you want to populate this with the USUBJID of the related subject. As we now know, it is possible that you might have a single associated person related to multiple study subjects in which case you would have a Pooled ID in RSUBJID with the related subjects defined in POOLDEF. Finally, RDEVID, will be populated with SPDEVID for the related device. RDEVID is only an expected variable for device studies.

STUDYID	APID	RSUBJID	SREL
AMR_001	A0101	AMR_001_01	HOUSEHOLD MEMBER
AMR_001	A101	AMR_001_01	SIBLING, FULL
AMR_001	A102	AMR_001_01	HOUSEHOLD MEMBER
AMR_001	A103	AMR_001_01	HOUSEHOLD MEMBER
AMR_001	A104	AMR_001_01	HOUSEHOLD MEMBER
AMR_001	A005	AMR_001_02	CAREGIVER
AMR_001	A005	AMR_001_02	AUNT, BIOLOGICAL PATERNAL
AMR_001	A006	AMR_001_02	MOTHER
AMR_001	A006	AMR_001_03	COUSIN, BIOLOGICAL PATERNAL
AMR_001	A027	AMR_001_07	CAREGIVER
AMR_001	A027	AMR_001_07	AUNT, BIOLOGICAL MATERNAL
AMR_001	A027	AMR_001_07	MOTHER, STEP
AMR_001	A030	AMR_001_07	TWIN, DIZYGOTIC
AMR_001	A030	AMR_001_07	BROTHER, FULL
AMR_001	A067	AMR_001_56	SISTER, BIOLOGICAL
AMR_001	A068	AMR_001_56	MOTHER, BIOLOGICAL
AMR_001	A456	AMR_001_58	CAREGIVER
AMR_001	A456	AMR_001_58	FAMILY MEMBER
AMR_001	A501	AMR_001_86	CAREGIVER
AMR_001	A501	AMR_001_86	HUSBAND
AMR_001	A551	AMR_001_51	DONOR, ORGAN
AMR_001	A551	AMR_001_82	DONOR, ORGAN
AMR_001	A551	AMR_001_82	RELATIVE BIOLOGICAL

Figure 3. APRELSUB table

**EXAMPLES OF HOW AP DOMAINS ARE USED**

**APMH – Associated Persons Medical History:**

History of:	Subject's Father's Family			Subject's Father		
	Yes	No	Unknown	Yes	No	Unknown
Heart Disease						
Stroke						
Cancer						
Depression						
Diabetes						
Sexually Transmitted Disease						
Alzheimer's Disease						
Obesity						
Blindness						
Deafness						
Allergies						
Asthma						
Congenital Abnormalities						
Learning Disabilities						
Psychomotor Retardation						
Subfertility/Infertility						
Miscarriages						
Stillbirths						
Infant deaths						

Figure 4. Family Medical History CRF

In Figure 4 there is a family history CRF. Note that the information being collected is in regard to the Subject's Father and Subject's father's family. These two values (circled in blue) translate to SREL using controlled terminology. The MHTERM becomes the medical history event (circled in red) for which we are collecting information. Then, just as we would for traditional SDTM, since all of these medical history terms are pre-specified, they would all have MHPRESP equal to "Y" and then the actual value collected would be in MHOCUR.

**APRP – Associated Persons Reproductive Findings:**

Currently pregnant?        YES    NO  
 Previous Pregnancies:  
 1. Number of full term live births:        \_\_\_\_\_  
 2. Number of premature births:        \_\_\_\_\_  
 3. Number of stillbirths:        \_\_\_\_\_  
 4. Number of late fetal deaths:        \_\_\_\_\_  
 5. Number of spontaneous abortions:        \_\_\_\_\_  
 6. Number of elective abortions:        \_\_\_\_\_  
 7. Number of ectopic pregnancies:        \_\_\_\_\_  
 Any congenital abnormalities / birth defects? YES    NO

**Figure 5. Partner Pregnancy Status**

In Figure 5 we can see the CRF collecting information for if the subject’s female partner became pregnant during the study. So, this associated person would be a sexual partner. This dataset’s structure from RPTESTCD, seen in Figure 6, is exactly how you would structure an RP domain. The only difference here are the AP variables which show that this information is not about the subject themselves, but rather their sexual partner, and the ID for that associated person. This is another popular domain to have as we always want to ensure that any drug given to a Male is not going to cause pregnancy complications for his partner or birth defects for his child. You can imagine that a study for HIV would find this domain particularly useful.

