PhUSE EU Connect 2018

Paper SI13

Easing Your Pain with Biomedical Concepts

Kirsten Walther Langendorf, S-CUBED & A3 Informatics, Copenhagen, Denmark

ABSTRACT

Does your organisation spend significant resources on:

- Study group meetings to clarify what data to collect?
- Specifying study setup: eCRF/lab specs/CRF SDTM annotation?
- Creating SDTM data?
- Managing standard CRFs?

If so, then using Biomedical Concepts (BCs) are a solution. A BC is a definition of an 'observation', something to measure in a study. The presentation will demonstrate how BCs allow:

- Unambiguous communication between the clinical 'world' and the technical 'world';
- Greater flexibility in CRF design while ensuring standardisation to CDISC;
- A more seamless process where mapping is no longer required; silos are removed, and we gain automated traceability

The presentation will include examples of BCs in action (using the CDISC pilot study) and how such constructs can be used within studies today implementing in an iterative manner while still supporting current study work.

INTRODUCTION

In a pharmaceutical company the 'study setup' is typically the step that translates study specifications in the protocol to actual implementations of data collection (e.g. EDC, Lab) and reporting (e.g. SDTM datasets). Questions like:

- Which units to use – do we allow for multiple units or just one? Which one are applicable for the test?
- Which tests are being measured in BLOOD, SERUM, or PLASMA
- Is the blood pressure to be measured STANDING, SITTING or SUPINE – what should be on the CRF? Do we need a new CRF?
- Which SDTM domains are needed for this study – do we need a new or where does this information belong in SDTM domains? Do we need a new annotation?

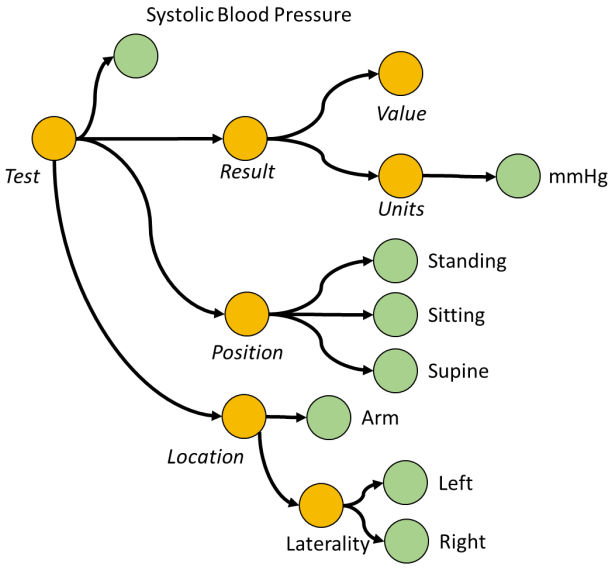
The data collection and reporting are handled in electronic systems (EDC systems, Lab loading systems, SAS programs, ePROs) and therefore needs to be specific and detailed. Many details beyond what is described in protocol need to be decided upon for the systems to work properly.

Often standards are in place to allow for re-use but typically the standards are not connected, i.e. units and specimen are kept in separate unit and specimen code lists, not linked to the list of --TESTCDs. E.g. the list of possible units associated to a TESTCD is typically something done when creating Value Level Metadata in the define.xml. The domains are kept in a list of standard domains but there is no linkage to which --TESTCDs goes into which domain, this is specified in the standard annotation which is typically in pdf.

Despite the standards, quite a bit of effort is used in every study specifying what to collect and how.
The definition of a Biomedical Concept (BC) ties these standard 'silos' together and facilitates a consistent way of collecting the data across studies.

WHAT IS A BIOMEDICAL CONCEPT?

