

SDTM-ETL 5.0 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2025-02-04

Handling Unscheduled Visits

Table of Contents

Table of Contents	1
Introduction.....	1
Post-processing for VISITNUM for unscheduled visits.....	2
Aligning with SV-VISITNUM for unscheduled visits.....	8
Unscheduled visits before the first planned visit	8
Trying to cope with data quality issues.....	10
Alternative approaches for generating SV.	11
Conclusions.....	12

Introduction

The way the SDTMIGs recommend to handle unscheduled visits is at least ... bizarre. First of all, for most domains, they require a variable VISITNUM (visit number) as a derivative of VISIT (visit name) with the notice "Numeric version of VISIT, used for sorting". Very probably meant is "for chronological sorting". As the chronology is however already provided by the –DTC (date/time of collection), it looks that the only reason for VISITNUM is that reviewers are incapable to sort on date/time-s, maybe that the only thing they can sort on is numeric values¹. Unfortunately, it looks as CDISC has once again given in to demands of the regulatory agencies to compensate for the primitiveness of their review tools.

Also very strange is that VISITNUM must come, though it is a derivative of VISIT, before VISIT, which is ... bad design ...

The way VISITNUM is defined, as a number, makes it complicated when there are unscheduled visits, i.e. visits that are executed between planned ("scheduled") visits. Essentially, such visits should not get a VISITNUM when the data is already coming (on basis of –DTC) in chronological order. Well, ideally, VISITNUM should not be present anyway when the data is already coming in chronological order. Essentially, the SDTM requirement for VISITNUM reminds me of the punch cards of the 80ies, where [we used column 73-80 for a sequence number](#) for the worst case that we dropped our stack of cards.

One popular way of treating VISITNUM for an unscheduled visit is to assign it a high number like "99" or "999". Essentially, this should be more than sufficient when the data is coming in chronological order, as there is no reason at all for the reviewer to have to "sort" the data.

I consider this as the best practice.

Remark also that VISIT and VISITNUM for unscheduled visits should not appear in the trial design

¹ As SDTM and SEND are using ISO-8601 notation, date/time-s can also simply be sorted alphabetically, giving the same result as when first converting to "calendar" date/time-s and then sorting.

TV (Trial Visits) dataset, as this domain should only contain "planned" visits.

Another approach, also recommended by an, in the mean time retired, [Phuse paper](#), is to use fractional numbers (like "3.1", "3.2" ...) for VISITNUM for unscheduled visits. This also appears in the examples of the SDTMIGs, and is explained in the section "Clinical Encounters and Visits" (here from the SDTMIG-3.4):

Sponsor practices for populating visit variables for unplanned visits may vary.

- VISITNUM should generally be populated, even for unplanned visits, as it is expected in many Findings domains, as described above. The easiest method of populating VISITNUM for unplanned visits is to assign the same value (e.g., 99) to all unplanned visits, although this method provides no differentiation between the unplanned visits and does not provide chronological sorting. Methods that provide a one-to-one relationship between visits and values of VISITNUM, that are consistent across domains, and that assign VISITNUM values that sort chronologically require more work and must be applied after all of a subject's unplanned visits are known.
- VISIT may be left null or may be populated with a generic value (e.g., "Unscheduled") for all unplanned visits, or individual values may be assigned to different unplanned visits.
- VISITDY must not be populated for unplanned visits; VISITDY is, by definition, the planned study day of visit. The actual study day of an unplanned visit belongs in a --DY variable.

The statement "... that sort chronologically require more work and must be applied after all of subject's unplanned visits are known" is a bit confusing, it essentially means: "You will need to do a post-processing step".

Essentially, this requires the mappers to sort the data chronologically, then, for unscheduled visits, (re)assign the visit number. This also means (like for assigning the --LOBXFL flag) that a post-processing must be applied, which can easily become yet-another source for errors.

Our own recommendation is to avoid post-processing steps as much as possible, but some of our customers want to have it anyway. So we implemented it.

Post-processing for VISITNUM for unscheduled visits

As of version 4.4 of the software, when executing the mappings, one will notice an additional checkbox in the "Execute Transformation (XSLT) code ..." dialog:

Execute Transformation (XSLT) Code

ODM file with clinical data:
C:\ Browse...

MetaData in separate ODM file
D:\SDTM-ETL\TestFiles\ODM1-3-1\CES_Metadata_with_unscheduled_visit.xml Browse...

Administrative data in separate ODM file
D:\SDTM-ETL\TestFiles\ODM1-3-1\CES_Metadata_with_unscheduled_visit.xml Browse...

