

Considerations in Data Modeling when Creating Supplemental Qualifiers Datasets in SDTM-Based Submissions

Fred Wood, Accenture Life Sciences, Wayne, PA

ABSTRACT

The Supplemental Qualifiers datasets described in the Study Data Tabulation Model (SDTM) and SDTM Implementation Guide (SDTMIG) provide a standard structure for the submission of non-standard variables (NSVs). Their naming follows the convention of SUPP--.xpt, where the two hyphens represent the two-letter domain code of the parent domain to which the NSVs belong. Despite the intended content of the SUPP-- datasets, instances have occurred where sponsors often include data in these that would actually be more accurately represented as either a separate domain or as additional records in the parent domain. Another challenge for sponsors lies in determining the merge key (represented in the IDVAR variable in SUPP-- datasets). When this is not chosen correctly, the values for the NSVs may appear on inappropriate parent records when data in the NSVs get merged onto the parent domain. In addition, cases have also been observed where sponsors have made extensive efforts to use multiple IDVAR values in the same SUPP-- dataset, when the resulting merge of the NSVs onto the parent records looks no different than if a simpler approach had been taken. This paper will present a number of actual examples where merging of the data in SUPP-- datasets onto the parent record has provided some unintended and possibly erroneous data representations.

INTRODUCTION

The Study Data Tabulation Model (SDTM) is the foundation for a number of implementation guides, including those for human clinical trials (the SDTMIG), nonclinical studies (the Standard for the Exchange of Nonclinical Data Implementation Guide), and the medical devices (the Medical Device Implementation Guide). Tables 2.1.1-2.2.5 of the SDTM describe standard variables that can be used in creating domains based upon the three General Observation Classes (Interventions, Events, and Findings). Table 2.2.6 describes the standard variables for the Demographics domain.

The Supplemental Qualifiers structure, described in the Table 4.1.2 of the SDTM and in sections of the implementation guides, was created to address the need to represent non-standard variables (NSVs -- those not found in SDTM Tables 2.2.1-2.2.6). The current submission format for NSVs is as separate Supplemental Qualifiers datasets named SUPP--, where the two hyphens represent the two-letter domain code of the "parent domain" to which the NSVs belong. When the SDTM and SDTMIG were first created, the thinking was that all NSVs would be submitted in one datasets per study, and this would be called SUPQUAL. Although it was soon realized a single dataset would not be practical, this name is still often used to describe both NSVs ("SUPQUALS") as well as the SUPP-- datasets themselves.

SUPPLEMENTAL QUALIFIER BASICS

In order to understand some of the potential modeling issues described in this paper, it's helpful to understand the structure of the SUPP-- datasets and how the data in them can be related to (and/or merged onto) the appropriate records in the parent datasets. The structure uses STUDYID (Study ID), USUBJID (Unique Subject Identifier), RDOMAIN (Related Domain), IDVAR (Identifying Variable), and IDVARVAL (Identifying Variable Value) to point to the related record (the relationship). The QNAM and QLABEL columns describe the name and label, respectively, of the NSV in the sponsor's original dataset. In the example below, QVAL contains the value on the record in the MH (Medical History) domain for Subject 026-0001 in Study 2006-026 whose MHSEQ value is 8. There are two additional columns, in the SUPP-- dataset structure: QORIG (Origin), which is really value-level metadata, and QEVAL (Evaluator), which is really an additional piece of data equivalent to the --EVAL variable in the Findings general observation class.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2006-026	MH	026-0001	MHSEQ	8	BASIS	Basis for Diagnosis	HISTOPATH	CRF
2006-026	MH	026-0001	MHSEQ	8	STAGE	Disease Stage	STAGE 3	CRF
2006-026	MH	026-0001	MHSEQ	8	THRSTAT	Therapy Status	RELAPSED	CRF

The Relationship

The Data

Metadata

There are several key principles to keep in mind when working with data in the SUPP-- datasets:

- Supplemental Qualifier records are merged onto one or more parent records defined in the Relationship variable shown above.
- Supplemental Qualifier records are related only to the parent record(s). There is no unambiguous machine-readable mechanism for Supplemental Qualifier records to be related to each other.
- Supplemental Qualifiers qualify the parent record in exactly the same way that the standard Qualifiers in SDTM Tables 2.1.1-2.2.3 do.
- Most validation tools will check to ensure that every record in a SUPP-- dataset relates back to a valid parent domain record. No orphans are allowed.
- The use of "Supplemental" in the name is a bit misleading, and has often resulted in sponsors finding creative ways to misappropriate standard variables for what are really their own NSVs in order to avoid the connotation that their NSVs might not be all that important.

SUPPLEMENTAL QUALIFIER CONSIDERATIONS

WHEN TO USE SUPPLEMENTAL QUALIFIERS

When the timing of the data represented in the SUPP-- datasets is different from that of the parent record, then creating Supplemental Qualifier records may not be appropriate. There will be times, however, when limited data precludes any alternative. An example is the following regarding an AE rechallenge:

suppae.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2008-0010	AE	0010-123	AESEQ	14	RECHALLFL	Rechallenge Conducted?	Y	CRF

It would not be expected that the rechallenge would have taken place during the span of time of the AE, but rather would have occurred after the AE has resolved. Thus, the timing is not shared, but there is little alternative if this is all the data one has. If, however, one had collected additional data, such as the rechallenge date and rechallenge result, the rechallenge data would likely be better represented in a Findings About domain, described in Section 6.4 of the SDTMIG.

