

## Considerations in the Submission of Holter (EG) Data in an SDTM Compliant Format

Sophie Arnould, Novartis Pharma AG, Basel, Switzerland  
Gbenga Ogunsulire, Novartis Pharma AG, East Hanover, US  
Elsa Lozachmeur, Novartis Pharma AG, Basel, Switzerland  
Joseph Rowley, Novartis Pharma AG, Basel, Switzerland

### ABSTRACT

Major types of Electrocardiogram procedure include: Resting ECG and Ambulatory ECG. A Resting ECG usually takes 10 minutes while an Ambulatory or Holter ECG is performed using a portable device that is typically worn for 24 hours.

The SDTM EG domain is well suited for collecting resting ECG data. Mapping Holter data into the SDTM EG domain poses a number of challenges: (1) in EG and other findings domains timing variables describe collection time, but for Holter timing variables are also needed to describe timing of events; (2) certain Holter findings and measurements do not have a match in the NCI/CDISC terminology and it is left to the individual sponsor's discretion to decide if it is advantageous to use additional qualifier variables as an alternative to updating existing list of controlled terms.

This paper focuses on some of these challenges: the use of timing variable, supplemental qualifier domain and grouping variables are presented.

### INTRODUCTION – STDM BASICS

The SDTM is built around the concept of observations collected about subjects who participated in a clinical study. Each observation can be described by a series of variables. Each variable can be classified according to its **role**. A role determines the type of information conveyed by the variable about each distinct observation and how it can be used. Variables can be classified into five major **roles**:

- **Identifier variables**, such as those that identify the study, subject, domain, and sequence number of the record
- **Topic variables**, which specify the focus of the observation (such as the name of a ECG test or Examination)
- **Timing variables**, which describe the timing of the observation.
- **Qualifier variables**, which include additional illustrative text or numeric values that describe the results or additional traits of the observation (such as units or descriptive adjectives)
- **Rule variables**, which express an algorithm or executable method to define start, end, and branching or looping conditions in the Trial Design model

CDISC has also introduced the concept of **core** variable. A variable can be categorized as:

- **Required variable**: must be included in the dataset and cannot be null for any record
- **Expected variable**: any variable necessary to make a record useful, may contain some null values but in most cases will not contain null value for every record
- **Permissible variable**: variable should be used in a domain as appropriate when collected or derived.

Observations about study subjects are normally collected for all subjects in a series of **domains**. A domain is defined as a collection of logically related observations with a common **topic**. Example includes Electrocardiogram Test results (EG domain). Electrocardiogram data can be obtained from different procedures. Major types of Electrocardiogram procedure used in clinical trials include: **resting ECG**, this type of ECG usually takes 10 minutes; and **Ambulatory or Holter ECG**, this type of ECG is performed using a portable device that is typically worn for 24 hours. Resting and ambulatory ECG procedure do not differ in **Topic** and thus data collected via both procedure should go in the EG domain. Of note, assumption provided in the SDTMIG v3.1.2 for the EG domain only cover resting ECG procedure.

There are three general domain classes also referred to as **general observation classes**. These classes are as follows:

- **Interventions:** investigational treatments, therapeutic treatments and surgical procedures administered to the subject.
- **Events:** occurrences or incidents independent of planned study evaluations occurring during the trial or prior to the trial.
- **Findings:** observation resulting from planned evaluations (e.g. ECGs).

ECGs recordings include measurement of the electric activity of the heart such as Heart Rate measurements. ECG is also used to monitor the occurrence of certain cardiac events. As both types of parameters are collected from planned evaluations, it makes sense indeed to collect this information into the **Finding EG domain**.

In addition to these general domain classes, the SDTM defines several special-purpose domains that do not use the general-observation-class models. Examples relevant to this paper include the following:

- The Supplemental Qualifiers (SUPP--) datasets play an important role in the submission of variables that cannot be mapped into the standard domain variables. In this domain a set of variable is used to establish a **relationship** with the parent general-observation-class model
- The Related Records (RELREC) dataset is used to describe collected **relationships** between records in two (or more) dataset.