Perform post-processing for assigning --LOBXFL Perform post-processing unscheduled VISITNUM

Split records > 200 characters to SUPP-- records

Move non-standard SDTM Variables to SUPP-- Move Comment Variables to Comments (CO) Domain

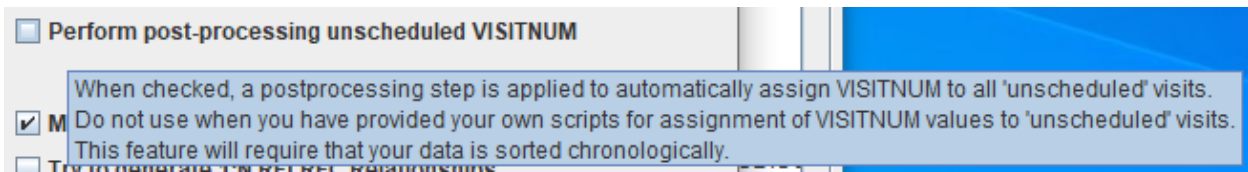
Move Relrec Variables to Related Records (RELREC) domain Try to generate 1:N RELREC Relationships

View Result SDTM tables Adapt Variable Length for longest result value

Generate 'NOT DONE' records for QS datasets Re-sort records using define.xml keys

Unique --SEQ values across 'split' domains Perform CDISC CORE validation on generated SDTM files

When one hovers the mouse over it, one obtains more information:



Also stating that this feature should not be used when one want to use it's own method for assigning VISITNUM for "unscheduled" visits, such as assigning "999" to them. It should only be used to have an automated assignment in a post-processing step.

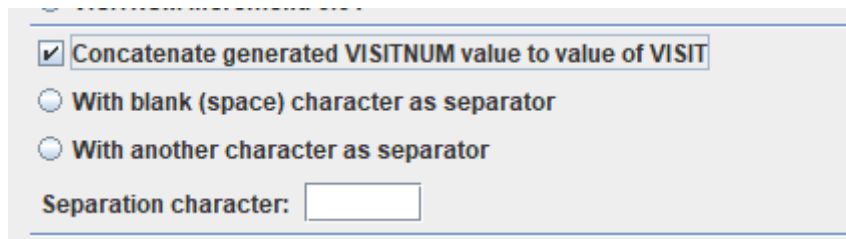
When one checks the checkbox, a dialog is presented:



It allows to choose whether the increment for assigning a decimal number for the unscheduled VISITNUM values should be "0.1" or "0.01". For the case of "0.1", this means that e.g. when an unscheduled visit chronologically comes after a planned visit with VISITNUM=3, it will be

assigned VISITNUM=3.1, any others before the next visit will then be assigned "3.2", "3.3" etc..

In case one wants to concatenate the assigned VISITNUM to the visit name in the variable "VISIT", check the checkbox "Concatenate generated VISITNUM value to value of VISIT":



The screenshot shows a configuration window with the following options:

- Concatenate generated VISITNUM value to value of VISIT
- With blank (space) character as separator
- With another character as separator

Below these options is a text input field labeled "Separation character:".

When none of the radiobuttons is selected, this will result in e.g. "UNSCHEDULED3.1" as the value for "VISIT". When selecting "with blank (space) character as separator", it will e.g. be "UNSCHEDULED 3.1". One can however also choose for another separator character by selecting the second radiobutton and then add the character of choice in the field for "Separation character:".

The choices for "For 'overlapping' visits" and "I have one or more unscheduled visits BEFORE the first scheduled visit" will be explained later in this document.

The **requirements** for correct use of this post-processing step for automated assignment for VISITNUM for unscheduled visits are important:

- The data needs to be sorted chronologically (see further)
- The value for "VISIT" (visit name) in the SDTM must be set to "UNSCHEDULED" (case-sensitive) in the mapping script, or at least be part of the visit name, otherwise the algorithm will not recognize that this is an unscheduled visit.