WHEN NOT TO USE SUPPLEMENTAL QUALIFIERS

There has been the temptation for some sponsors to include "extra information in as Supplemental Qualifiers, "just in case the FDA needs it." Examples seen have included separate day, month, and year components, SAS dates, and legacy data values that were replaced with controlled terminology. At no time has anyone from the FDA ever asked for these "extra" data values. In fact, this practice has likely contributed to the term, "trash can" sometimes being applied in a general way.

DATA VALUES SUBMITTED IN QVAL

Submission values for SDTM-based data are expected to be human interpretable. Sponsors should not be submitting values such as "1" or "2" for any variable. CDISC has controlled terminology for all standard variables based upon this principle. It is expected that this principle will be followed in Supplemental Qualifiers. The SUPPCM data below was created by one sponsor. Although there might have been a check box marked "X" on the CRF, "X" is not an acceptable submission value (even if we assume "X" means "yes" and not "wrong". Controlled terminology for Yes/No questions consists of primarily of "Y" and "N".

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2005-ABC	CM	ABC-0101	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0102	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0102	CMSEQ	2	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0104	CMSEQ	1	CMCONT	Continuing	X	CRF
2005-ABC	CM	ABC-0104	CMSEQ	3	CMCONT	Continuing	X	CRF

CHOICE OF IDVAR

In the above example, the IDVAR is MHSEQ, meaning that the data in each QVAL is related to a single parent record. This is because the --SEQ variable in all subject-based domains is a sponsor-defined numeric identifier that must be unique within study, subject, and domain. The --SEQ variable is the one most commonly used as an IDVAR in this author's experience. It allows for the process of creating and re-merging the NSV data to be consistent across all domains. It is possible, however, to use other Identifying Variables. The SDTMIG (Section 8.4.3) shows an example of an IDVAR being QSCAT, meaning that the value in QVAL (the language in which the questionnaire was presented) would be related to all records whose QSCAT value was that present in IDVARVAL (which in the example was SF36).

Caution is needed when choosing the IDVAR, in order to avoid the potential for unintended consequences. Consider the example below, in which the urine-temperature information will appear on every parent LB record whose VISITNUM value is 3.

supplb.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2008-0001	LB	2008-0001	VISITNUM	3	URINTEMP	Urine Temperature within Expected Range	Y	CRF

A few pros and cons with the selection of VISITNUM as the IDVAR:

Pro	Con
If the study collected only urine samples, the selection of VISITNUM as the IDVAR could be appropriate.	If the study also collected blood samples, the urine-temperature information would also appear on these records, and this would obviously not have been appropriate.
It may be desirable to know about the urine temperature when looking at each individual urine result.	The urine temperature will appear on all tests and might not be relevant to all of them.

IS THE USE OF SUPPLEMENTAL QUALIFIERS THE BEST SOLUTION?

There are cases when the use of Supplemental Qualifiers, while tempting, may not be the best solution. This section will present two case studies from the examination of actual data. The first consists of data collected on a physical examination CRF page, while the second involves the data collected during ambulatory blood-pressure monitoring.

CASE STUDY 1: PHYSICAL EXAMINATION CRF PAGE

Abdomen	_____
Extremities/Joints	_____
General Appearance	_____
Heart	_____
HEENT	_____
Lungs	_____
Lymph Nodes	_____
Mental Status	_____
Neurologic	_____
Reflexes	_____
Skin	_____
Disease Relapse Since Last Visit?	<input type="checkbox"/> Yes <input type="checkbox"/> No

Most sponsors follow the example in the SDTMIG and model the body-system data as shown below. Because the last question on the CRF is not a body system, some sponsors have been tempted to model it as a Supplemental Qualifier. In one case, a sponsor used an IDVAR of VISIT (Example 1) since they felt it applied to the Visit 1 physical examination as a whole. In another case, a sponsor used PESEQ (Example 2) as the IDVAR.

pe.xpt *

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC
2001-01	PE	2001-01-1008	1	ABDOMEN	Abdomen	NORMAL	NORMAL
2001-01	PE	2001-01-1008	2	EXTRJOIN	Extremities/ Joints	JOINTS SWOLLEN IN FINGERS	JOINTS SWOLLEN IN FINGERS
2001-01	PE	2001-01-1008	3	GENAPP	General Appearance	NORMAL	NORMAL
2001-01	PE	2001-01-1008	4	HEART	Heart	NORMAL	NORMAL
2001-01	PE	2001-01-1008	5	HEENT	HEENT	NORMAL	NORMAL
2001-01	PE	2001-01-1008	6	LUNGS	Lungs	NORMAL	NORMAL
2001-01	PE	2001-01-1008	7	LYMPNODE	Lymph Nodes	NORMAL	NORMAL
2001-01	PE	2001-01-1008	8	MENTSTAT	Mental Status	NORMAL	NORMAL
2001-01	PE	2001-01-1008	9	NEURO	Neurologic	NORMAL	NORMAL
2001-01	PE	2001-01-1008	10	REFLEXES	Reflexes	NORMAL	NORMAL
2001-01	PE	2001-01-1008	11	SKIN	Skin	NORMAL	NORMAL

* Some variables are not shown due to space limitations.

