

SDTM-ETL 4.4/4.6 User Manual and Tutorial

Author: Jozef Aerts, XML4Pharma

Last update: 2024-10-26

Handling Unscheduled Visits

Table of Contents

Table of Contents	1
Introduction.....	1
New feature in SDTM-ETL 4.4	2
Improvements in v.4.6.....	7
Limitations	9
Conclusions.....	10

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

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.

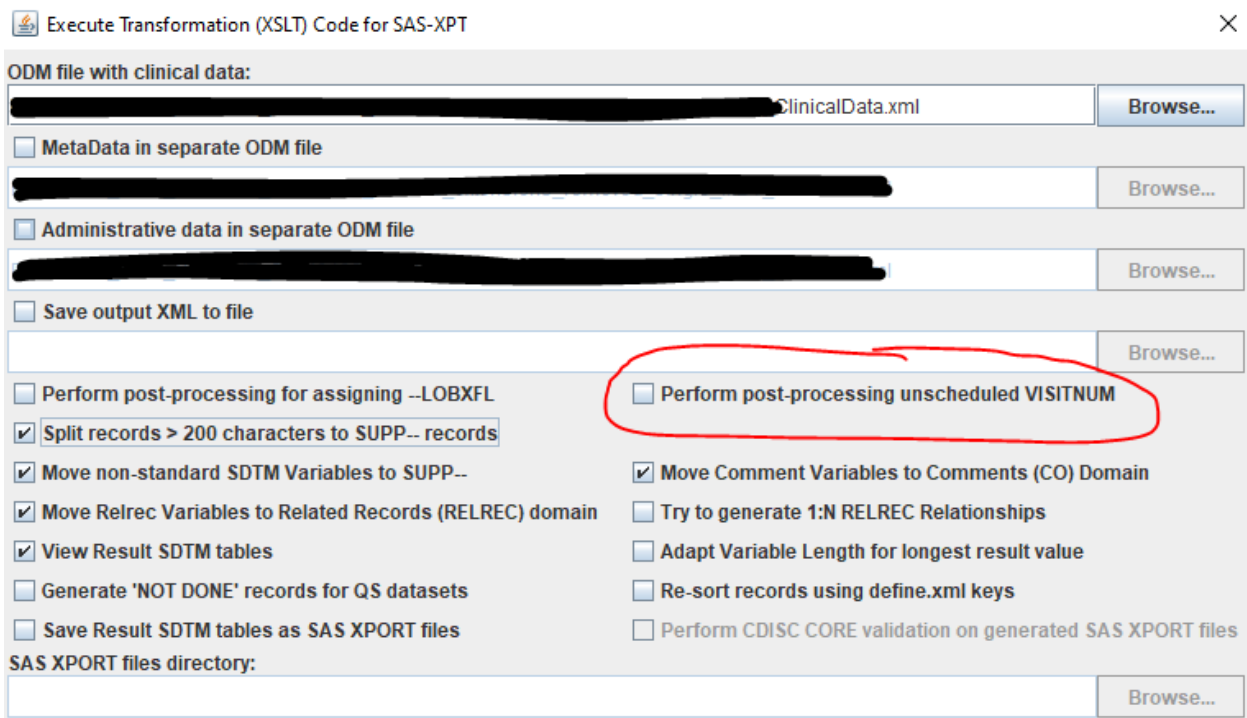
Another approach, also recommended by an, in the mean time retired, [Phuse paper](#), is to use fractional numbers (like "3.1", "3.2" ...) for VISITNUM for unscheduled visits.