For example:

```
The Transformation Script
1 # Mapping using ODM element StudyEventData using value from attribute StudyEventOID
2 # Generalized for all StudyEvents
3 $VISITOID = xpath(/StudyEventData/@StudyEventOID);
4 if($VISITOID = 'UNS') {
5     $VS.VISIT = 'UNSCHEDULED';
6 } else {
7     $VS.VISIT = $VISITOID;
8 }
```

- In the mapping for "VISITNUM", there must be a "null" placeholder value for the unscheduled visit. For example:

```

The Transformation Script
1 $VISITNAME = xpath(/StudyEventData/@StudyEventOID/);
2 if(starts-with($VISITNAME, 'V')) {
3     $VISITNAME = replace($VISITNAME, '_', '');
4     $VS.VISITNUM = substring-after($VISITNAME, 'V');
5 } elseif(starts-with($VISITNAME, '█')) {
6     $VISITNAME = substring-after($VISITNAME, '█');
7     $VS.VISITNUM = concat('10', $VISITNAME);
8 } elseif($VISITNAME = 'UNS') {
9     $VS.VISITNUM = '';
10 } else {
11     $VS.VISITNUM = "-999";
12 }

```

Where a "null" (empty) placeholder visit number is assigned when in the source data, a visit is declared as "unscheduled" by the identifier "UNS"².

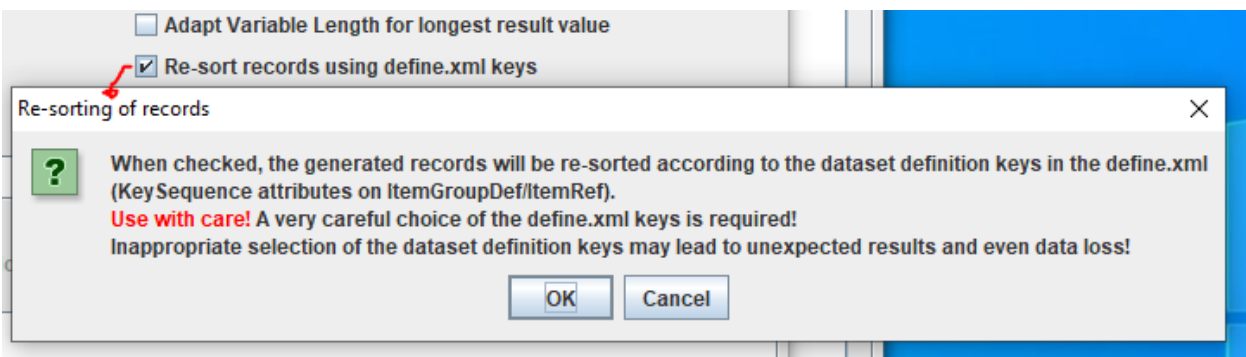
For the requirement that the data must be in chronological order, there are two possibilities:

- The source data is already ordered chronologically (as essentially required by the ODM standard, but not every EDC vendor adheres to this). In such a case, except for checking the checkbox "Perform post-processing unscheduled VISITNUM", nothing special needs to be done.
- The source data is **not** ordered chronologically (based on the date/time of collection). In that case, one will also need to check the checkbox "Re-sort records using define.xml keys"

Perform post-processing for assigning --LOBXFL
 Split records > 200 characters to SUPP-- records
 Move non-standard SDTM Variables to SUPP--
 Move Relrec Variables to Related Records (RELREC) domain
 View Result SDTM tables
 Generate 'NOT DONE' records for QS datasets
 Save Result SDTM tables as SAS XPORT files
 SAS XPORT files directory:

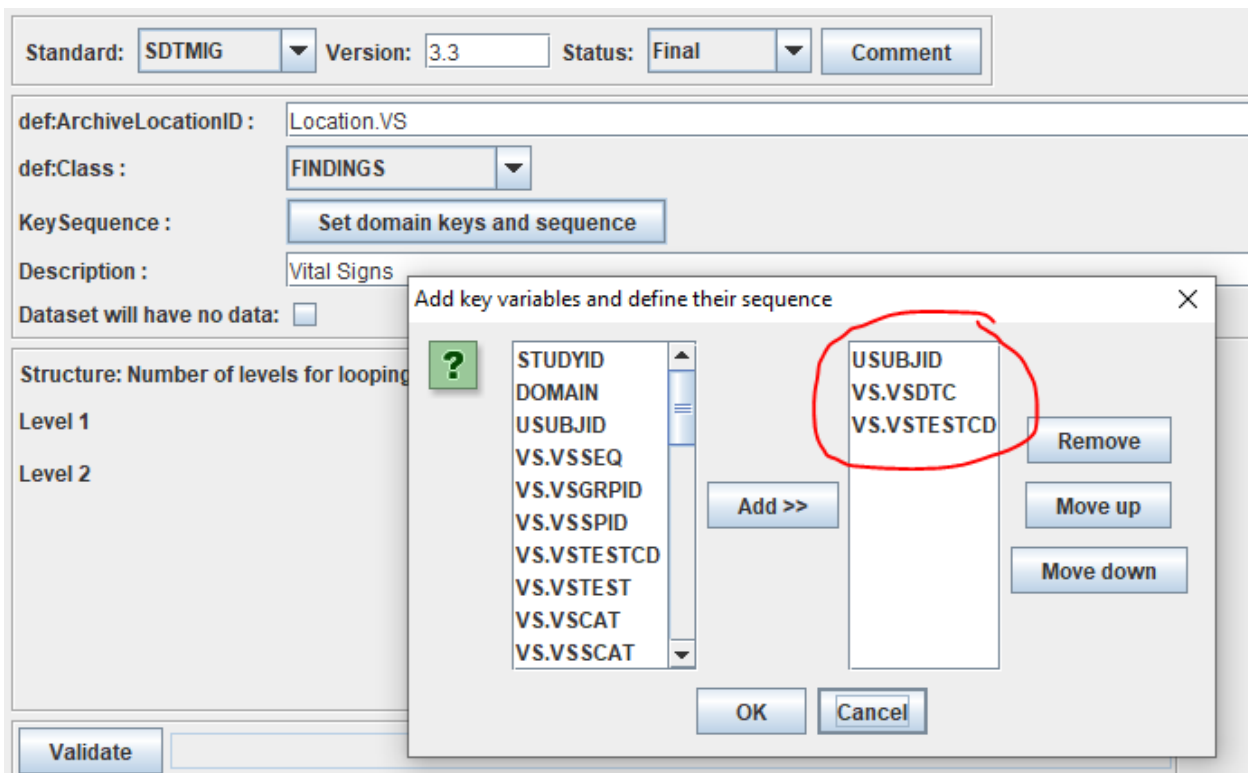
Perform post-processing unscheduled VISITNUM
 Move Comment Variables to Comments (CO) Domain
 Try to generate 1:N RELREC Relationships
 Adapt Variable Length for longest result value
 Re-sort records using define.xml keys
 Perform CDISC CORE validation on generated SAS XPORT files

which pops up an information dialog:



and take care that the "sorting keys" are well defined, meaning that e.g. -DTC is among them. For example:

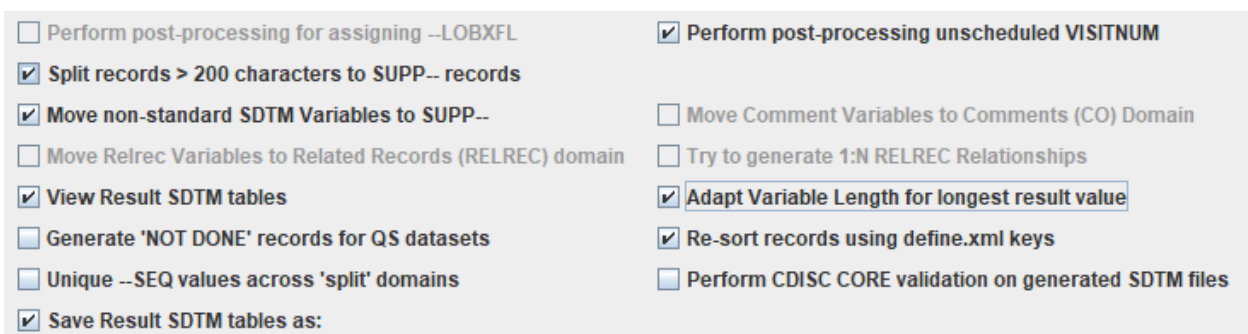
² There can be of course other ways the visit can be identified as "unscheduled" ...



IMPORTANT: Please take into account that using VISITNUM as a sort key is not a good idea here, as we want to generate VISITNUM ...³. Having it anyway may lead to undesired results, like duplicate records in the generated datasets.

Just as a reminder: setting the keys for a dataset (and for other dataset properties) can be achieved by a double-click on the first cell in the row for that dataset definition.

So, in our case, we would then e.g. have:



When then executing, this e.g. leads to:

³ In the final define.xml that is submitted, one can of course than use VISITNUM as a key again, though there is no extra advantage to do so.