STUDYID	DOMAIN	APID	RPSEQ	RSUBJID	SREL	RPTESTCD	RPTEST	RPORRES	RPSTRESC	RPDTC
2010_PR_34	APRP	PP_52	1	2010_PR_34_26	SEXUAL PARTNER	PREGST	Pregnant During the Study	Y	Y	2010-04-01

**Figure 6. APRP domain**

**APEX & APAE – Associated Persons Exposure and Associated Persons Adverse Events:**

Adverse events and study drug exposure are data that you may not expect to commonly see for an Associated Person, but they are both possible nonetheless. Consider the situation where a child may accidentally take a medicine prescribed to their parent or sibling. This would be a case for accidental exposure and certainly for any adverse events related to the exposure. In Figure 7 we can see how this might be collected in an APEX domain. The variables circled in green are the only structure differences compared to EX that you would expect to see in APEX. SREL tells us that the related person is a biological child and that they are related to subject ABC001-003.

Row	STUDYID	DOMAIN	APID	EXSEQ	RSUBJID	SREL
1	ABC001	APEX	ABC001-AP-001	1	ABC001-003	CHILD, BIOLOGICAL

Row	EXTRT	EXDOSE	EXDOSU	EXDOSFRQ	EXROUTE	EXSTDTC	EXENDTC
1 (cont)	TRT1	10	mL	ONCE	TOPICAL	2014-12-18	2014-12-18

**Figure 7. APEX domain**

Then, in Figure 8 we can see the documented adverse event that is related to this exposure.

Row	STUDYID	DOMAIN	APID	AESEQ	RSUBJID	SREL	AETERM
1	ABC001	APAE	ABC001-AP-001	1	ABC001-003	CHILD, BIOLOGICAL	RASH

Row	AESEV	AESER	AEREL	AEOUT	AESHOSP	AESTDTC	AEENDTC	AEENRF
1 (cont)	MILD	Y	DEFINITELY RELATED	RECOVERING/RESOLVING	Y	2014-12-18		ONGOING

**Figure 8. APAE domain**

Since this situation is never planned for, a CRF is generally not something that you would have prior to the study start. Due to this, domains such as APAE and APEX are rarely *expected* domains, but rather urgent and highly important domains to correctly program.

## CONCLUSION

Now you are an associated persons expert! With the Associated Persons implementation guide, we have an elegant and standardized method of collecting data that formerly had no clear place. Once there is an understanding of the associated persons specific variables used in SDTM, you can make almost any SDTM into an Associated Persons domain when the need arises.

## REFERENCES

- CDISC Submission Data Standards Team. (2013, December 12). Study Data Tabulation Model Implementation Guide: Associated Person Version 1.0.

## CONTACT INFORMATION

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

Name: Alyssa Wittle  
Enterprise: Chiltern  
Address: 1016 West Ninth Avenue  
City, State ZIP: King of Prussia, PA 19406  
Work Phone: (717) 449-9765  
E-mail: [Alyssa.Wittle@Chiltern.com](mailto:Alyssa.Wittle@Chiltern.com)  
Web: [www.Chiltern.com](http://www.Chiltern.com)

Name: Michael Stackhouse  
Enterprise: Chiltern  
Address: 4000 Centregreen Way  
City, State ZIP: Cary, NC 27513  
Work Phone: (610) 513-4609  
E-mail: [Michael.Stackhouse@Chiltern.com](mailto:Michael.Stackhouse@Chiltern.com)  
Web: [www.Chiltern.com](http://www.Chiltern.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.