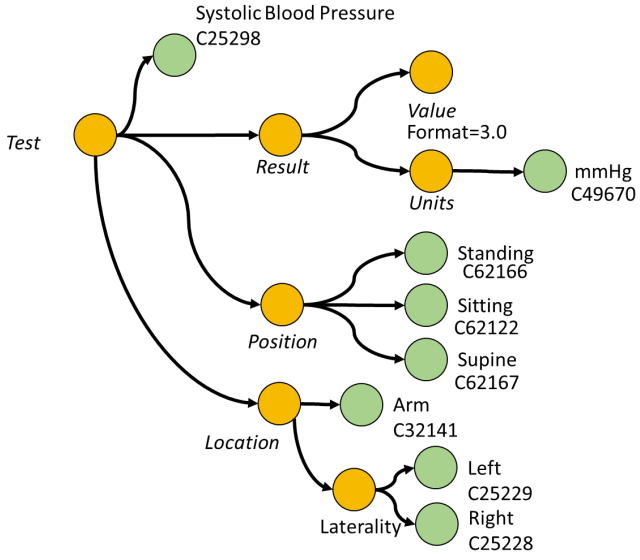
A biomedical concept (BC) is a definition of something measured on a subject, e.g. systolic blood pressure:



For those working with SDTM, think of a BC as a topic that would generate a record in SDTM. The SDTM IG refers to a 'Topic variables, which specify the focus of the observation', see ref. [1] (Section 2.1). The BC is more detailed defined than a single 'Topic variable' code. In the example above, think of the yellow nodes as variables and the green nodes as (possible) values.

The definition of the systolic blood pressure above may seem like a very detailed specification of the systolic blood pressure which is only used in phase 1 studies. For a phase 3 study this level of detail may not be needed. However, it provides a standardised view of what is needed for collecting a systolic blood pressure and all its varying factors.

The next step in the standardisation to CDISC would be to refer to the actual c-codes used. This is typically what a data manager or a SDTM programmer will need creating SDTM:



Now it specifies which of the terms and units from the different CDISC code lists are applicable for this test, and the expected format for the result of the systolic blood pressure. Below is the BC definition in an MDR

Show: Systolic Blood Pressure (BC C25298) BC C25298 (V3.1.0, 4, Incomplete)

Details

Identifer	Label	Owner	Status	Version	Version Label	
BC C25298	Systolic Blood Pressure (BC C25298)	ACME	Incomplete	3.1.0	0.3	Impact

Items

Show 10 entries Search:

Alias	Question Text	Prompt Text	Enabled	Collect	Datatype	Format	Terminology
Baseline (--BLFL)	Is this a baseline value?	Baseline	✔	✔	boolean		N (C49487) Y (C49488)
Body Position (--POS)	In what position was the subject during the measurement?	Position	✔	✔	string		SITTING (C62122) STANDING (C62166) SUPINE (C62167)
Date and Time (--DTC)	On what date were the measurement performed?	Date	✔	✔	dateTime		
Laterality (--LAT)	Which arm?	Arm	✔	✔	string		RIGHT (C25228) LEFT (C25229)
Location (--LOC)	The location of the measurement	Location	✔	✖	string		ARM (C32141)
Method Code (--METHOD)			✖	✖	string		
Result Units (--ORRESU)	The unit of the measurement	Units	✔	✔	string		mmHg (C49670)
Result Value (--ORRES)	What was the result of the measurement?	Result	✔	✔	float	3	
Test Code (--TESTCD)			✔	✖	string		SYSBP (C25298)
Test Name (--TEST)			✔	✖	string		Systolic Blood Pressure (C25298)

Finally, allowing the BC to be assigned to a domain will link the SDTM to the collected data (the BC):

Show: Vital Signs VS Domain (V1.0.0, 2, Standard)

Details

Identifer	Label	Owner	Internal Version	Version
VS Domain	Vital Signs	ACME	1.0.0	

Notes

Used Variables

Show 10 entries Search:

Ordinal	Name	Label	Datatype	CT	Format	Role	Sub Role	IG Notes	Notes	Core
1	STUDYID	Study Identifier	Char			Identifier		Unique identifier for a study.		Required
2	DOMAIN	Domain Abbreviation	Char	DOMAIN		Identifier		Two-character abbreviation for the domain.		Required
3	USUBJID	Unique Subject Identifier	Char			Identifier		Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.		Required
4	VSSEQ	Sequence Number	Num			Identifier		Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.		Required
5	VSGRPID	Group ID	Char			Identifier		Used to tie together a block of related records in a single domain for a subject.		Permissible
6	VSSPID	Sponsor-Defined Identifier	Char			Identifier		Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.		Permissible
7	VSTESTCD	Vital Signs Test Short Name	Char	VSTESTCD		Topic		Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g."1TEST"). VSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: SYSBP, DIABP, BMI.		Required
8	VSTEST	Vital Signs Test Name	Char	VSTEST		Qualifier	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in VSTEST cannot be longer than 40 characters. Examples: Systolic Blood Pressure, Diastolic Blood Pressure, Body Mass Index.		Required
9	VSCAT	Category for Vital Signs	Char			Qualifier	Grouping Qualifier	Used to define a category of related records.		Permissible
10	VSSCAT	Subcategory for Vital Signs	Char			Qualifier	Grouping Qualifier	A further categorization of a measurement or examination.		Permissible