VS.VSTESTCD	VS.VSTEST	VS.VSORRES	VS.VSORRESU	VS.VSSTAT	VS.VSREASND	VS.VSLOBXFL	VS.VISITNUM	VS.VISIT	VS.VSDTC
DIABP	Diastolic Blood Pres...	82	mmHg		Y		1 V1		2022-09-22T14:05
HR	Heart Rate	72	beats/min		Y		1 V1		2022-09-22T14:05
RESP	Respiratory Rate	16	breaths/min		Y		1 V1		2022-09-22T14:05
SYSBP	Systolic Blood Press...	129	mmHg		Y		1 V1		2022-09-22T14:05
TEMP	Temperature	36.8	C		Y		1 V1		2022-09-22T14:05
DIABP	Diastolic Blood Pres...	78	mmHg				1 V1		2022-09-22T15:30
HR	Heart Rate	70	beats/min				1 V1		2022-09-22T15:30
RESP	Respiratory Rate	14	breaths/min				1 V1		2022-09-22T15:30
SYSBP	Systolic Blood Press...	124	mmHg				1 V1		2022-09-22T15:30
TEMP	Temperature	36.7	C				1 V1		2022-09-22T15:30
DIABP	Diastolic Blood Pres...	86	mmHg				1.1 UNSCHEDULED 1.1		2022-09-26T11:00
HR	Heart Rate	77	beats/min				1.1 UNSCHEDULED 1.1		2022-09-26T11:00
RESP	Respiratory Rate	12	breaths/min				1.1 UNSCHEDULED 1.1		2022-09-26T11:00
SYSBP	Systolic Blood Press...	119	mmHg				1.1 UNSCHEDULED 1.1		2022-09-26T11:00
TEMP	Temperature	37.1	C				1.1 UNSCHEDULED 1.1		2022-09-26T11:00
DIABP	Diastolic Blood Pres...	87	mmHg				1.2 UNSCHEDULED 1.2		2022-09-26T12:00
HR	Heart Rate	78	beats/min				1.2 UNSCHEDULED 1.2		2022-09-26T12:00
RESP	Respiratory Rate	13	breaths/min				1.2 UNSCHEDULED 1.2		2022-09-26T12:00
SYSBP	Systolic Blood Press...	120	mmHg				1.2 UNSCHEDULED 1.2		2022-09-26T12:00
TEMP	Temperature	37.2	C				1.2 UNSCHEDULED 1.2		2022-09-26T12:00
DIABP	Diastolic Blood Pres...	80	mmHg				3 V3		2022-10-05
HR	Heart Rate	60	beats/min				3 V3		2022-10-05
RESP	Respiratory Rate	12	breaths/min				3 V3		2022-10-05

Where one sees that 2 unscheduled visit have been taken place between the planned visits V1 and V3, for which VISITNUM=1.1 and VISITNUM=1.2 has been assigned.

Important remark: if the data is not ordered chronologically, and one asks to assign VISITNUM for unscheduled visits automatically, incorrect values for VISITNUM will be assigned.

Some companies prefer to have VISITNUM for unscheduled visits still based on e.g. "99", but with a fraction added, e.g. "99.1", "99.2", meaning that it is not looked at what the value of VISITNUM was in the planned visit that took place immediately before the unscheduled visit. This is also possible in SDTM-ETL.

To do so, just assign "99" (or whatever other number one may want to use as the "base") to VISITNUM in the mapping script, e.g.:

```

0      $VISITNAME = SUBSTRING_INDEX($VISITNAME, '|', 1);
7      $VS.VISITNUM = concat('10', $VISITNAME);
8 } elseif($VISITNAME = 'UNS') {
9      $VS.VISITNUM = '99';
10 } else {
11     $VS.VISITNUM = "-999";
12 }

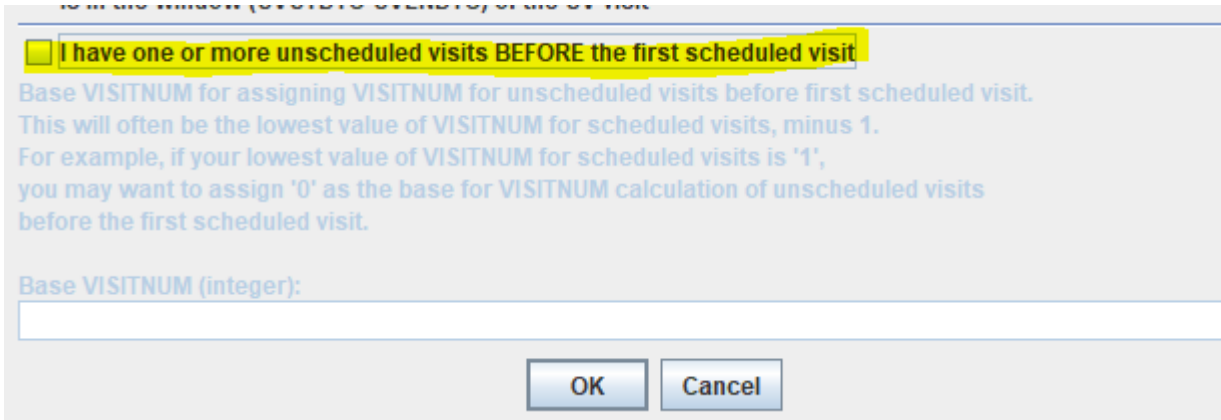
```

