MRC-NIHR Trials Methodology Research Partnership: Webinar recording

Data sharing and the potential role of CDISC

Presented by Sharon Kean, Victoria Watson, and Jonathan Gibb (University of Liverpool)

28 March 2022

On behalf of HDR UK

The slides are also available below.

For any queries, please contact uktmn@nottingham.ac.uk

https://www.youtube.com/watch?v=RHNFULSVAgI

FUNDED BY
NIHR | National Institute for Health Research

Data sharing and the potential role of CDISC

Victoria Watson[1], Jonathan Gibb[2], Sharon Kean[1,2], Professor Carrol Gamble[1]

[1]Liverpool Clinical Trials Centre, University of Liverpool

[2] Robertson Centre for Biostatistics, University of Glasgow

Webinar

- Apologies from Carrol

- Introduction - Sharon
- Applying data standards using a real example (CDSIC)– Vicky
- Demo of mapping tool – Jonno
- Questions (please put in chat during session)

Introduction:

NIHR project

- ‘Implementing data standards (Clinical Data Interchange Standards Consortium - CDISC) in Clinical Trials Units (CTUs)’ is a project funded by the NIHR Trials CTU support for efficiency/innovative delivery of research.

Data Sharing in Clinical Trials

Why?

- Expectation for data to be made available for secondary research after the primary statistical analyses have been completed
- Sharing clinical trials data offers advantages that can advance clinical research and benefit patients.
- Clinical trial units have a responsibility to support and facilitate this process

Data Sharing Challenges

- Governance
  - trial participant privacy (anonymisation)
  - consent
  - Contractual

- Potential Concerns
  - When to share?
  - Clash of publication planning for original consortium

Data Sharing Challenges

- Resource required
  - Prepare the data
    - Is it ready to be shared?
  - Provide metadata to explain context
  - Choose sharing platform or provide copies of data
  - Responses to questions regarding the data. Who?
  - Can data standards help?
    - Upfront
    - Map at end

Data Standards in Clinical Trials

- Ability to share data efficiently
- Facilitate knowledge transfer regarding the context and description of trial data if using known methods
- Will facilitate scientific collaborations and future use of trial data.

Data Standards in Clinical Trials

- CDISC standards are internationally recognised by regulators but have had little take up within academic CTUs in the UK.
- One of the barriers is the unfamiliar terminology used.
- This webinar aims to demystify the use of CDSIC standards within an academic CTU setting using a real trial example.

What is CDISC?

- CDISC stands for “Clinical Data Interchange Standards Consortium” and is commonly used to refer to a set of data standards primarily used in the pharmaceutical industry

- The outputs of applying CDISC are used in clinical trials to submit data for drug approvals to the United States Food and Drug Association (FDA) and other regulators

- Only 7% of CDSIC members are academic institutions[1]

- CDISC requires the production of SDTM and ADaM datasets, and the production of these requires technical programmers in Information Systems (IS), Data Management (DM) and statistics

[1] *https://www.cdisc.org/membership*

Benefits of CDISC standards

- Standardised way of presenting clinical trial data
- Harmonisation for data sharing
- Reusable programs can be set up for each SDTM with variable names and lengths prespecified, saving time and resources of staff
- Statistician will receive data in same format each time, also saving time and resources when deriving endpoints
- Potential to reduce the amount of data collected which doesn't map to an SDTM dataset as it could indicate the data is unnecessary
- From a management perspective, a CTU could be more attractive to pharma if they have CDISC experience and they can provide ADaM datasets as a service
- Using the UKCRC CTU network there is potential to provide an environment to support each other in the implementation of CDSIC

