Paper SI07

Gear SDTM up with Digital Health Devices

Vicky Poulsen, Novo Nordisk A/S, Copenhagen, Denmark

ABSTRACT

The advance of digital health technologies, such as continuous glucose monitoring (CGM), wearable actigraphy devices, wearable ECG monitors…to name a few, in recent years has significantly broadened the prospect of real-time, high-quality health data collection. This has been witnessed by the growing number of clinical studies that engage the use of voluminous device and/or omics data.

Getting large volume of device data into SDTM, however, is not without its challenges. Worsening performance, validation issues and use of Device Domains, all require extra attentions. In this paper, we will share Novo Nordisk's experience and learnings from preparing submission ready SDTM data that include multiple millions of CGM readings.

INTRODUCTION

In recent years, the advance of digital health devices that enable continuous monitoring has significantly changed the way data are collected, processed and analysed in clinical studies. This paper will discuss the challenges/considerations of including voluminous real-time continuous glucose monitoring (CGM) readings in SDTM and illustrate Novo Nordisk's (NN) approach that balances technical operationality, data usefulness as well as standard and submission compliance through the example of *Study A*.

Study A is a Phase 3a study with 1000 participants and a primary objective to investigate the effect of *COMPOUND X* over 78 weeks. The study includes ten 14-day periods where glucose values are collected at 5-minute intervals using CGM *Model TX* to assess glycaemic control, a confirmatory secondary endpoint, of individual participants.

CONTINUOUS GLUCOSE MONITORING

The increasing adoption of CGM devices, which collect real-time glucose readings at regular intervals, and accompanying ambulatory glucose profile (AGP)[1], which serves as a fast and easy way to understand CGM readings, in daily diabetes self-management has extended into clinical studies enabling evaluation of glycaemic regulation with targets other than HbA_{1c}.

A CGM device typically consists of three components:

1) A sensor placed underneath the skin into the subcutaneous tissue
2) A transmitter attached to the sensor
3) A receiver that displays and stores glucose information

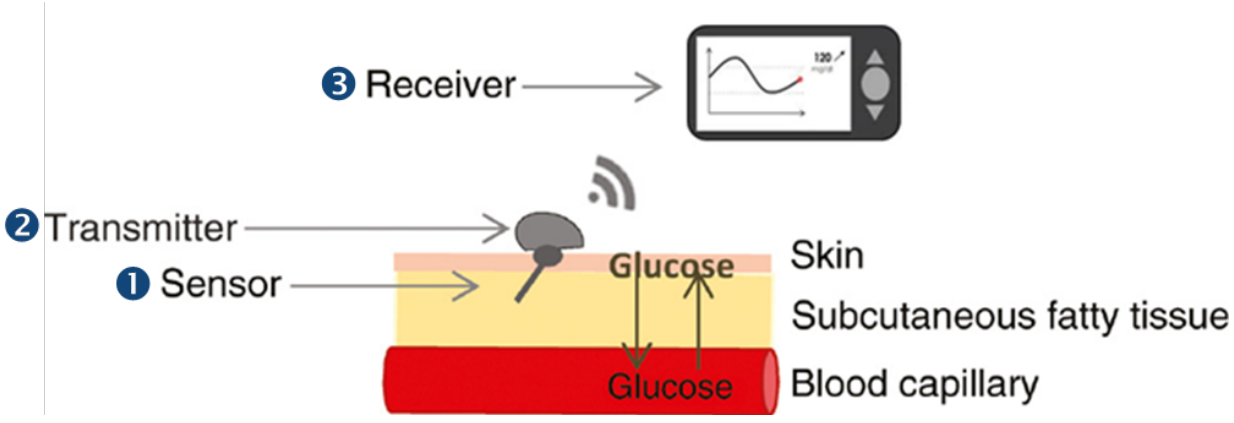


Figure 1 Components of a Continuous Glucose Monitoring System

The sensor detects glucose levels via enzymatic or non-enzymatic reaction with glucose in the interstitial fluid and creates a signal that is converted in the transmitter into a glucose value and sent to the receiver. In a clinical study, all glucose values stored in the receiver (CGM readings) are transferred to the sponsor for inclusion in SDTM and subsequent analysis.

An AGP contains a statistical summary of glucose metrics, based on CGM readings, for assessing glycaemic control. At least 10 to 14 days of CGM readings are required to establish an AGP. Furthermore, a minimum of 70% activity of the overall CGM period[2] is needed to achieve reliable analysis of glycaemic indices.

CGM READINGS IN SDTM: CHALLENGES AND CONSIDERATIONS

In a clinical study, CGM readings can be employed to assess glycaemic control of an intervention at planned periods. The inclusion of CGM readings, however, poses a number of operational challenges and implementation considerations to the creation and use of SDTM data.