Showing 1 to 10 of 32 entries Previous 1 2 3 4 Next

Unused Variables

Show 10 entries Search:

Ordinal	Name	Label	Datatype	CT	Format	Role	Sub Role	IG Notes	Notes	Core
No data available in table										

Showing 0 to 0 of 0 entries Previous Next

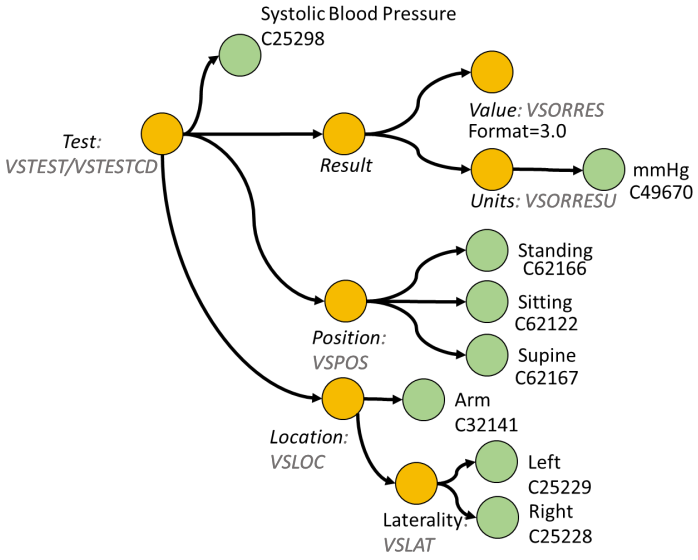
BCs added to the VS domain

Biomedical Concepts

Show 10 entries Search: pressure

Identifer	Label	Owner	Version	Version Label
BC C25298	Systolic Blood Pressure (BC C25298)	ACME	2.0.0	0.2
BC C25298	Systolic Blood Pressure (BC C25298)	ACME	3.0.0	0.3
BC C25299	Diastolic Blood Pressure (BC C25299)	ACME	2.0.0	0.2

By making the relationship of the SDTM domain to the BC then defines where in SDTM this collected information belongs:

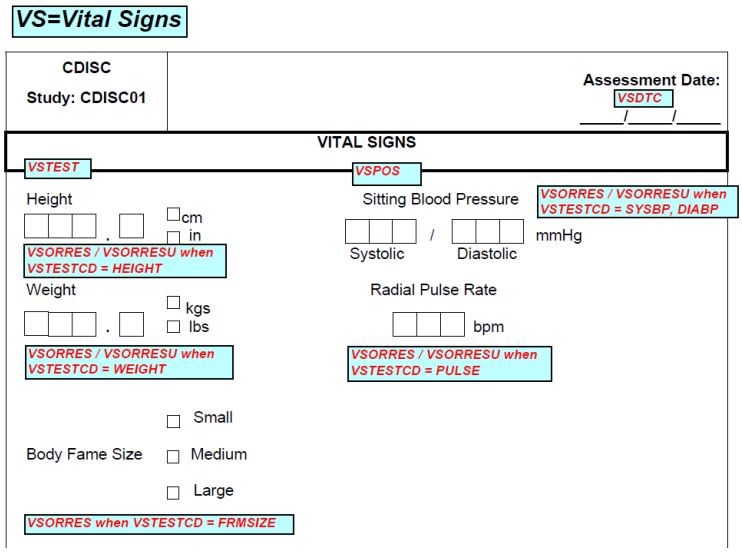


Note that multiple domains can refer to the same BC. The context of the study will then determine which domain is used.

Considering the amount of data collected across studies, there will be quite many BC definitions. So why bother doing all that definitions work and managing yet another 'standard'? Answer: a BC is a standardised definition that can be used for various purpose and facilitates faster and more flexible study setup. This is described in the following sections.