Aims

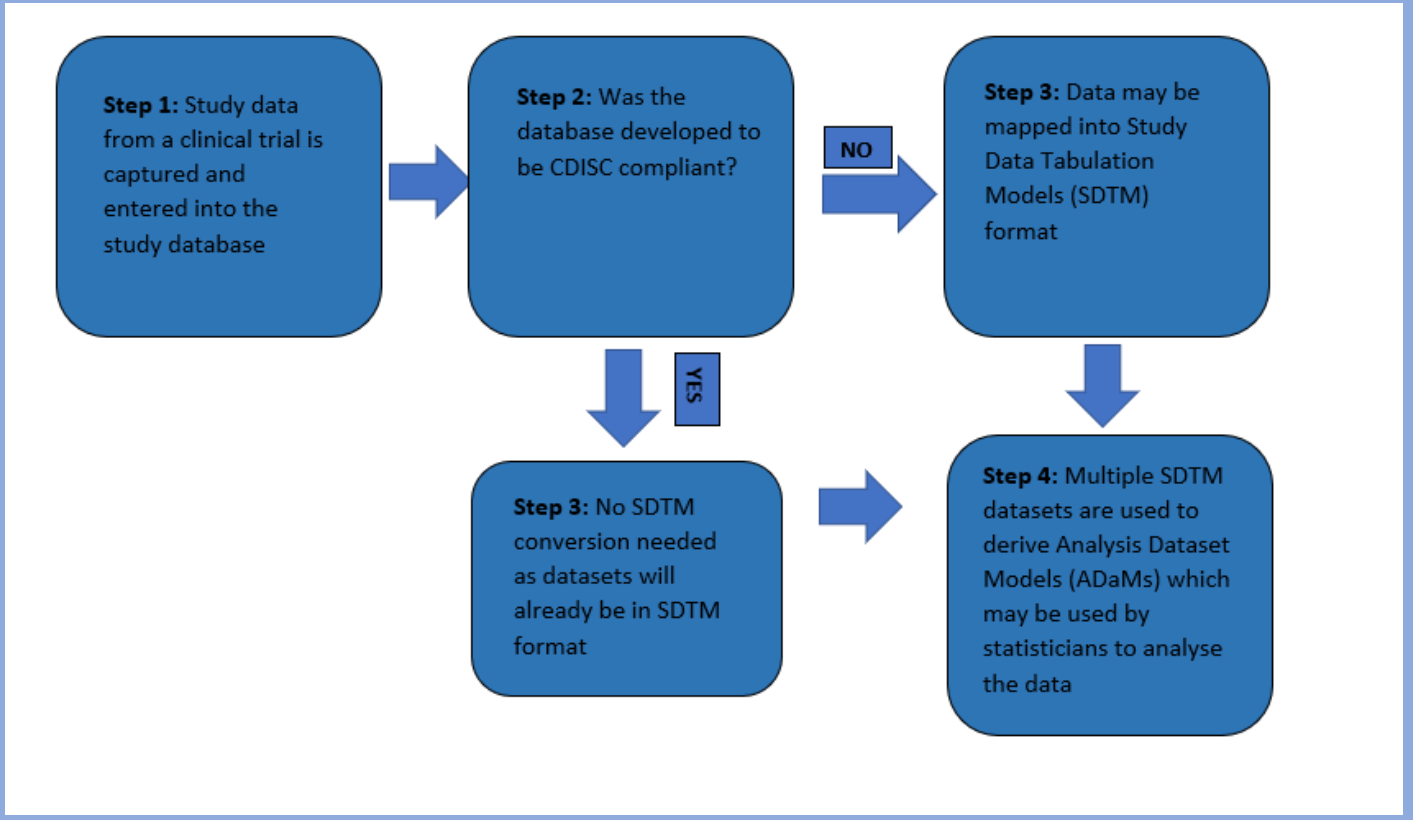
- To introduce the:
  - Clinical Data Interchange Standards Consortium (CDISC)
  - Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets
  - SDTM Implementation Guide (SDTM IG)
  - SDTM Controlled Terminology Document

- To provide a worked example of conversion of clinical trial data into an SDTM dataset by:
  - Demonstrating how the SDTM IG and SDTM Controlled Terminology Document are used in practice
  - Annotating Case Report Forms (aCRFs) to denote SDTM mappings

Applying data standards using a real example (CDSIC)

Steps for an SDTM Conversion

- The flow chart below demonstrates an overview of the steps involved for converting clinical trial data into SDTM format



Example SDTM datasets

- SDTM domains are commonly known as datasets or tables
- Below is an example of the output of applying the SDTM model to Laboratory Test Results data (LB) and Demographics data (DM)