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

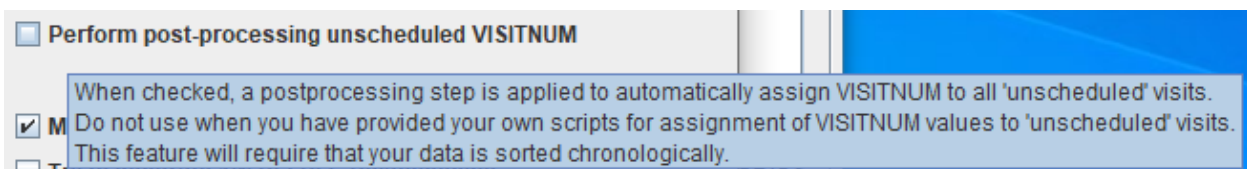
Our own recommendation is to not apply this additional post-processing step when the data already comes in chronological order, but some of our customers want to have it anyway. So we implemented it.

New feature in SDTM-ETL 4.4

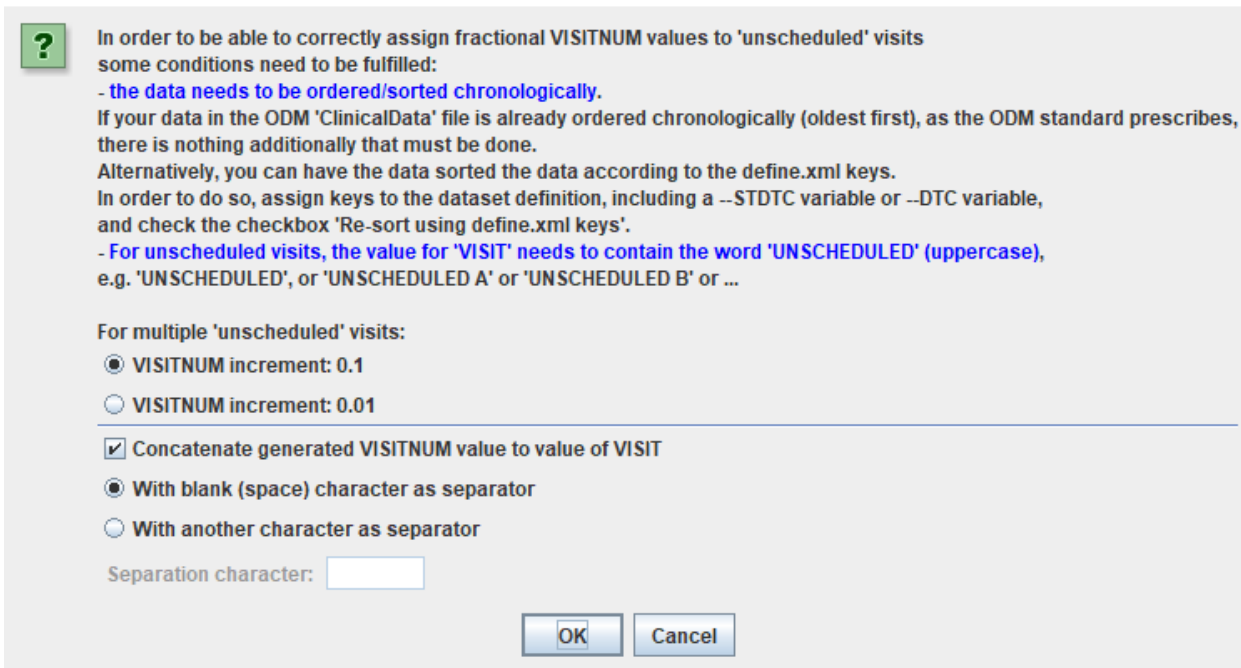
As of version 4.4, when executing the mappings, one will notice an additional checkbox in the "Execute Transformation (XSLT) code ..." dialog:



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



And when checks the checkbox, a dialog is presented:



As of version 4.6, update 1, the dialog contains more field. See section "Improvements in v.4.6".

Displaying the requirements for correct use, and an option to choose between "0.1" and "0.01" as the increment value, i.e. the increase in number value VISITNUM should get for each subsequent unscheduled visit.

By default, the value for VISIT will be extended with the visit number, separated by a blank.