**Relationships** make use of the standard domain identifiers, STUDYID, DOMAIN and USUBJID. In addition variable IDVAR and IDVARVAL are used for identifying the record-level merge/join keys. Example of variables that could be used in IDVAR includes the following variables:

- --SEQ: uniquely identify a record for a given USUBJID within a domain
- --REFID: used to capture a sponsor-defined or external identifier
- --GRPID: used to link related records for a subject within a dataset or across dataset

Four possible methods can be used to link records from two datasets. For example if we want to link records from the EG domain to the DU (Device In-Use) domain we have

- **Method A:** many to many, using EGGRPID and DUGRPID
- **Method B:** one to many, using EGSEQ and DUGRPID
- **Method C:** many to one, using EGGRPID and DUSEQ
- **Method D:** one to one, using EGSEQ and DUSEQ

These methods could also be used to link EG and SUPPEG domains.

## CONSIDERATION FOR GROUPING VARIABLE IN THE EG DOMAIN

**Grouping data across subjects: --CAT, --SCAT**

Variable --CAT (category) and --SCAT (sub-category) are used to group records across subjects.

In the STD MIG, variable EGCAT is used to subset Measurement (e.g. Summary (Min) Heart rate, Summary (Max) Heart rate, Summary (Mean) Heart rate) or Findings (e.g. cardiac events such as AV 2:1 BLOCK, PREMATURE ARTRIAL COMPLEX,...). EGCAT values for these two subset of records are EGCAT = MEASUREMENT and EGCAT = FINDINGS respectively.

Measurements and findings from resting ECG procedure are collected at discreet time points. Holter devices are typically worn continuously for 24 hours. Measurement and findings from Holter recording can be summarized overall over the full period the Holter is worn, but more often a higher level of granularity is desired, typically hourly summaries will be provided. Extract of Holter recording at representative discreet time point may also be collected, these extract are called strip labels.

We have used variable EGSCAT to differentiate between ECG procedures, but also to differentiate between these modes of data collection. In this specific setting, EGSCAT can take up to 4 values: RESTING ECG, HOLTER ECG – OVERALL, HOLTER ECG – INTERVAL, HOLTER ECG - STRIP LABEL.

We have also evaluated if variable EGMETHOD could be used to differentiate between resting and ambulatory/Holter ECG procedure. EGMETHOD values must comply with the NCI/CDISC controlled terminology "EGMETHOD". For

both procedures the most relevant ECG method that we could find in the available list of term is “12 LEAD STANDARD”, accordingly from current definition of the NCI/CDISC controlled terminology EGMETHOD it is not possible to differentiate between ECG procedures. Accordingly we think it makes sense indeed to have a reference to the ECG procedure defined in EGSCAT as presented in the above paragraph.

**Grouping data within subjects: --GRPID, --SEQ, --GRPID, --REFID, --SPID**

Variable --GRPID can be used to group records within subject. Variable --SPID and --REFID are identifier variables, usually not considered to be grouping variables.

In our trials, EGREFID capture ECG **assessment identifier**; EGSPID capture **device identifier**. EGREFID can be used to establish relationship between ECG third party data and e-CRF ECG cover page. Usage of EGSPID variable could be used to establish relationship between device properties (DO domain/DI domain) and parent EG domain, but in our trials EGSPID was most useful to track if an ambulatory device was replaced or possibly interrupted during a given ECG assessment. Such interruption and device change are common with ambulatory procedure.

We have made use of EGGRPID to tie together several observations from the same strip label in the EG domain, as shown in example 1. In this case, a grouping variable is needed notably because timing variable may not coincide between observation from such set of related records, but also to facilitate review by medical personnel.

We have also made use of EGGRPID to establish relationship between EG domain and other domains (e.g. SUPPEG, RELREC). Beside the USUBJID and DOMAIN variables, variable IDVAR is needed to establish the relationship between a children dataset (e.g. SUPPxx, RELREC) and a parent dataset (e.g. EG). IDVAR values may contain the name of one of the following merge/key variables: --SEQ (unique subject record sequence number), --GRPID, --REFID, --SPID. Care should be taken to ensure that related children records do not accidentally join back to incorrect parent records.