DATA VOLUME

For each participant, a CGM period of 14 days with glucose level measured every five minutes will incur 4 032 CGM readings. In *Study A*, a minimum of 28 224 000 CGM readings, corresponding to 70% of full CGM coverage required to achieve reliable analysis of glycaemic indices, would be added in SDTM. The number will mount up to 40.3 million if all participants attain 100% activity in all CGM periods. The data volume of *Study A* will increase by over 2 000 times compares with studies using only blood glucose measurements. Such increase in data volume inflicts pressure on NN's CDMS as well as the overall workflow as they are designed originally to manage relatively small data collected on case report forms (CRF).

PROCESSING TIME

Glucose values from CGM readings are to be reported in Laboratory Test Results (LB) domain together with results from other 2600 tests present in the Laboratory Test (LBTEST and LBTESTCD) CDISC codelists. On top of the increased processing time required to load tens of millions CGM readings together with other laboratory test results, the addition of these results to LB domain will also adversely impact the use of the domain in the analysis process of *Study A*. A lot of processing time will be required for data filtering when analysis or reporting is to be performed on non CGM readings which apart from being inefficient also put NN's Statistical Computing Environment (SCE) as a whole under stress.

STANDARD COMPLIANCE

In addition to glucose measurements in LB domain, the use of CGM needs also to be reported in a number of device domains in accordance with SDTMIG-MD version 1.1. Which device domains to use and how to represent CGM, a device that comprises of 3 components, are amongst considerations that are important for *Study A* where CGM is employed to collect data crucial for assessment of the study's confirmatory secondary endpoint.

NOVO NORDISK'S APPROACH

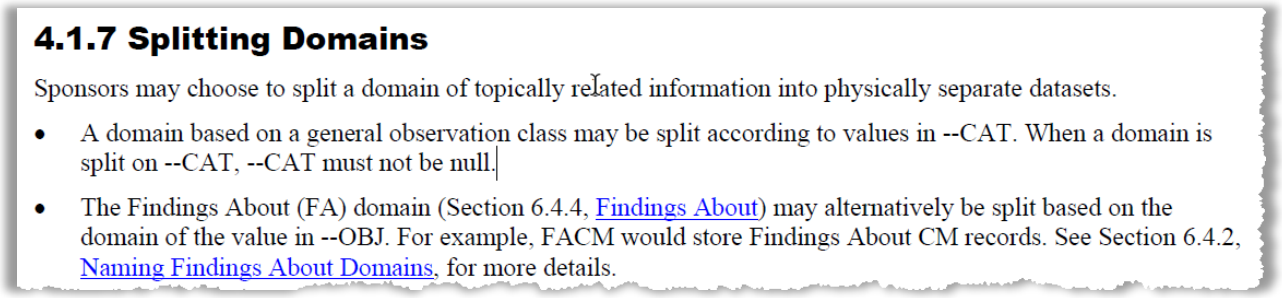
In light of the apparent operational challenges of processing CGM readings and the prospect of adopting other digital health technologies, NN's Global Standards Team (GST), the governance body of global standard implementation, sets up a Device Subgroup to identify and implement a viable solution that balances technical operationality, data usefulness as well as standard and submission compliance to handle voluminous data.

RESOLVING WORKFLOW AND CDMS CHALLENGES

Like eDiary, data captured by digital health devices such as CGM readings are not subject to data cleaning routines hence there is no actual need of loading the data into CDMS. The obvious answer would be to not use a system when there is no need to. By developing routines to QC and process CGM readings directly in SCE, the overall workflow of device data handling can be streamlined and the unnecessary operation pressure on CDMS removed.

OVERCOMING PROCESSING TIME CHALLENGES

The solution to processing time challenge is found in SDTMIG 3.3, Section 4.1.7 Splitting Domains (Figure 2).



4.1.7 Splitting Domains

Sponsors may choose to split a domain of topically related information into physically separate datasets.

- A domain based on a general observation class may be split according to values in --CAT. When a domain is split on --CAT, --CAT must not be null.
- The Findings About (FA) domain (Section 6.4.4, Findings About) may alternatively be split based on the domain of the value in --OBJ. For example, FACM would store Findings About CM records. See Section 6.4.2, Naming Findings About Domains, for more details.

Figure 2 Splitting Domains (Section 4.1.7, SDTMIG 3.3)

The processing time required for 1) loading CGM readings into LB domain can be reduced substantially and 2) data filtering eliminated entirely by splitting records in LB domain into 2 separate physical datasets (Figure 3):

- LBC: CGM readings (LBCAT="CGM")
- LB: all other laboratory test results

The separate physical datasets are referred to as *Stack* datasets to avoid being confused with the split of transport file requirement detailed in FDA's Technical Conformance Guide (Figure 4).