For example, if, for the unscheduled visits, one has "\$VS.VISIT = 'UNSCHEDULED'", the final result will e.g. be "UNSCHEDULED 2.1", thus taking care of the 1:1 relation between VISIT and VISITNUM.

One can however switch this off by unchecking the checkbox "Concatenate generated VISITNUM value to value of VISIT. One can also provide another character (e.g. "-") As the separator instead of the blank character.

The requirements for correct use 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) by 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 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**
Save Result SDTM tables as SAS XPORT files Perform CDISC CORE validation on generated SAS XPORT files
SAS XPORT files directory: Browse...

which pops up an information dialog:

Adapt Variable Length for longest result value Re-sort records using define.xml keys

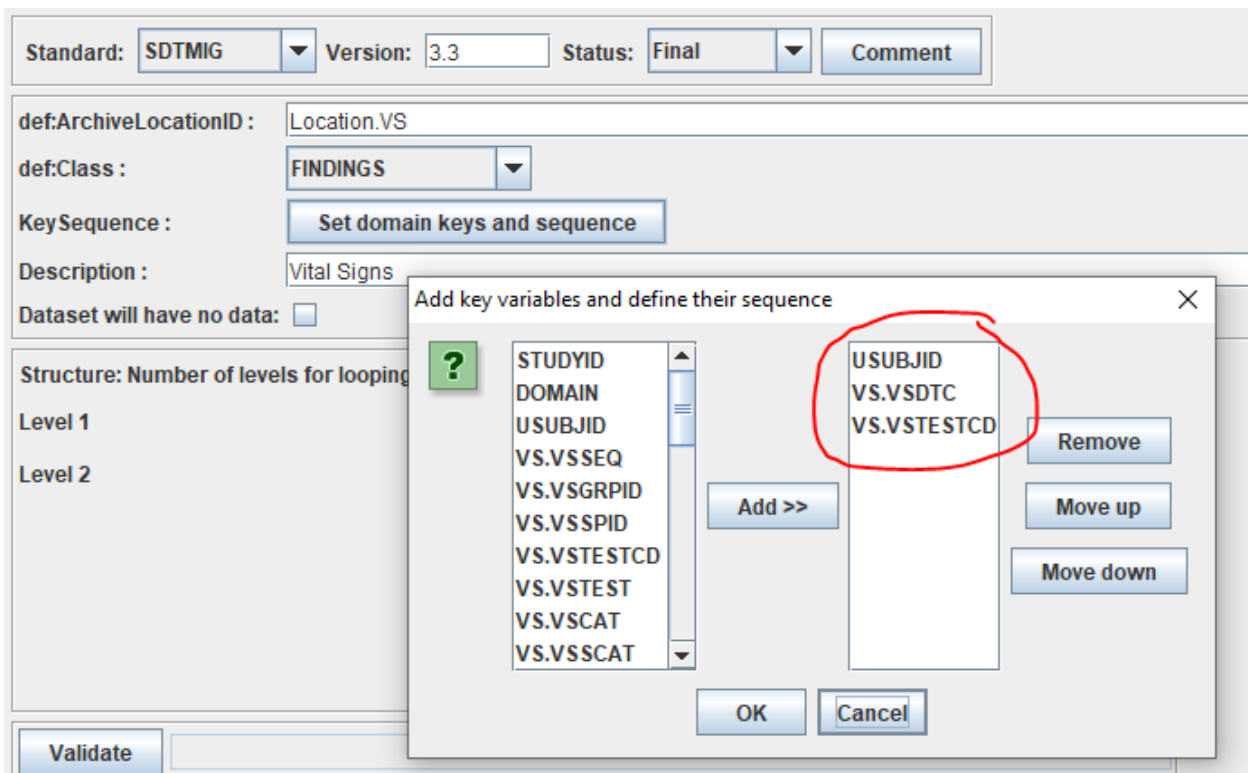
Re-sorting of records [X]

? When checked, the generated records will be re-sorted according to the dataset definition keys in the define.xml (KeySequence attributes on ItemGroupDef/ItemRef).
Use with care! A very careful choice of the define.xml keys is required!
Inappropriate selection of the dataset definition keys may lead to unexpected results and even data loss!