Example 1

supppe.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2001-01	PE	2001-01-1008	VISIT	1	PERELFL	Disease Relapse Since Last Visit	N	CRF

Example 2

supppe.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
2001-01	PE	2001-01-1008	PESEQ	1	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	2	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	3	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	4	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	5	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	6	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	7	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	8	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	9	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	10	PERELFL	Disease Relapse Since Last Visit	N	CRF
2001-01	PE	2001-01-1008	PESEQ	11	PERELFL	Disease Relapse Since Last Visit	N	CRF

The (likely) unintended consequence of both selections of IDVAR is that a question that was asked only once will appear on eleven records for every subject (for every visit) when the Supplemental Qualifier records for QNAM = PERELFL are merged onto the parent domain to create the view shown below.

Physical Exam Domain View with SUPPPE Data Merged onto PE Records*

STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PEORRES	PESTRESC	PERELFL
2001-01	PE	2001-01-1008	1	ABDOMEN	Abdomen	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	2	EXTRJOIN	Extremities/ Joints	JOINTS SWOLLEN IN FINGERS	JOINTS SWOLLEN IN FINGERS	N
2001-01	PE	2001-01-1008	3	GENAPP	General Appearance	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	4	HEART	Heart	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	5	HEENT	HEENT	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	6	LUNGS	Lungs	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	7	LYMPNODE	Lymph Nodes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	8	MENTSTAT	Mental Status	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	9	NEURO	Neurologic	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	10	REFLEXES	Reflexes	NORMAL	NORMAL	N
2001-01	PE	2001-01-1008	11	SKIN	Skin	NORMAL	NORMAL	N

* Some variables are not shown due to space limitations.

The more accurate solution might have been to retain the last question as the twelfth question in the pe.xpt dataset, eliminating the need for the SUPPPE dataset altogether.

CASE STUDY 2: AMBULATORY BLOOD-PRESSURE MONITORING (ABPM) DATA

Note to Readers: Due to page-width limitations, most of the datasets presented in this section consist of tables that wrap. Row numbers are included where needed, to facilitate viewing individual records. Most of the datasets presented as table structures in this section have corresponding screenshots of Excel data in Appendix A in an attempt to show as many of the columns as contiguously as possible. The presentation of the datasets in wrapped tables in this section allows for a larger-text view of the data. The example presented in this case study is a relatively complex one, and probably not for the faint of heart. It is presented here, however, because it is data such as this that often presents the biggest challenge to sponsors. For those not wanting to navigate through all the data examples, the key thoughts are summarized at the end of this section.

ABPM data often contains information about the monitoring period, often 24 hours, and the measurements of blood pressure and heart rate at set intervals during the monitoring period. These measurements are usually recorded at 10- to 15-minute intervals. Table 1 (replicated in Table A1 of Appendix A) shows four interval records from a legacy dataset from a 24-hour period for one subject. The data in the first two sections of the dataset are related to the interval measurements. The data in the second two sections of the table are related to the 24-hour measurement period.

Table 1. Original ABPM Dataset
abpm.xpt

Row	PROTNO	FILEREF	RECSEQNO	PATNO	VISITNO	BPSYS	BPDIA
1	Protocol Number	File Reference	Record Sequence Number	Patient Number	Visit Number	Systolic BP (mmHg)	Diastolic BP (mmHg)
2	2010-198	567-123	1101	2010-198-001	1	161	104
3	2010-198	567-123	1102	2010-198-001	1	175	103
4	2010-198	567-123	1103	2010-198-001	1	163	99
5	2010-198	567-123	1104	2010-198-001	1	162	96

Row	BPDIFF	HRATE	BPDT	BPTM	ERROR
1	Diff Systolic Diastolic BP (mmHg)	Heart Rate (bpm)	Date BP Reading	Time of BP Reading	Error Decode
2	57	74	2010-10-24	9:30	No Error
3	72	86	2010-10-24	9:45	No Error
4	64	78	2010-10-24	10:00	No Error
5	66	79	2010-10-24	10:15	Intermittent Connection

Row	BPSTDT	BPSTTM	BPENDT	BPENTM	BPMEAN	BPDUR
1	BP Start Date	BP Start Time	BP End Date	BP End Time	Mean BP Systolic (mmHg)	Total Duration
2	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00
3	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00
4	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00
5	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00

Row	QCALL	QCMINDUR	QCMAXDUR	NUMMEAS	TESTCOMP
1	Overall QC Code for 24H	QC Minimum Duration for 24H	QC Consecutive Hours for 24H	Num Measurements 24H Int	Test Period Successful Completion
2	PASS	PASS	PASS	94	YES
3	PASS	PASS	PASS	94	YES
4	PASS	PASS	PASS	94	YES
5	PASS	PASS	PASS	94	YES

Most sponsors submitting ABPM data usually map the interval data without much problem to the SDTMIG Vital Signs (VS) domain, as diagrammed in Figure 1. Table 2 below and Table A2 in Appendix A show representations of the corresponding SDTM-based Vital Signs dataset (*vs.xpt*). The FILEREF is the sponsor identifier for the period, and has been mapped to VSREFID (Reference Identifier). The RECSEQNO is the sponsor number for the interval within the period, and has been mapped to VSSPID (Sponsor ID).