When then having the checkbox "Perform post-processing unscheduled VISITNUM" and having taken care that the data is ordered chronologically, the outcome will e.g. be:

One could state that "0" would then be the "anchor", and that for unscheduled visits before the first scheduled visit get visit numbers like 0.1, 0.2, ...

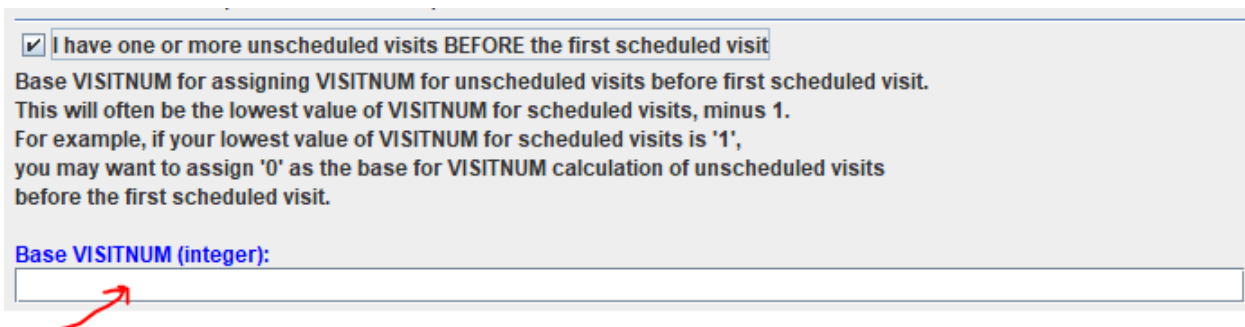
However, there is no rule at all that visit numbers for scheduled visits must start at 1. Some people e.g. assign VISITNUM=0 to the first scheduled visit, so in that case the "anchor" for unscheduled visits before first scheduled visit would e.g. be -1, or even -2, why not?

Therefore, the dialog about unscheduled visits has been extended:



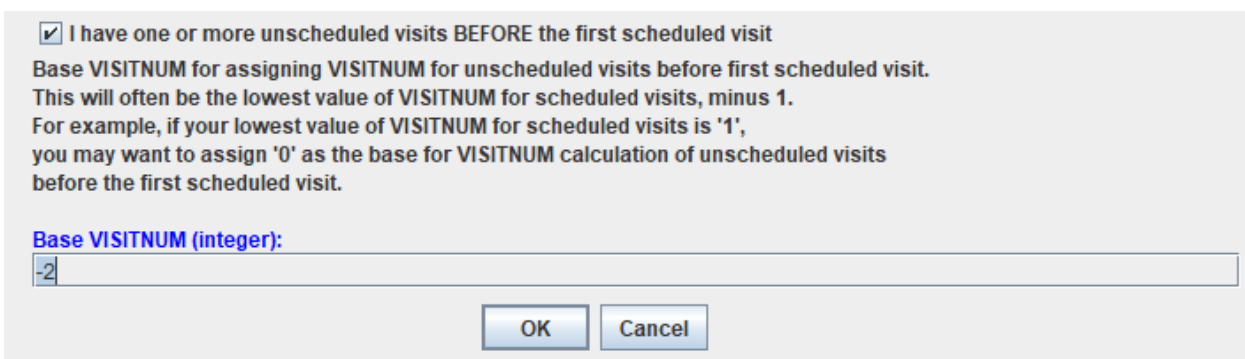
The screenshot shows a dialog box with a title bar. The first line is a checkbox labeled "I have one or more unscheduled visits BEFORE the first scheduled visit", which is currently unchecked and disabled. Below this is explanatory text: "Base VISITNUM for assigning VISITNUM for unscheduled visits before first scheduled visit. This will often be the lowest value of VISITNUM for scheduled visits, minus 1. For example, if your lowest value of VISITNUM for scheduled visits is '1', you may want to assign '0' as the base for VISITNUM calculation of unscheduled visits before the first scheduled visit." Below the text is a text input field labeled "Base VISITNUM (integer):" which is also disabled. At the bottom are "OK" and "Cancel" buttons.

when the checkbox "I have one or more unscheduled visits BEFORE the first scheduled visit" is checked, the other part becomes enabled:



This screenshot is identical to the previous one, but the checkbox "I have one or more unscheduled visits BEFORE the first scheduled visit" is now checked and enabled. Consequently, the text input field "Base VISITNUM (integer):" is also enabled and ready for input. A red arrow points to the input field.