OK Cancel

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

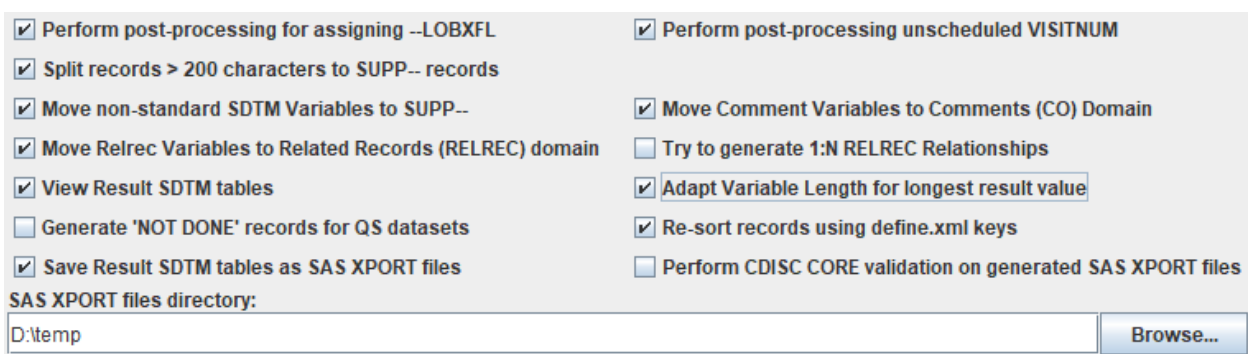
and take care that the "sorting keys" are well defined, meaning that e.g. -DTC is among them. For example:



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      $VISITNAME = SUBSTRING-ALICE($VISITNAME, FIND ),
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:

VS.VISITNUM	VS.VISIT	
1	V1	2022-01-01
1	V1	2022-01-01
1	V1	2022-01-01
1	V1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.1	UNSCHEDULED 99.1	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
99.2	UNSCHEDULED 99.2	2022-01-01
2	V2	2022-01-02

Improvements in v.4.6

We improved the algorithm for calculating VISITNUM for unscheduled visits in SDTM-ETL v.4.6. The old algorithm had the problem that when for a specific domain/dataset there was no data for a specific visit, it was not taken into account for the calculation of VISITNUM for unscheduled visits. For example, when lab data was collected in visits 4 and 6, but not in visit 5, but there was an unscheduled visit with lab data collected after visit 5 and before visit 6, then these rows would get VISITNUM=4.1 in the LB dataset, as "4" is the last value for VISITNUM available in the dataset. As of version 4.6, when also SV (Subject Visits) is loaded, the software will compare the date/times of SVSTDTC (visit start date/time) with the collection date/time (--DTC) of the rows for the unscheduled visits for the other datasets, and use the information from SV for assigning the VISITNUM for the unscheduled visits. So, for the above case, when the collection date/time in LB is before SVSTDTC of visit 5 in SV, then 4.1 will be assigned, but when it is after SVSTDTC of visit 5 in SV, then 5.1 will be assigned. Like this, VISITNUM values for unscheduled visits are completely synchronized between SV and the other datasets.

Conditions to have this working correctly are that the data is ordered or sorted chronologically within each subject, usually meaning that the first keys in the define.xml for the dataset are STUDYID, USUBJID and --DTC. Also, as said, SV must be loaded and come **before** the other datasets for which VISITNUM must be generated automatically. For SV, mappings must be present for VISITNUM and for SVSTDTC, as the latter is used for comparison of the date/times in order to correctly rank the record between the visits in SV.

A further improvement in v.4.6 (update 1) is for the case of unscheduled visits before the first scheduled visit. Essentially, there should be no visits at all before the first scheduled visit, as such visits should not belong to the study anyway. But we do also see that e.g. lab tests are done as "pre-screening" and then an "artificial", unscheduled visit is assigned to these which comes before the "screening visit" which is then taken as the first scheduled visit.