Figure 1. Legacy Data that Can Be Mapped to SDTM-Based Vital Signs (VS) Variables

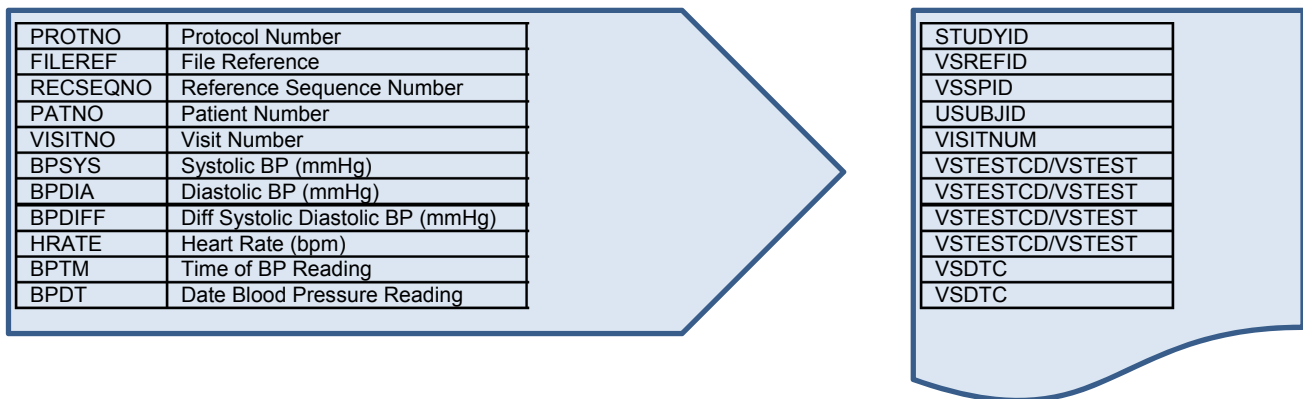


Table 2. Converted SDTM-Based Vital Signs Dataset vs.xpt

Row	STUDYID	DOMAIN	USUBJID	VSSEQ	VSREFID	VSSPID	VSTESTCD	VSTEST	VSORRES	VSORRESU
1	2010-198	VS	2010-198-001	1	567-123	101	BPSYS	Systolic Blood Pressure	161	mmHG
2	2010-198	VS	2010-198-001	5	567-123	101	BPDIA	Diastolic Blood Pressure	104	mmHG
3	2010-198	VS	2010-198-001	9	567-123	101	BPDIFF	BP Difference	57	mmHG
4	2010-198	VS	2010-198-001	13	567-123	101	HR	Heart Rate	74	BEATS/MIN
5	2010-198	VS	2010-198-001	2	567-123	102	BPSYS	Systolic Blood Pressure	175	mmHG
6	2010-198	VS	2010-198-001	6	567-123	102	BPDIA	Diastolic Blood Pressure	103	mmHG
7	2010-198	VS	2010-198-001	10	567-123	102	BPDIFF	BP Difference	72	mmHG
8	2010-198	VS	2010-198-001	14	567-123	102	HR	Heart Rate	86	BEATS/MIN
9	2010-198	VS	2010-198-001	3	567-123	103	BPSYS	Systolic Blood Pressure	163	mmHG
10	2010-198	VS	2010-198-001	7	567-123	103	BPDIA	Diastolic Blood Pressure	99	mmHG
11	2010-198	VS	2010-198-001	11	567-123	103	BPDIFF	BP Difference	64	mmHG
12	2010-198	VS	2010-198-001	15	567-123	103	HR	Heart Rate	78	BEATS/MIN
13	2010-198	VS	2010-198-001	4	567-123	104	BPSYS	Systolic Blood Pressure	162	mmHG
14	2010-198	VS	2010-198-001	8	567-123	104	BPDIA	Diastolic Blood Pressure	96	mmHG
15	2010-198	VS	2010-198-001	12	567-123	104	BPDIFF	BP Difference	66	mmHG
16	2010-198	VS	2010-198-001	16	567-123	104	HR	Heart Rate	79	BEATS/MIN

Row	VSSTRESC	VSSTRESN	VSSTRESU	VISITNUM	VSDTC
1	161	161	mmHG	1	2010-10-24T:09:30
2	104	104	mmHG	1	2010-10-24T:09:30
3	57	57	mmHG	1	2010-10-24T:09:30
4	74	74	BEATS/MIN	1	2010-10-24T:09:30
5	175	175	mmHG	1	2010-10-24T:09:45
6	103	103	mmHG	1	2010-10-24T:09:45
7	72	72	mmHG	1	2010-10-24T:09:45
8	86	86	BEATS/MIN	1	2010-10-24T:09:45
9	163	163	mmHG	1	2010-10-24T:10:00
10	99	99	mmHG	1	2010-10-24T:10:00
11	64	64	mmHG	1	2010-10-24T:10:00
12	78	78	BEATS/MIN	1	2010-10-24T:10:00
13	162	162	mmHG	1	2010-10-24T:10:15
14	96	96	mmHG	1	2010-10-24T:10:15
15	66	66	mmHG	1	2010-10-24T:10:15
16	79	79	BEATS/MIN	1	2010-10-24T:10:15

The data related to the monitoring period itself sometimes seem to be generically viewed as non-standard and, as a result, have been mapped by some into Supplemental Qualifiers, as diagrammed in Figure 2. The SUPPVS datasets that would result from this process are represented in Table 3 and Table A3 of Appendix A. The disadvantages of this approach are as follows:

- The timing for the interval is different from the timing of the individual measurements.
- Data that was collected once per period ends up appearing on every interval record in merged views of the data (shown only in Table A4 of Appendix A); however, the period data don't really qualify the interval records.

