

# Mapping of Gastrointestinal Tolerance Data to SDTM

Lieke Gijsbers, OCS Life Sciences, 's-Hertogenbosch, the Netherlands

## ABSTRACT

In nutrition studies data regarding gastrointestinal tolerance is often collected. Gastrointestinal tolerance involves among others 1) stool characteristics, such as the frequency and consistency and 2) the occurrence and severity of gastrointestinal symptoms, such as diarrhoea, constipation, cramps and regurgitation. In CDISC SDTM there is no modelled domain available in which you can fit all these data and it is not described how data collected in a diary need to be processed. This paper explains using a case study how the gastrointestinal data are mapped to the CDISC SDTM domains Clinical Events and Findings About, what the challenges are and how these are countered.

## INTRODUCTION

Data standardisation is increasingly important in the field of (clinical) nutrition research. It enables aggregation of data and comparison of data across studies, which accelerates clinical nutrition research and the development of safe food products. Gastrointestinal tolerance is often evaluated in clinical nutrition studies as either a primary or secondary outcome. However, nutrition concepts such as gastrointestinal tolerance are not (completely) addressed by the currently existing CDISC standards and implementation guide, and although volunteers are working on developing the therapeutic area guide 'nutrition', it is not yet available to the public.

Gastrointestinal tolerance may include the assessment of stool characteristics in terms of frequency, consistency, colour and volume. Also, the occurrence, severity and frequency of gastrointestinal symptoms, such as diarrhoea, constipation, vomiting, cramps and flatulence, may be assessed. Gastrointestinal tolerance data can be collected at visits using a case report form (CRF), but are often recorded in a study diary by the subject or by the subject's parents/guardians to gather more detailed information.

Mapping gastrointestinal tolerance data to CDISC SDTM is challenging, because there is no single modelled domain available to represent these data. There is also a large variety in the CRFs of legacy studies and in the collected gastrointestinal (diary) data. Another challenge is that these data are often collected by means of diaries which are prone to reporting errors. Therefore, having information about the planned timing of the diaries may be useful. This paper describes how gastrointestinal tolerance data and the planned timing of the diaries can be mapped to SDTM.

## STOOL CHARACTERISTICS DATA

In Figure 1 an example CRF on stool characteristics and the mapping of its data is shown. We mapped the stool characteristics data to the SDTM domain Findings About (FA). This domain, which is a specialisation of the general observation class 'findings', should be used in the following cases:

1. Data or observations about an event or intervention for which no event or intervention record has been collected or created
2. Data or observations about an event or intervention which have qualifiers of their own that can be represented in findings variables (e.g., units, method)
3. Data or observations that have different timing from an associated event or intervention as a whole
4. Data or information about an event or intervention that indicate the occurrence of related symptoms or therapies
5. Data or information that indicate the occurrence of pre-specified adverse events

The stool characteristics data can be interpreted as data about the event 'having a stool' for which no event record is created, and therefore these data reflect case 1 in the abovementioned list, and hence should be mapped to the domain Findings About. In Figure 1 it is shown - using annotations and colour coding - how the stool characteristics data collected on the CRF are mapped to SDTM variables. In general, the following SDTM variables are used to map the following data:

- FATESTCD: short name of the stool characteristic assessment
- FATEST: name of the stool characteristic assessment
- FAOBJ: the object of the assessment, in this case "Stool"
- FAORRES: result of the stool characteristic assessment as originally collected
- FAORRESU: original unit in which the data were collected, i.e. unit for FAORRES
- FAEVINTX: the evaluation interval associated with the assessment

# PhUSE 2017

A CRF on stool characteristics data with annotations

## Visit 1

FATESTCD = "A\_COUNT"

FAORRESU = "/day"

What was the average number of stools per day last week?

3

FAOBJ = "Stool"

FAEVINTX = "LAST WEEK"

FAORRES

What was the average consistency of the stools last week?

Watery

FATESTCD = "A\_CONSIS"

Soft

Formed

Hard

Mapping of data of this CRF into domain Findings About

DOMAIN	USUBJID	FATESTCD	FATEST	FAOBJ	FAORRES	FAORRESU	FAEVINTX
FA	321-001	A_COUNT	Average Count	Stool	3	/day	LAST WEEK
FA	321-001	A_CONSIS	Average Consistency	Stool	Soft		LAST WEEK

Figure 1. Representation of the stool characteristics data recorded on the CRF in domain Findings About

Stool characteristics data collected by means of a diary is often more detailed than data collected by means of a CRF, for example the consistency per stool is assessed instead of the average stool consistency. In Figure 2 an example of a diary is shown and it is displayed how these data would be represented in SDTM domain Findings About. In addition to the variables used to store the CRF data, the following SDTM variable is used:

- FASPID: the number of the stool

Although it is not explicitly collected how many stools the subject had, this information is implicitly available by counting the number of stools for which the consistency and colour has been recorded. Therefore, a record with the number of stools (i.e. FATESTCD = "COUNT") is included in domain Findings About.