A problem here is that, in such a case, there is no "anchor" to assign a fractional number based on prior and next scheduled visit, as there is no prior scheduled visit.

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:

- For **unscheduled visits**, the value for 'VISIT' needs to contain the word 'UNSCHEDULED' (uppercase), e.g. 'UNSCHEDULED', or 'UNSCHEDULED A' or 'UNSCHEDULED B' or ...

For multiple 'unscheduled' visits:

- VISITNUM increment: 0.1
- VISITNUM increment: 0.01

Concatenate generated VISITNUM value to value of VISIT

- With blank (space) character as separator
- With another character as separator

Separation character:

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

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.

Base VISITNUM (integer):

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

With another character as separator

Separation character:

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

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.

Base VISITNUM (integer):

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":

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

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.

Base VISITNUM (integer):

the final result e.g. becomes:

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:

NRHI	LB.LBNRIND	LB.LBBLFL	LB.VISITNUM	LB.VISIT	LB.LBDTC
	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.

Limitations

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.

Another limitation is that when not involving SV into the mapping, there may be inaccuracies in assigning VISITNUM for unscheduled visits, when there is more than one such between two visits or before the first visit. As in such case, the start- and end date of the visit are not known (as these are in SV), the system cannot know whether data points of allegedly subsequent unscheduled visits belong together or not. For example, when we have mappings loaded for DM, LB (with "unscheduled visit" data points) and DM, the result of the assignments for LB is:

LB.LBSTNRHI	LB.LBNRIND	LB.LB...	LB.VISITNUM	LB.VISIT	LB.LBDTC	LB.LBDY
6.3	NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44	-57
10.7	NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T07:18:44	-57
6.3	NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T22:13:05	-57
10.7	NORMAL		-1.9	UNSCHEDULED -1.9	2010-01-01T22:13:05	-57
6.1	NORMAL	Y	0	BASELINE	2010-02-27T10:49:23	1
10.8	NORMAL	Y	0	BASELINE	2010-02-27T10:49:23	1
6.1	NORMAL		1	WEEK 1	2010-03-06T11:22:24	8
10.8	NORMAL		1	WEEK 1	2010-03-06T11:22:24	8
6.4	NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23	9
10.7	NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T09:22:23	9
6.3	NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T14:06:21	9
10.4	NORMAL		1.1	UNSCHEDULED 1.1	2010-03-07T14:06:21	9
6.1	NORMAL		2	WEEK 2	2010-03-13T09:22:23	15
10.8	NORMAL		2	WEEK 2	2010-03-13T09:22:23	15
6.4	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T13:00	15
10.7	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T13:00	15
6.3	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T14:00	15
10.4	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T14:00	15
6.3	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T15:00	15
10.4	NORMAL		2.1	UNSCHEDULED 2.1	2010-03-13T15:00	15

Without SV, the system cannot know whether the first 4 data points belong to the same unscheduled visit or not, especially as they are on the same day (but with different times). If we then look at the same results with SV being involved (see higher up), we see that these are two different unscheduled visits, so for the third and fourth row, "-1.8" should have been assigned. So, in order to get the most accurate assignments for VISITNUM for unscheduled visits, especially when there are several of them, it is of importance to have the mappings for SV loaded. Remark again that SV must then come **before** the other dataset definitions for which VISITNUM must be assigned automatically for unscheduled visits.

And not a limitation, but an important remark: if one wants to have VISITNUM generated for unscheduled records in an automated way, one should NOT have VISITNUM as a key for sorting the records for that domain/dataset or datasets. It must be clear that one cannot sort on something that needs to be generated in a post-processing step.

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

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.

sponsors want to have fractional numbers for VISITNUM for unscheduled visits, based on the VISITNUM of the last scheduled visit.

The feature 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.