Figure 2. Non-Standard Legacy Data Mapped to Supplemental Qualifiers

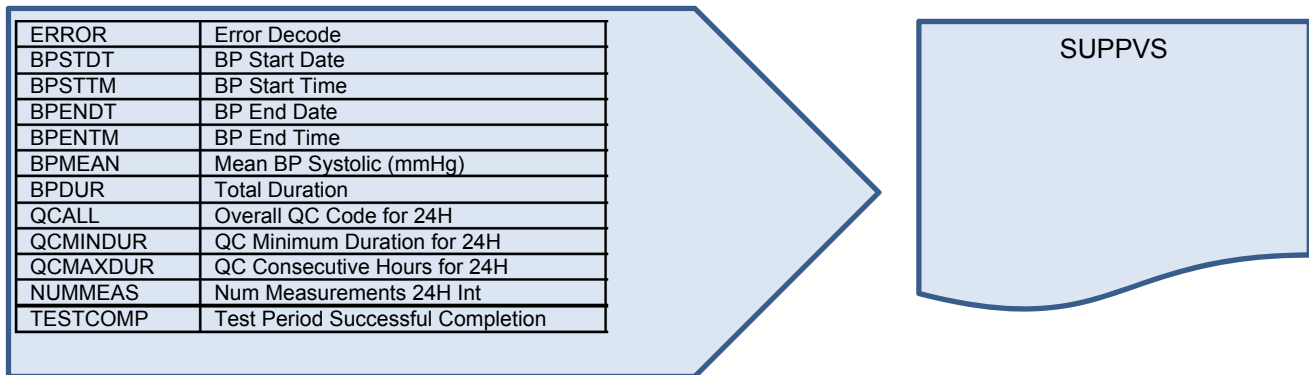


Table 3. Converted Supplemental Qualifiers Dataset

suppvs.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
2010-198	VS	2010-198-001	VSSPID	101	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	102	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	103	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	104	ERROR	BP Error	Intermittent Connection	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPSTDT	BP Start Date/Time	2010-10-24T09:30	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPENDT	BP End Date/Time	2010-10-25T09:45	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPMEAN	Mean BP Systolic (mmHg)	156	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPDUR	Total Duration	PT24H15M	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCALL	Overall QC Code for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMINDUR	QC Minimum Duration for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMAXDUR	QC Consecutive Hours for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMINPCT	QC Minimum Percentage for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCINCL	Reading Incl. Period	Y	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	NUMMEAS	Num Measmnts 24H Int	94	eDT	

Because of these potential issues around the SUPPVS representation, a better option might be to map the period data to a separate Findings observation-class domain, as diagrammed in Figure 3.

Figure 3. : Non-Standard Legacy Data Mapped to Another Findings Dataset

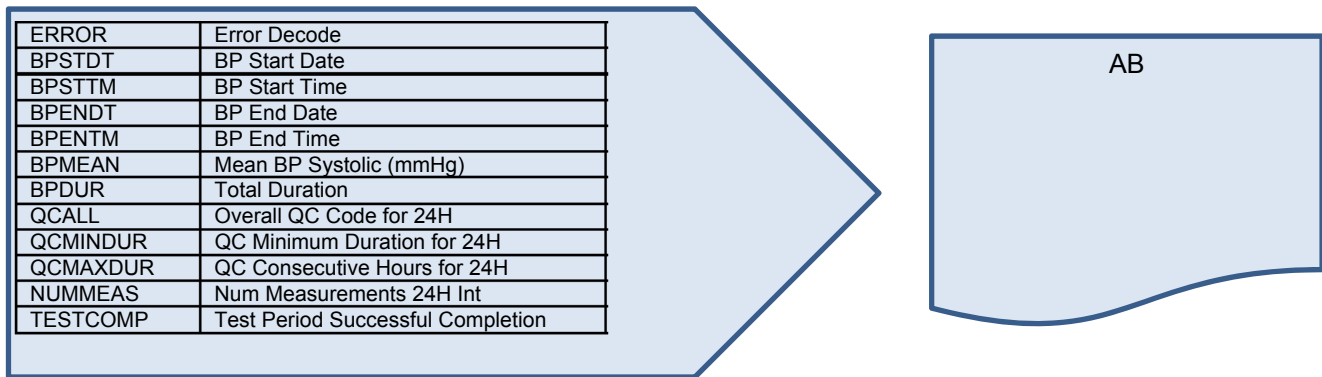


Table 4 shows what this dataset would look like. The ab.xpt dataset is also shown as part of Table A5 in Appendix A. This dataset represents many of the SUPPVS QNAM/QVAL values as ABTESTCD/ABTEST values. Each one is associated with the start and end of the interval, represented by ABDTC and ABENDTC. Each maintains the value of what was in VSREFID in ABREFID so that the measurements of an interval can be associated with the period from which the interval came. The fact that VSREFID and ABREFID are used to establish this relationship for all measurements and intervals is reflected in the RELREC dataset, shown in Table 5 and in Table A5 in Appendix A.