A diary on stool characteristics data with annotations

## Diary week 1 - day 1

FAEVINTX = "TODAY"

No stool today

FATESTCD = "CONSIS"

FATESTCD = "COLOUR"

FASPID	Stool consistency				Stool colour		
No.	Watery	Soft	Formed	Hard	Brown	Green	Yellow
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FATESTCD = "COUNT"

FAOBJ = "Stool"

Mapping of data of this diary into domain Findings About

DOMAIN	USUBJID	FASPID	FATESTCD	FATEST	FAOBJ	FAORRES	FAORRESU	FAEVINTX
FA	321-001	1	CONSIS	Consistency	Stool	Watery		TODAY
FA	321-001	1	COLOUR	Colour	Stool	Brown		TODAY
FA	321-001	2	CONSIS	Consistency	Stool	Soft		TODAY
FA	321-001	2	COLOUR	Colour	Stool	Green		TODAY
FA	321-001		COUNT	Count	Stool	2	/day	TODAY

Figure 2. Representation of the stool characteristics data recorded on the diary in domain Findings About

## PhUSE 2017

### GASTROINTESTINAL SYMPTOMS DATA

In Figure 3 an example CRF on gastrointestinal symptoms and the mapping of its data to SDTM domain Clinical Events is shown. Gastrointestinal symptoms are events of interest that are not classified as adverse events (i.e. events that do not constitute adverse events in themselves) and should therefore be represented in the SDTM domain Clinical Events (CE). In general, the following SDTM variables are used to map the following data:

- CETERM: name of the gastrointestinal symptom
- CEOCCUR: occurrence of the gastrointestinal symptom
- CESEV: severity or intensity of the gastrointestinal symptom
- CEPRESP: to indicate that the gastrointestinal symptom was pre-specified
- CEEVINTX: evaluation interval associated with the observation of the gastrointestinal symptom

A CRF on gastrointestinal symptoms data with annotations

**Visit 1** CEEVINTX = "LAST WEEK"

In general, did the subject suffer from the following symptoms last week:

CETERM	Absent	Mild	Moderate	Severe
Vomiting	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Flatulence	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diarrhoea	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Constipation	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cramps	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

CEPRES = "Y"      CEOCCUR      CESEV

Mapping of data of this diary into domain Findings About

DOMAIN	USUBJID	CETERM	CEPRES	CEOCCUR	CESEV	CEEVINTX
CE	321-001	VOMITING	Y	N		LAST WEEK
CE	321-001	FLATULENCE	Y	Y	MILD	LAST WEEK
CE	321-001	DIARRHOEA	Y	Y	MODERATE	LAST WEEK
CE	321-001	CONSTIPATION	Y	Y	MILD	LAST WEEK
CE	321-001	CRAMPS	Y	Y	SEVERE	LAST WEEK

**Figure 3.** Representation of the gastrointestinal symptoms data recorded on the CRF in domain Clinical Events

Besides the occurrence and severity, also the frequency of gastrointestinal symptoms may be assessed (see Figure 4 for an example diary). The occurrence of the gastrointestinal symptom is, as explained above, stored in the domain Clinical Events. The frequency data is, however, stored in the domain Findings About, see Figure 4. We did consider mapping of the frequency data to the supplemental qualifier Clinical Events (SUPPCE) domain, but based on the cases when SDTM domain Findings About should be used (see section "Stool Characteristics"), we decided to store the data in SDTM domain Findings About. The frequency data can be seen as data about the event 'gastrointestinal symptom' and since frequency data does have a unit it meets case 2 of when to use domain Finding About: i.e. data or observations about an event or intervention which have qualifiers of their own that can be represented in findings variables (e.g., units, method), hence it should be represented in the domain Findings About. Moreover, regulatory authorities, like the Food & Drug Administration (FDA), prefer the usage of domain Findings About above supplemental qualifiers.

To store the frequency data the same variables are used as for the storage of stool characteristics data. However, in this case FAOBJ is populated with the name of the gastrointestinal symptom.