LB (Laboratory Test Results) LBC (CGM Readings) PE (Physical Examination) QS (Questionnaires) RP (Reproductive System Findings) VS (Vital Signs) XL (Laboratory Test Results-US Units) XV (Vital Signs-US Units)	LB	Laboratory Test Results	FINDINGS	One record per analyte per planned time point number per time point reference per visit per subject	Tabulation
	LBC	CGM Readings (Laboratory Test Results)	FINDINGS	One record per reading per subject	Tabulation

Figure 3 Excerpt from define.xml demonstrating split of LB domain into 2 datasets

3.3.2 Dataset Size

Each dataset should be provided in a single transport file. The maximum size of an individual dataset that FDA can process depends on many factors. Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB. Sponsors should submit these smaller datasets, in addition to the larger non-split datasets, to better support regulatory reviewers. The split datasets should be placed in a separate sub-directory labeled 'split' (See section 7.2). A clear explanation regarding how these datasets were split needs to be presented within the relevant data RG.

Figure 4 FDA's requirement on splitting transport file larger than 5 GB

The *Stack* datasets concept has been adopted as part of NN global standard implementation and its use governed by NN's GST accordingly. In principle, all data captured by digital health devices could be in scope for *Stack* datasets implementation. In practice, an end-to-end assessment would be conducted per data type and source by the Device Subgroup to evaluate the suitability of *Stack* datasets implementation in order to ensure standard compliance and data integrity. Currently, *Stack* datasets are allowed in LB domain for CGM and FGM (Flash Glucose Monitoring) data and in XO domain (a custom domain for OMICS Findings) for non-genomics data.

ENSURING STANDARD COMPLIANCE

One of the main challenges in the use of device domains (Table 1) is the construct of SPDEVID (Sponsor Device Identifier). According to SDTMIG-MD, SPDEVID is "a surrogate identifier that represents all the characteristics of a device" and used to identify a device uniquely and consistently across domains.

Table 1 Device Domains

Study Reference

- *Device Identifiers (DI)*: contains consistent sponsor-defined variable (SPDEVID) for linking data across domains. DEVTYPE DIPARMCD must be included as a minimum.

Events

- *Device Events (DE)*: contains information about various kinds of device-related events, such as device malfunctions.
- *Device Tracking (DT)*: represents a record of tracking events for a given device (e.g., initial shipment, deployment, return, destruction).

Findings

- *Device-In-Use (DU)*: contains the values of measurements and settings that are intentionally set on a device when it is used.
- *Device Properties (DO)*: contains characteristics of the device.

Interventions

- *Device Exposure (DX)*: records the details of a subject's exposure to a medical device under study.

Relationships

- *Device-Subject Relationships (DR)*: links each subject to devices to which they have been exposed.

Of the 3 components that make up a CGM device, only the receiver and transmitter can be uniquely and consistently identified via the serial number provided by the manufacturer. It is thus established that assigning a SPDEVID to each and every component of a CGM device used in *Study A* is not feasible. Given that 1) all data are stored in the receiver and 2) all

components function as a whole to collect CGM readings, it is evaluated that being able to identify uniquely and consistently each and every component of a CGM device does not add value to the study hence should not be pursued.

Instead, it would be sufficient to represent a CGM device by concatenating SUBJID and serial number of the CGM receiver in SPDEVID and reported in DI and DR domains (Illustration 1 and 2 respectively). In this way, the use of CGM devices by individual participants in *Study A* can be related to the CGM readings in LBC dataset and across device domains for traceability. For studies that include multiple digital health devices, for example both CGM and FCM, however, such construct may not be sufficient. A global naming convention would be beneficial to ensure standardisation and alignment across studies.

STUDYID	DOMAIN	SPDEVID	DISEQ	DIPARMCD	DIPARM	DIVAL
STUDY A	DI	103-TX3101722	1	DEVTYPE	Device Type	CGM Receiver
STUDY A	DI	103-TX3101722	2	MANUF	Manufacturer	TX Manu.
STUDY A	DI	103-TX3101722	3	MODEL	Model Number	TX
STUDY A	DI	103-TX3101722	4	SERIAL	Serial Number	TX3101722
STUDY A	DI	119-TX03300629	5	DEVTYPE	Device Type	CGM Receiver
STUDY A	DI	119-TX03300629	6	MANUF	Manufacturer	TX Manu.
STUDY A	DI	119-TX03300629	7	MODEL	Model Number	TX
STUDY A	DI	119-TX03300629	8	SERIAL	Serial Number	TX03300629

Illustration 1 DI Domain

STUDYID	DOMAIN	USUBJID	SPDEVID
STUDY A	DR	STUDY A/103	103-TX3101722
STUDY A	DR	STUDY A/119	103-TX3101722

Illustration 2 DR Domain

REPORTING DATA USEFULNESS

As discussed earlier, a minimum of 70% activity of the overall CGM periods is needed to achieve reliable analysis of glycaemic indices required. This makes CGM exposure, in terms of activity, of each participant an important piece of information in *Study A* that needs to be reported in SDTM to substantiate analysis decision. For example, exclusion of CGM readings of a particular participant due to insufficient CGM exposure, i.e. with activity throughout the 10 CGM periods below 70%, from analysis.