Table 4. Converted Findings Dataset for Period-Level Data

ab.xpt

Row	STUDYID	DOMAIN	USUBJID	ABSEQ	ABREFID	ABTESTCD	ABTEST	ABORRES	ABORRESU
1	2010-198	AB	2010-198-001	1	567-123	TESTCOMP	Test Period Successful Completion	Y	
2	2010-198	AB	2010-198-001	2	567-123	BPMEAN	Mean BP Systolic (mmHg)	156	BEATS/MIN
3	2010-198	AB	2010-198-001	3	567-123	BPDUR	Total Duration		
4	2010-198	AB	2010-198-001	4	567-123	QCALL	Overall QC Code for 24H	PASS	
5	2010-198	AB	2010-198-001	5	567-123	QCMINDUR	QC Minimum Duration for 24H	PASS	
6	2010-198	AB	2010-198-001	6	567-123	QCMAXDUR	QC Consecutive Hours for 24H	PASS	
7	2010-198	AB	2010-198-001	7	567-123	QCMINPCT	QC Minimum Percentage for 24H	PASS	
8	2010-198	AB	2010-198-001	8	567-123	QCINCL	Reading Incl. Period	Y	
9	2010-198	AB	2010-198-001	9	567-123	NUMMEAS	Num Measurements 24H Int	94	

Row	ABSTRESC	ABSTRESN	ABSTRESU	VISITNUM	ABDTC	ABENDTC
1	Y			1	2010-10-24T:09:30	2010-10-25T:09:45
2	156	156	BEATS/MIN	1	2010-10-24T:09:30	2010-10-25T:09:45
3	PT24H15M			1	2010-10-24T:09:30	2010-10-25T:09:45
4	PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
5	PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
6	PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
7	PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
8	Y			1	2010-10-24T:09:30	2010-10-25T:09:45
9	94	94		1	2010-10-24T:09:30	2010-10-25T:09:45

Table 5. Using RELREC to Relate the Individual Measurements to the Measurement Period

relrec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
2010-198	VS		VSREFID		MANY	VSAB1
2010-198	VS		ABREFID		ONE	VSAB1

This case study shows that one cannot simply look at a single legacy dataset and conclude that everything that doesn't map into a single SDTM-based domain must be a Supplemental Qualifier. In this case, the mapping of NSVs into Supplemental Qualifiers results in data with different time frames and different data structures being shown in merged views of data. One needs to be familiar with the legacy data in order to represent in SDTM-based datasets as accurately as possible. In this case, the more appropriate mapping (shown in Table A5 of Appendix A) includes the following:

- One record per measurement per interval standard data in a Vital Signs dataset
- One record per measurement per interval NSVs in a SUPPVS dataset
- One record per measurement period (24 hours) in an Ambulatory BP dataset
- The use of RELREC to represent the relationship of the interval data to the corresponding period data.

CONCLUSIONS

This paper has presented some considerations for when Supplemental Qualifiers should and should not be used. Instances were presented where creating additional test and test code records in a Findings domain would be a better choice than Supplemental Qualifiers. In mapping legacy data to SDTM-based domains, one cannot simply conclude that everything that doesn't map into a single SDTM-based domain must be a Supplemental Qualifier.

Once one has chosen to use Supplemental Qualifiers to represent NSVs, caution is needed when choosing the IDVAR, in order to avoid the potential for unintended consequences when merged views of the data are created. Sponsors should ask

themselves this question: if the use Supplemental Qualifiers had not been necessary to represent NSVs, and the data could have been submitted in one or more relatively normalized dataset, what would it have looked like?

CONTACT INFORMATION

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

Fred Wood
Accenture Life Sciences
585 East Swedesford Road
Wayne, PA 19087
484-881-2297
f.wood@accenture.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.

Considerations in Data Modeling when Creating Supplemental Qualifiers Datasets in SDTM-Based Submissions (continued)

Appendix A
Dataset Examples for Case Study 2

Table A1. Original Dataset
abpm.xpt

PROTNO	FILEREF	RECSEQNO	PATNO	VISITNO	BPSYS	BPDIA	BPDIFF	HRATE	BPDT	BPTM	BPERR	BPSTDT	BPSTTM	BPENDT	BPENTM	BPMEAN	BPDUR	QCALL	QCMINDUR	QCMAXDUR	NUMMEAS	TESTCOMP
Protocol Number	File Reference	Record Sequence Number	Patient Number	Visit Number	Systolic BP (mmHg)	Diastolic BP (mmHg)	Diff Systolic Diastolic BP (mmHg)	Heart Rate (bpm)	Date BP Reading	Time of BP Reading	Error Decode	BP Start Date	BP Start Time	BP End Date	BP End Time	Mean BP Systolic (mmHg)	Total Duration	Overall QC Code for 24H	QC Minimum Duration for 24H	QC Consecutive Hours for 24H	Num Measurements 24H Int	Test Period Successful Completion
2010-198	567-123	1101	2010-198-001	1	161	104	57	74	2010-10-24	9:30	No Error	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00	PASS	PASS	PASS	94	YES
2010-198	567-123	1102	2010-198-001	1	175	103	72	86	2010-10-24	9:45	No Error	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00	PASS	PASS	PASS	94	YES
2010-198	567-123	1103	2010-198-001	1	163	99	64	78	2010-10-24	10:00	No Error	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00	PASS	PASS	PASS	94	YES
2010-198	567-123	1104	2010-198-001	1	162	96	66	79	2010-10-24	10:15	Intermittent Connection	2010-10-24	9:30	2010-10-25	9:45	156	24:15:00	PASS	PASS	PASS	94	YES