STANDARDISATION - FROM FORMS TO BCs

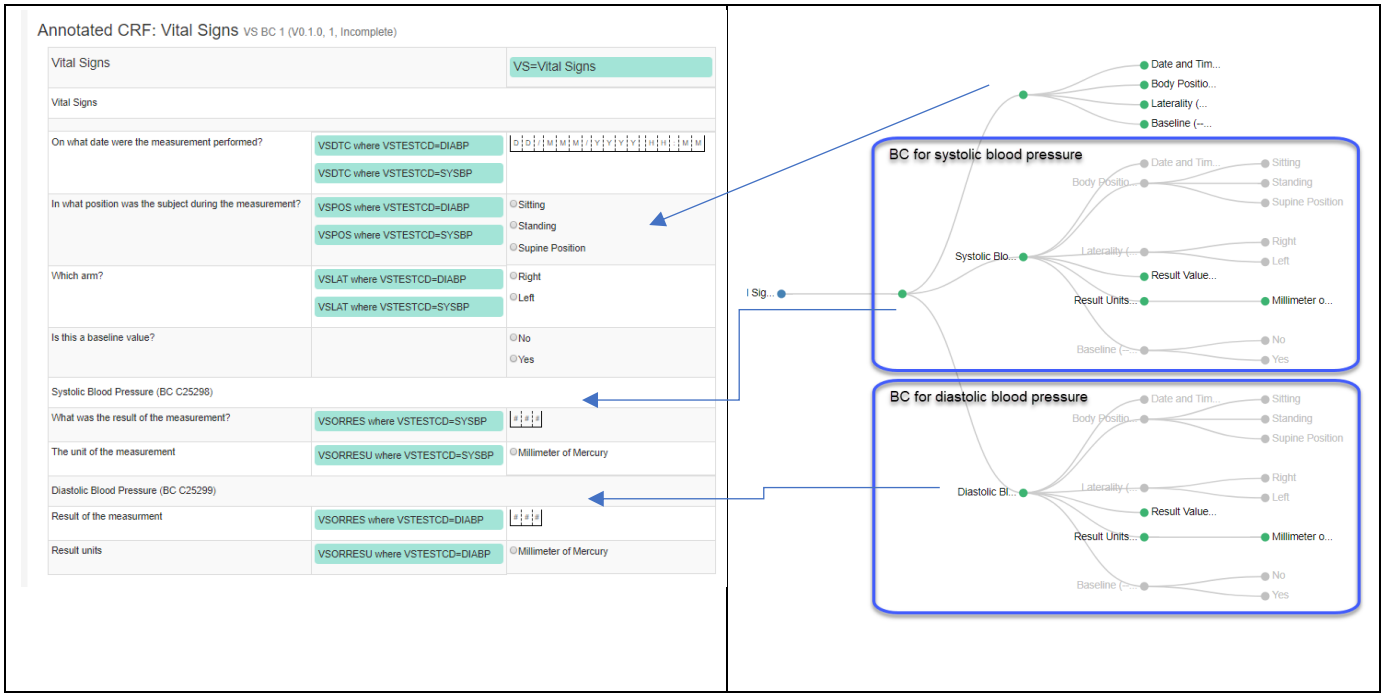
Traditionally, a form is built by a question and an answer type of setup on a 'page' (either paper or EDC):



Depending on what level of detail is needed different forms ('pages') are being developed and maintained. SDTM annotations per page/form is developed too and most likely also other mappings to internal databases. The fact that these questions (Height, Weight, Blood Pressure etc.) are put in the same box on a page with a headline VITAL SIGN determines that the data will end up in the VS domain. To ensure that CDISC standards are adhered to, a

governance structure is typically adopted. Making a new form for a study will require request/review/acceptance by a standards governance team. I will use the term 'paper world' for this approach in the following discussion.

Developing forms using BCs will allow for a more flexible approach while still ensuring adherence to CDISC standards. The following examples are from a graph-based implementation, but similar approach could be made using other tools. In the example below, a Vital Signs form has been created using BC for systolic and diastolic blood pressure:



Elements in the BC that are common across the BCs can be grouped together to avoid redundancy. Creating a form like this only takes a few minutes since it is only a selection of the BCs needed on the form. Also, both blank CRF, annotated CRF and export file to EDC can be made automatically by press of a button.