Since gastrointestinal tolerance data is now divided over two domains (Findings About with frequency data and Clinical Events with occurrence and severity data), identifier variables FALNKID and CELNKID are introduced to the domains Findings About and Clinical Events respectively in order to link the related data (see Figure 4). An identical value for FALNKID and CELNKID within a USUBJID indicates that the records are related. Also, a Related Records (RELREC) domain is created to store the relationship between the domains Clinical Events and Findings About (see Figure 4) with the following variables:

- RDOMAIN: codes of the related domains

## PhUSE 2017

- USUBJID: left blank, because whole datasets are linked instead of specific subject records
- IDVAR: name of the variable that links the two datasets and identifies the related records, in this case CELNKID and FALNKID
- IDVARVAL: left blank, because whole datasets are linked instead of records with a specific value for CELNKID and FALNKID
- RELTYPE: type of relationship between the datasets, in this case *one* record in domain Clinical Events is linked to *one* record in domain Findings About
- RELID: identifier of the relationship, can be chosen by the sponsor

A diary on gastrointestinal symptoms data with annotations

### Diary week 1 - day 1

CEEVINTX/FAEVINTX = "TODAY"

Did regurgitation or vomiting occur today?

CEPRESP = "Y"

FATESTCD = "COUNT"

Vomiting  yes How many times today?  2

Regurgitation  yes How many times today?  1

CETERM/FAOBJ

CEOCCUR

FAORRESU = "/day"

FAORRES

Mapping of the occurrence data into domain Clinical Events

DOMAIN	USUBJID	CELNKID	CETERM	CEPRESP	CEOCCUR	CEEVINTX
CE	321-001	CEFA1	VOMITING	Y	Y	TODAY
CE	321-001	CEFA2	REGURGITATION	Y	Y	TODAY

Mapping of the frequency data into domain Findings About

DOMAIN	USUBJID	FALNKID	FATESTCD	FATEST	FAOBJ	FAORRES	FAORRESU	FAEVINTX
FA	321-001	CEFA1	COUNT	Count	Vomiting	2	/day	TODAY
FA	321-001	CEFA2	COUNT	Count	Regurgitation	1	/day	TODAY

Representation of domain Related Records to indicate that domains Clinical Events and Findings About are related

RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
CE		CELNKID		ONE	GICOUNT
FA		FALNKID		ONE	GICOUNT

**Figure 4.** Representation of the gastrointestinal symptoms data recorded on the diary in domain Clinical Events, Findings About and Related Records

### TIMING OF DIARIES

In Figure 5 it is displayed how the (planned) timing of a diary can be mapped to SDTM variables. The date of the diary is mapped to CEDTC, which represents the date/time of event collection. The SDTM timing variables VISIT and VISITNUM are not populated, because diary data are not collected during a clinical encounter.

Diaries filled out by subjects or subject's parents/guardians are prone to reporting errors. Because of this, dates may be incomplete or incorrect. Most diaries contain fortunately information about when the diary should be filled out, i.e. the planned timing of the diary. Mapping of these planned timings is useful, because it gives some indication of the timing of the diary. The diary displayed in Figure 5 had to be filled out on day 1 of week 1. We mapped the planned time point using the following variables:

- CETPT: description of when the assessment should be taken, i.e. planned time point name
- CETPTNUM: numeric version of CETPT
- CETPTREF: description of the fixed reference pint referred to by CETPT and CETPTNUM

Note: in the example diary in Figure 5 a diary on gastrointestinal symptoms is displayed. However, these (planned) time points may also be used for stool characteristics data and accordingly the variables FADTC, FATPT, FATPTNUM and FATPTREF should be used.

## PhUSE 2017

A diary with (planned) timing information and annotations  
 Instruction to fill out the diary *daily* for 10 weeks

**CEDTC**

**Diary week 1 - day 1**      Date (dd/mm/yyyy): 08/10/2017

**CETPTREF**      **CETPT**

Did regurgitation or vomiting occur today?

Vomiting       yes      How many times today?     

Regurgitation       yes      How many times today?     

Mapping of the (planned) timing of diaries

USUBJID	VISITNUM	VISIT	CEDTC	CETPT	CETPTNUM	CETPTREF
321-001			2017-10-08	DAY 1	1	WEEK 1

**Figure 5.** Representation of (planned) timing of diaries

### CONCLUSION

In this paper, we have shown how gastrointestinal tolerance data can be mapped to CDISC SDTM. The occurrence and severity of gastrointestinal symptoms data fits within the domain Clinical Events, whereas the frequency of gastrointestinal symptoms fits within the domain Findings About. The domain Findings About is also suitable to store stool characteristics data, such as the stool frequency, consistency, colour and volume. We recommend the mapping of the planned time point of diaries, because diary dates are often incomplete or incorrect. The planned time points will give at least some indication of the timing of the diaries. Until the therapeutic area guide 'nutrition' is being published the examples shown in this paper may provide guidance on how to represent gastrointestinal tolerance data according to CDISC SDTM.

### CONTACT INFORMATION

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

Author Name      Lieke Gijsbers  
 Company          OCS Life Sciences  
 Address          Ruwekampweg 2G  
 City / Postcode    's-Hertogenbosch / 5222 AT  
 Work Phone:      0031 (0)73 523 6000  
 Email:            lieke.gijsbers@ocs-consulting.com  
 Web:              www.ocs-lifesciences.com