Table A2. Converted VS Dataset
vs.xpt

STUDYID	DOMAIN	USUBJID	VSSEQ	VSREFID	VSSPID	VSTESTCD	VSTEST	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VISITNUM	VSDTC
2010-198	VS	2010-198-001	1	567-123	101	BPSYS	Systolic Blood Pressure	161	mmHG	161	161	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	5	567-123	101	BPDIA	Diastolic Blood Pressure	104	mmHG	104	104	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	9	567-123	101	BPDIFF	BP Difference	57	mmHG	57	57	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	13	567-123	101	HR	Heart Rate	74	BEATS/MIN	74	74	BEATS/MIN	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	2	567-123	102	BPSYS	Systolic Blood Pressure	175	mmHG	175	175	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	6	567-123	102	BPDIA	Diastolic Blood Pressure	103	mmHG	103	103	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	10	567-123	102	BPDIFF	BP Difference	72	mmHG	72	72	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	14	567-123	102	HR	Heart Rate	86	BEATS/MIN	86	86	BEATS/MIN	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	3	567-123	103	BPSYS	Systolic Blood Pressure	163	mmHG	163	163	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	7	567-123	103	BPDIA	Diastolic Blood Pressure	99	mmHG	99	99	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	11	567-123	103	BPDIFF	BP Difference	64	mmHG	64	64	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	15	567-123	103	HR	Heart Rate	78	BEATS/MIN	78	78	BEATS/MIN	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	4	567-123	104	BPSYS	Systolic Blood Pressure	162	mmHG	162	162	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	8	567-123	104	BPDIA	Diastolic Blood Pressure	96	mmHG	96	96	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	12	567-123	104	BPDIFF	BP Difference	66	mmHG	66	66	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	16	567-123	104	HR	Heart Rate	79	BEATS/MIN	79	79	BEATS/MIN	1	2010-10-24T:10:15

Table A3. Converted SUPPVs Dataset
suppv.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
2010-198	VS	2010-198-001	VSSPID	101	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	102	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	103	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	104	ERROR	BP Error	Intermittent Connection	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPSTDT	BP Start Date/Time	2010-10-24T09:30	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPENDT	BP End Date/Time	2010-10-25T09:45	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPMEAN	Mean BP Systolic (mmHg)	156	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	BPDUR	Total Duration	PT24H15M	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCALL	Overall QC Code for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMINDUR	QC Minimum Duration for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMAXDUR	QC Consecutive Hours for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCMINPCT	QC Minimum Percentage for 24H	PASS	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	QCINCL	Reading Incl. Period	Y	eDT	
2010-198	VS	2010-198-001	VSREFID	567-123	NUMMEAS	Num Measurements 24H Int	94	eDT	

Appendix A (cont'd)
Dataset Examples for Case Study 2

Table A4. Merged View of VS and SUPPVS Datasets (continuous table)

Row	STUDYID	DOMAIN	USUBJID	VSSEQ	VSREFID	VSSPID	VSTESTCD	VSTEST	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VISITNUM	VSDTC
1	2010-198	VS	2010-198-001	1	567-123	101	BPSYS	Systolic Blood Pressure	161	mmHG	161	161	mmHG	1	2010-10-24T:09:30
2	2010-198	VS	2010-198-001	5	567-123	101	BPDIA	Diastolic Blood Pressure	104	mmHG	104	104	mmHG	1	2010-10-24T:09:30
3	2010-198	VS	2010-198-001	9	567-123	101	BPDIFF	BP Difference	57	mmHG	57	57	mmHG	1	2010-10-24T:09:30
4	2010-198	VS	2010-198-001	13	567-123	101	HR	Heart Rate	74	BEATS/MIN	74	74	BEATS/MIN	1	2010-10-24T:09:30
5	2010-198	VS	2010-198-001	2	567-123	102	BPSYS	Systolic Blood Pressure	175	mmHG	175	175	mmHG	1	2010-10-24T:09:45
6	2010-198	VS	2010-198-001	6	567-123	102	BPDIA	Diastolic Blood Pressure	103	mmHG	103	103	mmHG	1	2010-10-24T:09:45
7	2010-198	VS	2010-198-001	10	567-123	102	BPDIFF	BP Difference	72	mmHG	72	72	mmHG	1	2010-10-24T:09:45
8	2010-198	VS	2010-198-001	14	567-123	102	HR	Heart Rate	86	BEATS/MIN	86	86	BEATS/MIN	1	2010-10-24T:09:45
9	2010-198	VS	2010-198-001	3	567-123	103	BPSYS	Systolic Blood Pressure	163	mmHG	163	163	mmHG	1	2010-10-24T:10:00
10	2010-198	VS	2010-198-001	7	567-123	103	BPDIA	Diastolic Blood Pressure	99	mmHG	99	99	mmHG	1	2010-10-24T:10:00
11	2010-198	VS	2010-198-001	11	567-123	103	BPDIFF	BP Difference	64	mmHG	64	64	mmHG	1	2010-10-24T:10:00
12	2010-198	VS	2010-198-001	15	567-123	103	HR	Heart Rate	78	BEATS/MIN	78	78	BEATS/MIN	1	2010-10-24T:10:00
13	2010-198	VS	2010-198-001	4	567-123	104	BPSYS	Systolic Blood Pressure	162	mmHG	162	162	mmHG	1	2010-10-24T:10:15
14	2010-198	VS	2010-198-001	8	567-123	104	BPDIA	Diastolic Blood Pressure	96	mmHG	96	96	mmHG	1	2010-10-24T:10:15
15	2010-198	VS	2010-198-001	12	567-123	104	BPDIFF	BP Difference	66	mmHG	66	66	mmHG	1	2010-10-24T:10:15
16	2010-198	VS	2010-198-001	16	567-123	104	HR	Heart Rate	79	BEATS/MIN	79	79	BEATS/MIN	1	2010-10-24T:10:15