SDTM LB Domain

VIEWTABLE: Work.Lb

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	VISITNUM	VISIT	VISITDY
1	STUDY001	LB	000101	1	CRP	C Reactive Protein	Haematology	10.2	mg/L	1	Visit 1: Baseline	0
2	STUDY001	LB	000101	2	CRP	C Reactive Protein	Haematology	9.8	mg/L	2	Visit 2: 2 Weeks Post Randomisation	14
3	STUDY001	LB	000101	3	CRP	C Reactive Protein	Haematology	9.3	mg/L	3	Visit 3: 4 Weeks Post Randomisation	28
4	STUDY001	LB	000101	4	CRP	C Reactive Protein	Haematology	8.9	mg/L	4	Visit 4: 6 Month Follow-up	182
5	STUDY001	LB	000102	1	CRP	C Reactive Protein	Haematology	9.6	mg/L	1	Visit 1: Baseline	0
6	STUDY001	LB	000102	2	CRP	C Reactive Protein	Haematology	9.9	mg/L	2	Visit 2: 2 Weeks Post Randomisation	14
7	STUDY001	LB	000102	3	CRP	C Reactive Protein	Haematology	10.1	mg/L	3	Visit 3: 4 Weeks Post Randomisation	28
8	STUDY001	LB	000102	4	CRP	C Reactive Protein	Haematology	10.3	mg/L	4	Visit 4: 6 Month Follow-up	182

SDTM DM Domain

VIEWTABLE: Work.Dm

	STUDYID	DOMAIN	USUBJID	SUBJID	BRTHDTC	AGE	AGEU	SEX
1	STUDY001	DM	000101	000101	1978-08-17	43	YEARS	F
2	STUDY001	DM	000102	000102	1985-04-24	36	YEARS	M

Example ADaM dataset

- ADaMs combine multiple pieces of information from various SDTM datasets. In the image below, we have combined the SDTM LB with the SDTM DM dataset
- The variables for the trial analysis are also derived, which can be seen in the variables Change “CHG” and Percent change “PCHG”
- The variables “SEX” and “AGE” are from the SDTM DM dataset, so that they could be used in the analysis of this ADaM dataset
- This is the dataset that would be analysed by the statistician, submitted to regulatory authorities and potentially onwards shared

VIEWTABLE: Work.Adcrp

	STUDYID	USUBJID	PARAM	PARAMCD	AVISIT	AVISITN	SEX	AGE	AVAL	BASE	CHG	PCHG
1	STUDY001	000101	C Reactive Protein	CRP	Visit 1: Baseline	1	F	43	10.2	10.2	.	.
2	STUDY001	000101	C Reactive Protein	CRP	Visit 2: 2 Weeks Post Randomisation	2	F	43	9.8	10.2	-0.4	-3.9
3	STUDY001	000101	C Reactive Protein	CRP	Visit 3: 4 Weeks Post Randomisation	3	F	43	9.3	10.2	-0.9	-8.8
4	STUDY001	000101	C Reactive Protein	CRP	Visit 4: Follow-up	4	F	43	8.9	10.2	-1.3	-13
5	STUDY001	000102	C Reactive Protein	CRP	Visit 1: Baseline	1	M	36	9.6	9.6	.	.
6	STUDY001	000102	C Reactive Protein	CRP	Visit 2: 2 Weeks Post Randomisation	2	M	36	9.9	9.6	0.3	3.13
7	STUDY001	000102	C Reactive Protein	CRP	Visit 3: 4 Weeks Post Randomisation	3	M	36	10.1	9.6	0.5	5.21
8	STUDY001	000102	C Reactive Protein	CRP	Visit 4: Follow-up	4	M	36	10.3	9.6	0.7	7.29

Steps when performing an SDTM conversion of clinical trial data

- Usually, the first step when converting a full clinical trial database into SDTM datasets, is to create the **trial design domains**
- These are all administration data from the protocol, and no live data (results) are used to populate them
  - These are the Trial Summary (TS), Trial Inclusion/Exclusion Criteria (TI), Trial Elements (TE) and Trial Visits (TV) domains

Example TS, TI and TV Domains

The Trial Visits domain (TV) :

- This contains information on the visits for the trial, similar to the schedule of assessments in the study protocol
  - The variables “ARM” (Description of Planned Arm) and “ARMCD” (Planned Arm Code) may be included in this dataset if the visit days or start rules vary by treatment arm
- The variable “TVSTRL” indicates the start rule for each visit

VIEWTABLE: Work.Tv

	STUDYID	DOMAIN	VISITNUM	VISIT	VISITDY	TVSTRL
1	STUDY001	TV	1	Visit 1: Baseline	0	Start of Screen Epoch
2	STUDY001	TV	2	Visit 2: 2 Weeks Post Randomisation	14	2 Weeks (+/- 4 days) after randomisation
3	STUDY001	TV	3	Visit 3: 4 Weeks Post Randomisation	28	4 Weeks (+/- 4 days) after randomisation
4	STUDY001	TV	4	Visit 4: 6 Month Follow-up	182	6 Months (+/- 14 days) after randomisation