And we can assign a base (anchor) integer number for calculation of unassigned visits that come before the first scheduled visit. When we e.g. set this to "-2":



This screenshot shows the dialog box with the checkbox checked and the input field containing the value "-2". The "OK" and "Cancel" buttons are visible at the bottom.

the final result e.g. becomes:

CES:DM	CES:SV	CES:LB	CES:VS	STUDYID	DOMAIN	USUBJID	SV.VISITNUM	SV.VISIT	SV.SVSTDTCT
CES	SV					001	-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44
CES	SV					001	-1.8	UNSCHEDULED -1.8	2010-01-01T22:13:05
CES	SV					001	0	BASELINE	2010-02-27T10:27:33
CES	SV					001	1	WEEK 1	2010-03-06T11:02:23
CES	SV					001	1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23
CES	SV					001	1.2	UNSCHEDULED 1.2	2010-03-07T14:06:21
CES	SV					001	2	WEEK 2	2010-03-13T09:12:28
CES	SV					001	2.1	UNSCHEDULED 2.1	2010-03-13T13:00
CES	SV					001	2.2	UNSCHEDULED 2.2	2010-03-13T14:00
CES	SV					001	2.3	UNSCHEDULED 2.3	2010-03-13T15:00
CES	SV					002	-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44
CES	SV					002	0	BASELINE	2010-02-27T10:27:33
CES	SV					002	1	WEEK 1	2010-03-06T11:02:23
CES	SV					002	1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23
CES	SV					002	1.2	UNSCHEDULED 1.2	2010-03-07T14:06:21
CES	SV					002	2	WEEK 2	2010-03-13T09:12:28
CES	SV					002	2.1	UNSCHEDULED 2.1	2010-03-13T13:00
CES	SV					002	2.2	UNSCHEDULED 2.2	2010-03-13T14:00
CES	SV					002	2.3	UNSCHEDULED 2.3	2010-03-13T15:00

for SV, and for LB:

CES:LB	CES:VS	NRHI	LB.LBNRIND	LB.LBBLFL	LB.VISITNUM	LB.VISIT	LB.LBD1
			NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44
			NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44
			NORMAL		-1.8	UNSCHEDULED -1.8	2010-01-01T22:13:05
			NORMAL		-1.8	UNSCHEDULED -1.8	2010-01-01T22:13:05
			NORMAL	Y	0	BASELINE	2010-02-27T10:27:33
			NORMAL	Y	0	BASELINE	2010-02-27T10:27:33
			NORMAL		1	WEEK 1	2010-03-06T11:02:23
			NORMAL		1	WEEK 1	2010-03-06T11:02:23
			NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23
			NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23
			NORMAL		1.2	UNSCHEDULED 1.2	2010-03-07T14:06:21
			NORMAL		1.2	UNSCHEDULED 1.2	2010-03-07T14:06:21
			NORMAL		2	WEEK 2	2010-03-13T09:12:28

We see that for subject 001, there are two unscheduled visits (with lab data being collected) before the first scheduled visit, which is assigned VISITNUM=0. As we assigned "-2" as the base for the calculation of VISITNUM for unscheduled visits before the first scheduled visit, these two unscheduled visits have been assigned VISITNUM "-1.9" and "-1.8" automatically.