Continuous glucose monitoring in *Study A* is achieved by uninterrupted detection and report of glucose values every 5 minutes. This means that if a participant has 100% activity during a 14-day CGM period, there will be 4 032 CGM records at 5-minutes interval and a CGM exposure of 20 160 minutes. Following this principle, CGM exposure of each participant can be determined and reported in DX domain by identifying the disruptions, i.e. “holes” in 5-minute interval records, in CGM readings throughout the CGM periods (Illustration 3). With CGM exposure reported in DX domain, the use of CGM at participant, site and collectively at trial level throughout the study become transparent and informed decision on use of the data can be made accordingly.

STUDYID	DOMAIN	USUBJID	SPDEVID	DXSEQ	DXTRT	DXCAT	DXSTDTC	DXENDTC	DXSTDY	DXENDY	DXDUR
STUDY A	DX	STUDY A/103	103-TX3101722	1	GLUCOSE MEASUREMENT	CGM	2021-03-23T07:43:58	2021-03-30T12:13:54	1	8	PT10350M
STUDY A	DX	STUDY A/103	103-TX3101722	2	GLUCOSE MEASUREMENT	CGM	2021-03-30T12:34:01	2021-03-30T15:09:12	8	8	PT155M
STUDY A	DX	STUDY A/103	103-TX3101722	3	GLUCOSE MEASUREMENT	CGM	2021-03-30T17:39:12	2021-04-06T12:09:04	8	15	PT9750M
STUDY A	DX	STUDY A/103	103-TX3101722	4	GLUCOSE MEASUREMENT	CGM	2021-04-06T12:54:11	2021-04-13T08:09:13	15	22	PT9795M
STUDY A	DX	STUDY A/103	103-TX3101722	5	GLUCOSE MEASUREMENT	CGM	2021-04-13T08:44:20	2021-04-20T08:04:23	22	29	PT10040M
STUDY A	DX	STUDY A/103	103-TX3101722	6	GLUCOSE MEASUREMENT	CGM	2021-08-24T08:42:50	2021-09-01T22:37:58	155	163	PT12355M
STUDY A	DX	STUDY A/103	103-TX3101722	7	GLUCOSE MEASUREMENT	CGM	2021-09-02T05:07:50	2021-09-03T08:43:01	164	165	PT1655M
STUDY A	DX	STUDY A/103	103-TX3101722	8	GLUCOSE MEASUREMENT	CGM	2021-09-03T14:12:59	2021-09-03T23:43:00	165	165	PT570M
STUDY A	DX	STUDY A/103	103-TX3101722	9	GLUCOSE MEASUREMENT	CGM	2021-09-04T04:57:51	2021-09-06T01:53:14	166	168	PT2695M
STUDY A	DX	STUDY A/103	103-TX3101722	10	GLUCOSE MEASUREMENT	CGM	2021-09-06T04:28:14	2021-09-07T08:18:00	168	169	PT1670M

Illustration 3 DX Domain: One Row per Uninterrupted CGM Exposure Period per Subject

CONCLUSION

With the promising benefit of real-time, high-quality collection of health data, adoption of digital health devices in clinical studies is an unturnable tide that threatens to transform the way clinical data are processed and analysed. Given the diversity of digital health technologies and the varying use cases, there are no set rules as to how these voluminous data are to be fit into SDTM.

From the experience of *Study A*, NN learns that 1) end-to-end evaluation at early planning phase is the key to “survive” operational challenges and 2) dedicated effort to look into global implementation pays off in securing a viable solution. Lastly, even though it is not touched upon in this paper, challenges in the submission process are to be anticipated and planned accordingly.

REFERENCES

[1] http://agpreport.org/

[2] Ambulatory Glucose Profile (AGP) Report in Daily Care of Patients with Diabetes: Practical Tips and Recommendations *Leszek Czupryniak, Grzegorz Dzida, Piotr Fichna, Przemyslawa Jarosz-Chobot, Janusz Gumprecht, Tomasz Klupa, Malgorzata Mysliwiec, Agnieszka Szadkowska, Dorota Bomba-Opon, Krzysztof Czajkowski, Maciej T. Malecki, and Dorota A. Zozulinska-Ziolkiewicz*

CONTACT INFORMATION

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

Vicky Poulsen
Novo Nordisk A/S
Vandtårnsvej 114
DK-2860 Søborg
Denmark
Email:vicp@novonordisk.com