**EG domain - Example 1**

VISIT	EGSCAT	EGCAT	EGTPT	EGGRPID	EG-REFID	EG-SPID	EGTEST	EGSTRESC
DAY1	RESTING ECG	MEASUREMENT	PRE		331PT8		Summary Heart Rate (Mean)	75
DAY1	RESTING ECG	MEASUREMENT	1 HOUR POST		332PT8		Summary Heart Rate (Mean)	65
DAY1	RESTING ECG	...	...		...			
DAY1	RESTING ECG	MEASUREMENT	6 HOURS POST		336PT89		Summary Heart Rate (Mean)	55
DAY1	HOLTER ECG OVERALL	MEASUREMENT			55HO81		Summary Heart Rate (Mean)	
DAY1	HOLTER ECG INTERVAL	MEASUREMENT	PRE	DAY1 - PRE	55HO81	Dev991	Summary Heart Rate (Mean)	72
DAY1	HOLTER ECG INTERVAL	MEASUREMENT	1 HOUR POST	DAY1 - 1HR_POST	55HO81	Dev991	Summary Heart Rate (Mean)	62
DAY1	HOLTER ECG INTERVAL		...		...			
DAY1	HOLTER ECG OVERALL	FINDING					Technical condition	Device replaced 6 hours post-dose after an interruption of about 3 hours.
DAY1	HOLTER ECG INTERVAL	MEASUREMENT	23 HOURS POST	DAY1 - 23HR_POST	55HO81	Dev554	Summary Heart Rate (Mean)	75
DAY1	HOLTER ECG STRIP LABEL	MEASUREMENT		1	55HO81	Dev554	Summary Heart Rate (Mean)	55
DAY1	HOLTER ECG	FINDING		1	55HO81		Ectopy	PREMATURE ARTRIAL

	STRIP LABEL							COMPLEX
DAY1	HOLTER ECG -- STRIP LABEL	FINDING		1	55HO 81		Technical problem	ARTIFACT
DAY1	HOLTER ECG -- STRIP LABEL	FINDING		1	55HO 81		Conduction	AV 2:1 MOBITZ

In the section “CONSIDERATION FOR DATE/TIME VARIABLE IN THE EG DOMAIN”, we provide an example of relationship between SUPPEG and EG domain: in this example IDVAR was set to EGSEQ. We also provide a second example of relationship between EG and DU (device-in-use) domain and setting-up of the RELREQ dataset. In this second example, IDVAR value in the parent EG domain was set to EGGRPID. Only one variable name can be specified in IDVAR: in our case, variable EGREPID, EGSPID or natural key variables (i.e. STUDYID, USUBJID, EGTESTCD, VISITNUM, EGPTREF, EGPTNUM) could not be used in isolation to establish correct relationship between EG and DU via RELREC dataset: in this scenario EGGRPID is the outcome of combination from several variables

## CONSIDERATION FOR DATE/TIME VARIABLE IN THE EG DOMAIN

### -- DTC, --STDT, -- ENDTC

For any domain based on the Findings general observation class, if evaluation is performed at single-point in time, the date/time of the evaluation should go into the --DTC field. For example with resting ECG, the EGDTC capture the date/time of the discreet reading.

For evaluation performed over an interval of time, the start date/time of collection should go into the --DTC field and the end date/time of collection should go into the --ENDTC field. For example with HOLTER ECG, cardiac evaluation is performed over an interval of time, and the start date/time of each interval goes into EGDTC; the end date/time goes into EGENDTC.

The --STDT variable should not be used in a Finding domain.

A sample of date/time variables for the EG Domain is provided below (see “EG domain – Example 2”).

When additional dates come into play, it becomes confusing how to populate timing variables within a finding domain, for example in addition to the interval-collection-time, it may be relevant to collect event-time. With resting ECG, this does not pose problem as event-time coincide with the discrete collection- time of ECG. With Holter ECG, for certain parameters (e.g. Summary (Min) Heart rate, cardiac findings ...) , it may be relevant to collect an event-time (e.g. time of minimum Heart rate) but this event-time usually do not coincide with the interval-collection-time; for other parameters (e.g. Summary (Mean) Heart rate, hourly counts of certain cardiac events of interest,...), it is not relevant to capture an event-time.

We propose two alternative approaches to capture interval-collection-time and event-time all together.

The **first approach** is based on the SDTMIG assumption that for finding domain collection-time should be captured in --DTC and --ENDTC. As we could not map event-time to a standard EG domain variable, this additional information is collected in a supplemental qualifier (SUPPEG) dataset. The relationship between EG and SUPPEG is a one-to-one relationship. In one-to-one relationship IDVAR is set up to EGSEQ. This approach is illustrated by the following example:

### EG domain – Example 2

EGSEQ	EGSCAT	EGTEST	EGSTRESC	EGTPT Time point Name	EGDTC Collection start date/time	EGENDTC Collection End date/time
1	RESTING ECG	Summary (Min) Heart rate	55	1 HOUR POST	2013-04-01T8:12	
2	RESTING ECG	CONDUCTION	AV 2:1 BLOCK	1 HOUR POST	2013-04-01T8:12	
3	HOLTER ECG - INTERVAL	Summary (Mean) Heart rate	72	1 HOUR POST	2013-04-01T8:02	2013-04-01T9-15
4	HOLTER ECG - INTERVAL	Summary (Min) Heart rate	45	1 HOUR POST	2013-04-01T8:02	2013-04-02T9-15
5	HOLTER ECG - INTERVAL	CONDUCTION	AV 2:1 BLOCK	1 HOUR POST	2013-04-01T8:02	2013-04-02T9-15

**SUPPEG: supplemental qualifier for EG**

QNAM	QLABEL	QVAL	IDVAR	IDVAL
EVDTTC	Event start date/time	2013-04-01T15:52	EGSEQ	4
EVDTTC	Event start date/time	2013-04-01T16:12	EGSEQ	5
EVENDTC	Event end date/time	2013-04-01T16:30	EGSEQ	5

In the **second approach**, we have judged interval-collection-time could be considered machine generated information and should be capture in another appropriate general observation class domain. A Device supplement to the SDTMIG is available since January 2012. This implementation guide describes seven new device SDTM domains. Of these, the Device In-Use (DU) domain is a Finding domain that contains the values of measurements and settings that are intentionally set on a device. In the below example, event-time is captured in the EG domain and the DU domain captures Holter interval-collection-time for HOLTER ECG – OVERALL and for HOLTER ECG – INTERVAL observations; for other EGSCAT categories (i.e. resting ECG and HOLTER ECG – Strip label) collection-time and event-time coincide. Relationship between DU and EG domain are established in the RELREC dataset:

**EG domain (capture event-time):**

EG-SEQ	EGCAT	EGTEST	EGSTRESC	EGGRPID	EGDTC	EGENDTC
					start date/time	End date/time
1	RESTING	Summary (Mean) Heart rate	55		2013-04-01T7:50	
2	HOLTER	Summary (Mean) Heart rate	72	DAY1 1HR_POST		
3	HOLTER	Summary (Min) Heart rate	45	DAY1 1HR_POST	2013-04-01T9:32	
4	HOLTER	CONDUCTION	AV 2:1 BLOCK	DAY1 1HR_POST	2013-04-01T 9:42	2013-04-01T9:45

**DU domain (capture Holter interval-collection time):**

DUGRPID	DUDTC	DUENDTC	DUTPT
DAY1 – 1HR_POST	2013-04-01T8:52	2013-04-02T9-52	0-24 HR POST

**RELREC domain (establish relationship between EG and DU domains):**

RDOMAIN	IDVAR	IDVARVAL	RELTYPE	RELID
DU	DUGRPID	DAY1 – 1HR_POST	ONE	A
EG	EGGRPID	DAY1 – 1HR_POST	MANY	A

This approach has a number of pitfalls: (1) EGDTC is left blank for certain records (e.g. EGSEQ = 2), which could potentially be all right since EGDTC is an expected variable. However, we were concerned that allowing missing values in this case, may be an over-interpretation of the definition of an expected variable. (2) In this paper, we have made limited use of variable EGGRPID. It may turn out that EGGRPID may be required to establish other type of relationships. (3) With RESTING-ECG and HOLTER-ECG STRIP LABEL collection-time is captured; with HOLTER-ECG OVERALL and HOLTER-ECG PERIOD event time is captured: such inconsistencies between both ECG procedure and mode of data collection may ultimately confuse the reviewer.

Accordingly, we have opted for the first approach.

**CONTROLLED TERMINOLOGY FOR EG PARAMETERS**

In the NCI/CDISC controlled terminology, we have a codelist associated to ECG test (i.e. EGTEST) which is used to code ECG measurement, we also have a codelist associated to ECG results (i.e. EGSTRESC) which is used to code ECG findings.

ECG is employed to identify the presence of certain cardiac events, as an example “AV 2:1 BLOCK”. When such event is observed, Standardized result variable (i.e. EGSTRESC) is set-up to “AV 2:1 BLOCK”, this value is defined in the NCI/CDISC codelist EGSTRESC which associated to variable EGSTRESC.

“AV 2:1 BLOCK” that last less than 3 seconds don’t have the same clinical significance as compared to “AV 2:1 BLOCK” that last greater or equal to 3 seconds. With Holter procedure it is also of interest to collect the number of occurrence of these two types of event per Hour.