The flexibility is achieved on 3 levels:

1. The user can decide which BCs to add to a Vital Signs form, i.e. which questions are included in the vital sign page
2. The user can enable/disable data elements in a BC, e.g. decide not to use the baseline flag question if not relevant, i.e. what the level of response is needed for the question
3. The user can decide if questions and/or responses are optional/mandatory

The user defining the form does not have to consider the terminology used or where the questions are to be mapped to SDTM. That is all defined in BC and by linking the domain to the BC. Recall that the SDTM relationship to the BC also defines the --TEST/--TESTCD and other SDTM default terms (e.g. LOCATION=ARM in the example above or could be --CAT/--SCAT, if relevant for the BC). In this example the VS domain link to the BC so the --TEST is VSTEST. Assigning the BC to another domain will automatically change the annotation.

Building forms using BCs will allow for a more flexible and distributed approach compared to the typical 'paper world' approach. Creating forms does not have to be governed so tightly – but the BCs does. However, having a library of forms will still help in making study setup more effective (re-use), but it can be decentralised e.g. to the clinical project or on Therapeutic Area (TA) level.

UPSTREAM STANDARDISATION – BCs USED IN PROTOCOLS

The use of the BC can be taken one step further up in the data stream supporting the standardisation at protocol level.

The protocol typically specifies in an overview table (protocol flowchart or Schedule of Assessments (SoA)) the planned tests and the timing hereof (visits):

Assessments	V1	V2	V3	V4
Demographics	x			
Vital signs	x			x
CT Scan		x		
Dosing		x	x	x
EQ-5D	x			x
Lipids	x			x
Haemoglobin	x			x
Adverse Events	x	x	x	x

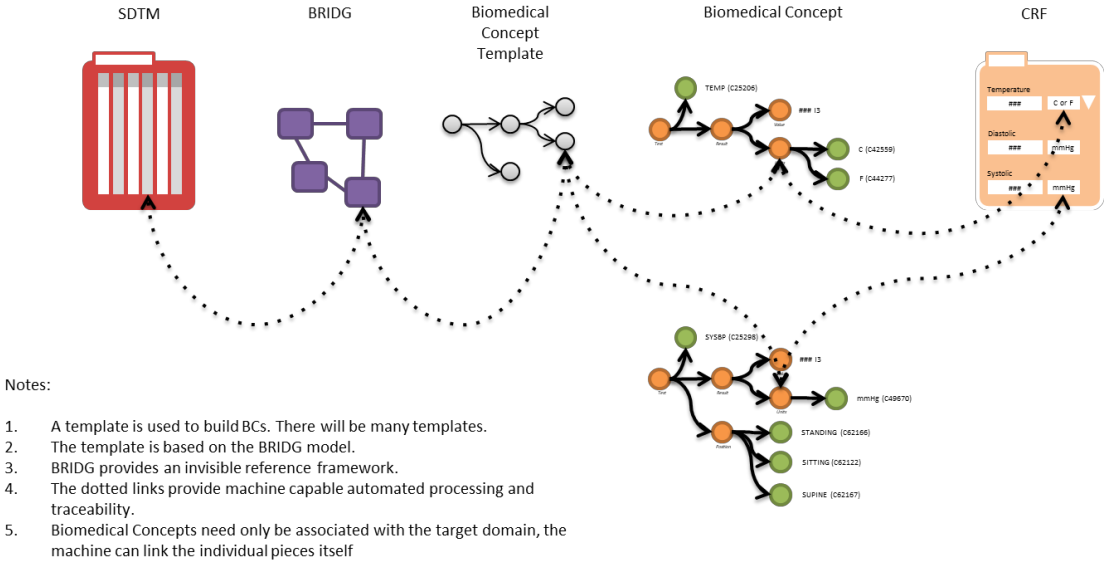
This overview could be made electronically by having a system/application that would allow the study group to select from a list of pre-defined test, i.e. the BCs.