Row	ERROR	BPSTDT	BPENDT	BPMEAN	BPDUR	QCALL	QCMINDUR	QCMAXDUR	QCMINPCT	QCINCL	NUMMEAS
1	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
2	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
3	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
4	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
5	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
6	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
7	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
8	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
9	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
10	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
11	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
12	No Error	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
13	Intermittent Connection	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
14	Intermittent Connection	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
15	Intermittent Connection	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94
16	Intermittent Connection	2010-10-24T09:30	2010-10-25T09:45	156	PT24H15M	PASS	PASS	PASS	PASS	Y	94

Appendix A (cont'd)
Dataset Examples for Case Study 2

Table A5. Final Solution: VS, SUPPVs, and AB Datasets, with RELREC to Show Relationship
vs.xpt

STUDYID	DOMAIN	USUBJID	VSSEQ	VSREFID	VSSPID	VSTESTCD	VSTEST	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VISITNUM	VSDTC
2010-198	VS	2010-198-001	1	567-123	101	BPSYS	Systolic Blood Pressure	161	mmHG	161	161	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	5	567-123	101	BPDIA	Diastolic Blood Pressure	104	mmHG	104	104	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	9	567-123	101	BPDIFF	BP Difference	57	mmHG	57	57	mmHG	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	13	567-123	101	HR	Heart Rate	74	BEATS/MIN	74	74	BEATS/MIN	1	2010-10-24T:09:30
2010-198	VS	2010-198-001	2	567-123	102	BPSYS	Systolic Blood Pressure	175	mmHG	175	175	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	6	567-123	102	BPDIA	Diastolic Blood Pressure	103	mmHG	103	103	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	10	567-123	102	BPDIFF	BP Difference	72	mmHG	72	72	mmHG	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	14	567-123	102	HR	Heart Rate	86	BEATS/MIN	86	86	BEATS/MIN	1	2010-10-24T:09:45
2010-198	VS	2010-198-001	3	567-123	103	BPSYS	Systolic Blood Pressure	163	mmHG	163	163	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	7	567-123	103	BPDIA	Diastolic Blood Pressure	99	mmHG	99	99	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	11	567-123	103	BPDIFF	BP Difference	64	mmHG	64	64	mmHG	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	15	567-123	103	HR	Heart Rate	78	BEATS/MIN	78	78	BEATS/MIN	1	2010-10-24T:10:00
2010-198	VS	2010-198-001	4	567-123	104	BPSYS	Systolic Blood Pressure	162	mmHG	162	162	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	8	567-123	104	BPDIA	Diastolic Blood Pressure	96	mmHG	96	96	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	12	567-123	104	BPDIFF	BP Difference	66	mmHG	66	66	mmHG	1	2010-10-24T:10:15
2010-198	VS	2010-198-001	16	567-123	104	HR	Heart Rate	79	BEATS/MIN	79	79	BEATS/MIN	1	2010-10-24T:10:15

suppv.s.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
2010-198	VS	2010-198-001	VSSPID	101	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	102	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	103	ERROR	BP Error	No Error	eDT	
2010-198	VS	2010-198-001	VSSPID	104	ERROR	BP Error	Intermittent Connection	eDT	

ab.xpt

Row	STUDYID	DOMAIN	USUBJID	ABSEQ	ABREFID	ABTESTCD	ABTEST	ABORRES	ABORRESU	ABSTRESC	ABSTRESN	ABSTRESU	VISITNUM	ABDTC	ABENDTC
1	2010-198	AB	2010-198-001	1	567-123	TESTCOMP	Test Period Successful Completion	Y		Y			1	2010-10-24T:09:30	2010-10-25T:09:45
2	2010-198	AB	2010-198-001	2	567-123	BPMEAN	Mean BP Systolic (mmHg)	156	BEATS/MIN	156	156	BEATS/MIN	1	2010-10-24T:09:30	2010-10-25T:09:45
3	2010-198	AB	2010-198-001	3	567-123	BPDUR	Total Duration			PT24H15M			1	2010-10-24T:09:30	2010-10-25T:09:45
4	2010-198	AB	2010-198-001	4	567-123	QCALL	Overall QC Code for 24H	PASS		PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
5	2010-198	AB	2010-198-001	5	567-123	QCMINDUR	QC Minimum Duration for 24H	PASS		PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
6	2010-198	AB	2010-198-001	6	567-123	QCMAXDUR	QC Consecutive Hours for 24H	PASS		PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
7	2010-198	AB	2010-198-001	7	567-123	QCMINPCT	QC Minimum Percentage for 24H	PASS		PASS			1	2010-10-24T:09:30	2010-10-25T:09:45
8	2010-198	AB	2010-198-001	8	567-123	QCINCL	Reading Incl. Period	Y		Y			1	2010-10-24T:09:30	2010-10-25T:09:45
9	2010-198	AB	2010-198-001	9	567-123	NUMMEAS	Num Measurements 24H Int	94		94	94		1	2010-10-24T:09:30	2010-10-25T:09:45

relrec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
2010-198	VS		VSREFID		MANY	VSAB1
2010-198	VS		ABREFID		ONE	VSAB1