This could be transcribed into the requirement of two news ECG tests/measurements: EGTEST = “AV 2:1 BLOCK < 3 Sec (counts)” and EGTEST = “AV 2:1 BLOCK >= 3 Sec (counts)”. Such requirement should be submitted to NCI/CDISC for approval which may be a lengthy process. Accordingly, the sponsor may decide to update in-house codelist at risk of not getting approval from NCI/CDISC.

Other alternative could be to use --RESCAT in combination with a non-standard variable in SUPPEG domain as described here:

**-- RESCAT**

The SDTMIG state that this variable can be used to categorize results in a standard format after the fact, i.e. after assessing the value of Findings results. We think that SDTMIG meant this to be associated with a result with say a 2 step process: you do a test and get the results and then using this result it is associated to a result categorization. For example this variable could be used to categorize tumor finding in Malignant or Benign. --RESCAT variable should not be used like another variable to add more information.

Accordingly we thought we may apply this concept in categorizing finding into ">= 3 sec" or "<3 sec" The same concept could have been applied to categorization of certain cardiac events (e.g. Supraventricular Tachycardia) into "Sustained" or "Non-sustained. An episode of Supraventricular Tachycardia that last < 30 seconds is considered non-sustained, an episode of Supraventricular Tachycardia that last >= 30 seconds is considered sustained. "SUSTAINED SUPRAVENTRICULAR TACHYCARDIA" is available in the controlled terminology EGSTRESC, which let us think that it may be more appropriate to have such characterization of finding as part of the results itself and not contained into an additional grouping/categorization variable.

**CONSIDER A NON-STANDARD VARIABLE TO COLLECT THE NUMBER OF EPISODES: USAGE OF SUPPEG DOMAIN**

The number of episodes per hour could be mapped into a non-standard variable into the SUPPEG domain as follows:

**EG domain**

VISIT	EGSCAT	EGCAT	EGTPT	EGGRPID	EGRESCAT	EGTEST	EGSTRESC
DAY1	HOLTER ECG - INTERVAL	FINDIN G	1 HOUR POST	DAY1_ 1HR_POST _>=3sec	>= 3 SEC	Conduction	AV 2:1 MOBITZ

**SUPPEG: supplemental qualifier for EG**

QNAM	QLABEL	QVAL	IDVAR	IDVAL
COUNT	Number of episode	52	EGGRPID	DAY1_1HR_POST _>=3sec

In the present instance, our preference was to submit the new EGTEST = "AV 2:1 BLOCK < 3 Sec (counts)" and EGTEST = "AV 2:1 BLOCK >= 3 Sec (counts)" to NCI/CDISC at risk. In the perspective this new codelist values get rejected by NCI/CDISC, we would still have the chance to go for the alternative proposal (i.e. combined usage of EGRESCAT and SUPPEG to collect number of episode).

**CONCLUSION**

This paper highlights the difficulties of applying SDTM assumption when both interval-collection-time and event-time are of interest in a Finding domain. This paper is an opportunity for us to get advice and possibly influence key opinion leader in further implementing the SDTM model: our key message is that we think it would be beneficial to have two set of timing variables defined in finding domains one for event-time the other one for collection-time.

This paper also presents some challenges difficulties with collecting additional information about findings. We could not find a standard way to collect this additional information: either we had to update existing controlled terminology or we had to make usage of SUPPEG to collect non-standard variables: only one non-standard variable was sufficient to collect this complementary information, if indeed the usage of EGRESCAT described in this paper is appropriate. Our message is that it would give flexibility to have new standard variables for the purpose of collecting complementary information about findings. The solution that we have adopted instead was to update existing controlled terminology at risk: about 50 new values have been added and still requires to be approved by NCI/CDISC.

**REFERENCES**

www.cdisc.org  
 Study Data Tabulation Model (SDTM) Version 1.2  
 SDTM Implementation Guideline (SDTMIG) Version 3.1.2  
 CDISC SDTM Controlled Terminology, 2012-12-21

## **ACKNOWLEDGMENTS**

We would like to thank Sebastien Bostel and Kathy Mellars from the standard governance board in Novartis for their support in implementing SDTM aspect for Holter data.

## **CONTACT INFORMATION**

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

Sophie Arnould

Novartis Pharma AG  
Postfach  
CH-4002 Basel  
SWITZERLAND

Work Phone: +41 61 3245176

Email: [sophie.arnould@novartis.com](mailto:sophie.arnould@novartis.com)