Assessments	BC	V1	V2	V3	V4
Demographics	Age Race Date of Birth Sex	x			
Vital signs	Height Weight Heart Rate Temperature Systolic Blood Pressure Diastolic Blood Pressure	x			x
CT Scan	CT scan timing		x		
Dosing	Number of dosing units		x	x	x
EQ-5D	Mobility	x			x
Lipids	LDL HDL Total Cholesterol	x			x
Haemoglobin	Haemoglobin	x			x
Adverse Events	Adverse Events	x	x	x	x

Each BC having a detailed definition like the systolic blood pressure on page 2. For the study CRF setup, the defined forms containing these BCs will then be listed and the CRF designer would then pick the appropriate CRF (or make a new one). The CRF and aCRF could then be made by a push of a button and reviewed early by the study group allowing for adjustments to the protocol if needed.

SHARING BCs

The BCs shown in the graph-based example above are based on the BRIDG definitions, see ref. [2]. Each BC is based on a BC template that links to the BRIDG model which has links/annotation to the SDTM model.



Notes:

1. A template is used to build BCs. There will be many templates.
2. The template is based on the BRIDG model.
3. BRIDG provides an invisible reference framework.
4. The dotted links provide machine capable automated processing and traceability.
5. Biomedical Concepts need only be associated with the target domain, the machine can link the individual pieces itself

Different types of BC templates have been defined to allow for definition of different types of BCs, e.g. BCs with numeric response and BCs with categoric (code list) response.

Graph-based approach or not, if the BCs are defined using structured templates, then BC definitions (in multiple format such as MS Excel or RDF format) can be shared in the industry. Contributors could be both pharmaceutical companies, CROs as well as regulators. Sharing the definition of a BC does not disclose any intellectual property or any sensitive data that needs de-identification.

BCs FOR DEFINING THERAPEUTIC AREAS (TA)

Having a common way of defining BCs could also support a more precise definition of the Therapeutic Areas. The TAs could be defined as a set of BCs, as illustrated on the use of BCs for protocol definition

Assessments	BC	Screening	Baseline	Treatment	EndOfTrial
Demographics	Age Race Date of Birth Sex	x			
Vital signs	Height Weight Heart Rate Temperature Systolic Blood Pressure Diastolic Blood Pressure	x			x
CT Scan	CT scan timing		x		
Dosing	Number of dosing units		x	x	x
EQ-5D	Mobility	x			x
Lipids	LDL HDL Total Cholesterol	x			x
Haemoglobin	Haemoglobin	x			x
Adverse	Adverse Events	x	x	x	x

Assessments	BC	Screening	Baseline	Treatment	EndOfTrial
Events					

Each of the BCs would then have the detailed description:

Show: Systolic Blood Pressure (BC C25298) BC C25298 (V3.1.0, 4, Incomplete)

Details

Identifer	Label	Owner	Status	Version	Version Label	
BC C25298	Systolic Blood Pressure (BC C25298)	ACME	Incomplete	3.1.0	0.3	Impact

Items

Show 10 entries Search:

Alias	Question Text	Prompt Text	Enabled	Collect	Datatype	Format	Terminology
Baseline (--BLFL)	Is this a baseline value?	Baseline	✔	✔	boolean		N (C49487) Y (C49488)
Body Position (--POS)	In what position was the subject during the measurement?	Position	✔	✔	string		SITTING (C62122) STANDING (C62166) SUPINE (C62167)
Date and Time (--DTC)	On what date were the measurement performed?	Date	✔	✔	dateTime		
Laterality (--LAT)	Which arm?	Arm	✔	✔	string		RIGHT (C25228) LEFT (C25229)
Location (--LOC)	The location of the measurement	Location	✔	✖	string		ARM (C32141)
Method Code (--METHOD)			✖	✖	string		
Result Units (--ORRESU)	The unit of the measurement	Units	✔	✔	string		mmHg (C49670)
Result Value (--ORRES)	What was the result of the measurement?	Result	✔	✔	float	3	
Test Code (--TESTCD)			✔	✖	string		SYSBP (C25298)
Test Name (--TEST)			✔	✖	string		Systolic Blood Pressure (C25298)

and should allow for the same options (mandatory/options, enable/disable) as described on page 5 when creating forms.