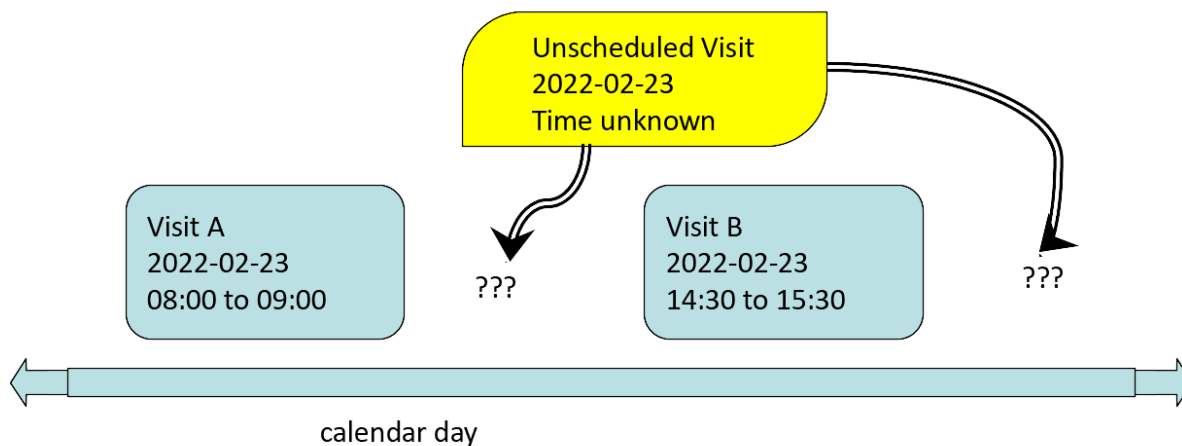
Trying to cope with data quality issues

Regularly, we see that users of the SDTM-ETL software obtain data for which there is no date available for the visit, which I regard as a "data quality issue". In such a case, there is nothing the software can do to correct for this. If the visit date is not provided, i.e. --DTC cannot be populated, then for each subject, when sorting by the collection date (as a key in the define.xml), these visits (of the data for them) will appear first in the list, before the visits (or data for them) for which a date is available.

We also often see that visit dates are delivered without a time part. This can become problematic for assigning VISITNUM, especially for unscheduled visits, when two (or more) have the same date provided. For example, when visit A and visit B come with the same date, there is no way to know whether visit A started before, or after, or even during, visit B.

In such a case, the sorting delivers the order in which visits A and B come in the source data⁴.

We have also seen cases where data of planned visits come with date and time, but for the unscheduled visits, only a date part without a time is available. If then planned visits A and B are on the same date (time known), and the unscheduled visit is on the same date, but the time is unknown, there is no way to know whether the unplanned visit became after visit A, or that it came after visit B. This may then easily lead to misalignment between VISITNUM in the dataset (usually a "Findings" domain) and in SV.



If the mapper however knows which one applies, he/she can use the checkboxes coming with "for overlapping visits":

Separation character: _____

For 'overlapping' visits

Base the 'unscheduled' VISITNUM on the **last** SV visit for which the --DTC of the non-SV visit is in the window (SVSTDTC-SVENDTC) of the SV visit

Base the 'unscheduled' VISITNUM on the **first** SV visit for which the --DTC of the non-SV visit is in the window (SVSTDTC-SVENDTC) of the SV visit

I have one or more unscheduled visits BEFORE the first scheduled visit

This is of course not an ideal solution, but it has helped some of our users to get alignment between VISITNUM for unscheduled visits in some of their Findings domains and in SV: one should be aware that in many findings domains, there will not always be data for each visit in SV, which complicates VISITNUM assignment for unscheduled visits when data quality (e.g. missing time part in --DTC variable values) is not good.

Alternative approaches for generating SV.

An alternative approach for generating SV with inclusion of unscheduled visits, is to generate it starting from all the other datasets when all the other datasets are ready and finalized. Essentially, this is just a pragmatic approach, even when it is not always scientifically sound (e.g. "propagating" assignment errors to SV).

⁴ Remark that the ODM standard requires that the data comes in chronological order in the ODM-XML file (oldest first). We do however observe many violations to this rule, especially when the data does not come from an EDC system.

We haven't implemented this in SDTM-ETL (yet), as it can also easily be done in R or SAS, as these are better suited for manipulating SAS-XPT datasets.

As soon however as FDA allows the use of Dataset-JSON 1.1, which is a much better, more modern and more flexible data format, and especially, really vendor neutral, we will however provide such a program together with SDTM-ETL.

Conclusions

VISITNUM should not be an SDTM variable. It looks as it is only there because reviewers at regulatory authorities wanted it as they are incapable to sort data by collection data (–DTC or –STDTC values). The introduction of VISITNUM (for "sorting") causes additional problems, rather than solving some, especially for unscheduled visits. When the data is already sorted chronologically, the use of "99" or "999" for VISITNUM surely makes sense. However, many sponsors want to have fractional numbers for VISITNUM for unscheduled visits, based on the VISITNUM of the last scheduled visit.

Different features to do so, using a post-processing step has now been added to SDTM-ETL, including for the case of unscheduled visits before the first scheduled visit.