Currently CDISC TAs come with a CDASH definition (here VS example from the CV TAUG – cdash-ta-cv-taug.xls, see ref. [1]):

Domain	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	Codelist	CRF Completion Instructions	Information for Sponsors	Core	Data Type
VS	Were vital signs collected ?	Vital signs collected ?	VSPERF		General prompt question regarding whether or not any VS were collected during the study. This provides verification that all other fields on the CRF were deliberately left blank. {NY} (See Section 2.2.)	NY	'Indicate if the vital signs were collected. If yes, include the appropriate details where indicated on the CRF.'	'The intent/purpose of collecting this field is to help with data cleaning and monitoring. See Best Practice Section 3.4, FAQ #6. For the SDTM-based dataset, SDTMIG variable VSSTAT is derived from a No value in VSPERF. This field does not map directly to an SDTM variable.'	Optional	Char
VS	On what date were the measurements performed?	Date	VSDAT	PerformedActivity.PerformedObservation.dateRange	Date of measurements.		Record date of measurements using this format (DD-MON-YYYY).	'The date of measurement can be derived from a collected date of visit and in such cases a separate measurement date field is not required. For the SDTM-based dataset, the SDTM IG variable VSDTC is derived by concatenating CDASH Date (VSDAT) and Time (VSTIM) of Vital Sign Measurements (if time is collected) into VSDTC using the ISO 8601 format. For more detail see the Best Practice section. This field does not map directly to an SDTM variable. *See the BRIDG model for complete path.'	Recommended/ Conditional	Date (dd-MON-yyyy)
VS	At what time were the measurements performed?	Time	VSTIM	PerformedActivity.PerformedObservation.dateRange	Time of measurements.		Record time of measurement (as complete as possible).	'For the SDTM-based dataset, the SDTM IG variable VSDTC is derived by concatenating CDASH Date (VSDT) and Time (VSTIM) of Vital Sign Measurements (if time is collected) into VSDTC using the ISO 8601 format. For more detail see the Best Practice section. This field does not map directly to an SDTM variable. *See the BRIDG model for complete path.'	Recommended/ Conditional	Time (24 hour)
VS	What is the Sponsor-Defined Identifier ?	<line number> or <VS number>	VSSPID	not needed	A sponsor-defined identifier that can be used for pre-printed numbers on the CRF.		Not applicable.	It can be beneficial to use an identifier in a data query to communicate clearly to the site the specific record in question.	Optional	Char

Domain	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	Codelist	CRF Completion Instructions	Information for Sponsors	Core	Data Type
VS	What is the planned time point for this measurement?	<Planned Time Point>	VSTPT	PerformedActivity.PerformedObservation > PlannedActivity > PlannedCompositionRelationship > PlannedActivity.PlannedSubjectActivityGroup.name WHERE PlannedActivity.PlannedSubjectActivityGroup > StudyActivity > DefinedActivity.DefinedSubjectActivityGroup.categoryCode = 'Study Subject Encounter' and subcategoryCode = 'Timepoint'	Text description of time when measurement should be taken.		Not applicable.	'If applicable, this will be pre-printed on CRF when measurements are required at multiple time points within a visit day. If the form is laid out as a grid, then words such as Planned Time Point can be included as the column header.'	Recommended/ Conditional	Char
VS	What is the vital sign test name?	<Test>	VSTEST	PerformedActivity.PerformedObservation > PlannedActivity > StudyActivity > DefinedActivity.DefinedObservation.nameCode.CD.originalText	Verbatim name of the test or examination used to obtain the measurement or finding. {VSTEST} (See Section 2.2.)	VSTEST	Record the name of the vital sign test if not pre-printed on the CRF.	'It is recommended that the test names be pre-printed on the CRF. If the form is laid out as a grid, then words such as Test can be included as the column header. *See the BRIDG model for complete path.'	Highly Recommended	Char
VS	What was the result of the measurement?	Result	VSORRES	PerformedObservation > PerformedObservationResult.value WHERE value.ANY => originalText or translation in original units for whatever data type the concept is captured as, IF such originalText or translation attribute exists; if no originalText or translation attribute exists (e.g. BL or ST) for that data type, then there is no mapping	Result of the vital signs measurement as originally received or collected.		Record vital sign results.	None. *See the BRIDG model for complete path.	Highly Recommended	Char
VS	What was the unit of the measurement?	<Units>	VSORRESU	PerformedObservation > PerformedObservationResult.value.ANY =>PQ.translation.PQR.code	Original units in which the data were collected. {VSRESU} (See Section 2.2.)	VSRESU	'Record or select the unit of measure associated with the test, if not pre-printed on the CRF.'	It is recommended that the units be pre-printed on the CRF when possible. *See the BRIDG model for complete path.	Recommended/ Conditional	Char

Domain	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	Codelist	CRF Completion Instructions	Information for Sponsors	Core	Data Type
VS	Was the result clinically significant?	Clinically Significant	VSCLSIG	PerformedObservation (vital sign) > PerformedObservationResult.> AssessedResultRelationship > PerformedObservation (significance evaluation) > PerformedObservationResult.PerformedClinicalInterpretation.clinicallySignificantIndicator	Whether vital sign result was clinically significant.		Record whether vital sign result was clinically significant.	'If this level of information is needed, it may be added to the CRF. This field does not map directly to an SDTM variable.'	Optional	Char
VS	At what location was the measurement taken?	Location	VSLOC	not needed	Location relevant to the collection of Vital Signs measurement. Example: LEFT ARM for blood pressure. {LOC) (See Section 2.2)	LOC	'Record or select location on body where measurement was performed, if not pre-printed on CRF.'	Location may be pre-defined as part of CRF label.	Optional	Char
VS	In what position was the subject during the measurement?	Position	VSPOS	not needed	Position of the subject during a measurement or examination. {POSITION} (See Section 2.2)	POSITION	Record the position of subject at time of test (e.g. SITTING).	'Results may be affected by whether conditions for vital signs as specified in the protocol were properly met. One common condition is the subject's position (e.g., SUPINE, STANDING). If the protocol requires this type of information, then this question may be included to confirm that the subject's position matches the protocol. The following are examples of when it is not necessary to collect these data on the CRF: Position of the subject is not pertinent to the protocol, or The protocol specifies only one possible position and the sponsor does not feel there is significant risk of the sites measuring the vital signs with the subject in the wrong position *See the BRIDG model for complete path.'	Recommended/ Conditional	Char

This is providing some of the BC information and the relationship to BRIDG model, but it does not provide the specific detail on actual --TEST/--TESTCDs used in VS and it does not specify the actual allowed/expected units and response codes. Also, these detailed CDASH definitions are not provided for all SDTM domains, see ref. [4]. The above is more like a BC template.

Representing the TA by using BCs would provide more specific definition and consistent usage to the consumer of the TAs.

CONCLUSION

The definition of Biomedical Concepts (BCs) provides a detailed metadata definition of something measured and registered on a subject in a study as exemplified with the definition of what a systolic blood pressure consists of.

A BC is defined once and can be reused for various purposes:

- Protocol specification – facilitating early data flow decisions.
- CRF build – facilitating flexible and distributed CRF build while still ensuring adherence to SDTM standards.
- Definition of Therapeutic Areas (TAs) – facilitating more precise guidance on the content to the users.

To ensure consistent definition of BCs a set of templates can be defined. BCs should also be shared which would expand the reuse beyond a single company and hereby minimise the variance in SDTM implementations.

REFERENCES

[1] Study Data Tabulation Model Implementation Guide: Human Clinical Trials, Version 3.2
[2] Iberson-Hurst, Dave (PhUSE 2015). *TT09 - CDISC Standards and the Semantic Web*
[3] Cardiovascular TAUG Metadata: https://www.cdisc.org/system/files/gold/eshare/taug-cv-v1.zip
[4] CDASH Model v1.0: https://www.cdisc.org/standards/foundational/cdash/cdash-model-10#Bookmark7

ACKNOWLEDGMENTS

Thanks to my colleague Dave Iberson-Hurst (dih@A3Informatics.com) for providing review and content to this paper.

RECOMMENDED READING

General book about graph databases: Robinson, Webber & Eifrem (2013). *Graph Databases.* Get free copy at https://neo4j.com/graph-databases-book/

CONTACT INFORMATION

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

Kirsten Walther Langendorf, S-CUBED, A3 Informatics
Lille Strandstraede 20C 5., DK-1254 Copenhagen K
Work Phone: +45 31 65 76 55
Email:
kl@s-cubed.dk
kl@a3informatics.com

Web:
www.s-cubed.dk
www.a3informatics.com

Brand and product names are trademarks of their respective companies.