The Trial Inclusion domain (TI) :

- This contains a full list of inclusion and exclusion criteria as per the study protocol
- The variable TIVERS refers to the study protocol version

VIEWTABLE: Work.Ti

	STUDYID	DOMAIN	IETESTCD	IETEST	IECAT	TIVERS
1	STUDY001	TI	INCL01	Those with a confirmed diagnosis of CD	INCLUSION	1
2	STUDY001	TI	INCL02	Identified as having vitamin D deficiency < 50 nmol/L 25(OH)D in the screening study	INCLUSION	1
3	STUDY001	TI	INCL03	Must be at least 18 years of age	INCLUSION	1
4	STUDY001	TI	EXCL01	Currently taking over the counter vitamin D, fish oil or multi-vitamin supplementation and unwilling to stop this to participate in the feasibility trial	EXCLUSION	1
5	STUDY001	TI	EXCL02	Currently receiving Vitamin D containing supplementation prescribed by a healthcare professional	EXCLUSION	1
6	STUDY001	TI	EXCL03	Currently receiving Bisphosphonates	EXCLUSION	1

Trial Summary (TS) Domain

- The TS domain contains information which the majority can be populated from the study protocol
- In our example, we can see information such as the trial title, the length of the trial, is the trial randomised and the sex of participants
- This information is captured in the variable “TSVAL” (Parameter Value), and the variables “TSPARMCD” (Trial Summary Parameter Short Name) and “TSPARM” (Trial Summary Parameter) indicate what TSVAL is capturing
- The variables “TSVALCD” (Parameter Value Code), “TSVCDREF” (Name of the Reference Terminology) and “TSVCDVER” (Version of the Reference Terminology) refer to if a Codelist is used when populating TSVAL, Codelists are explained later on

VIEWTABLE: Work.Ts

	STUDYID	DOMAIN	TSSEQ	TSPARMCD	TSPARM	TSVAL	TSVALCD	TSVCDREF	TSVCDVER
1	STUDY001	TS	1	AGEMAX	Planned Maximum Age of Subjects	P65Y		ISO 8601	
2	STUDY001	TS	1	AGEMIN	Planned Minimum Age of Subjects	P18Y		ISO 8601	
3	STUDY001	TS	1	LENGTH	Trial Length	P6M		ISO 8601	
4	STUDY001	TS	1	PLANSUB	Planned Number of Subjects	150			
5	STUDY001	TS	1	RANDOM	Trial is Randomized	Y	C49488	CDISC	2011-06-10
6	STUDY001	TS	1	SEXPOP	Sex of Participants	BOTH	C49636	CDISC	2011-06-10
7	STUDY001	TS	1	TBLIND	Trial Blinding Schema	DOUBLE BLIND	C15228	CDISC	2011-06-10
8	STUDY001	TS	1	TSCNTRL	Control Type	PLACBEO	C49648	CDISC	2011-06-10
9	STUDY001	TS	1	INDIC	Trial Disease/Condition Indication	Tonic-Clonic Epilepsy (Disorder)	352818000	SNOMED	
10	STUDY001	TS	1	TITLE	Trial Title	A 6 Month Study of Oral Gabapentin vs. Placebo in Subjects with Epilepsy due to Neurofibromatosis			
11	STUDY001	TS	1	TPHASE	Trial Phase Classification	Phase II Trial	C15601	CDSIC	2011-06-10
12	STUDY001	TS	1	TTYPE	Trial Type	EFFICACY	C49666	CDISC	2011-06-10
13	STUDY001	TS	2	TTYPE	Trial Type	SAFETY	C49667	CDISC	2011-06-10

Steps when performing an SDTM conversion of clinical trial data

- **Setting up Patient IDs and Visit Schedules**
  - STUDYID is a required variable in every SDTM dataset and is a unique identifier for the trial, for example “STUDYID01”. This is often the trial abbreviation which would also be located in the TS domain
  - Across all SDTM datasets, unique subject identifiers map to a variable called “USUBJID”
    - This variable is a required variable as per the SDTM IG in every SDTM dataset
    - This may often be the registration or randomisation number
  - To distinguish between different visits on the trial, the variables “VISITNUM” (Visit Number), “VISIT” (Visit Name) and “VISITDY” (Visit Day -days since randomisation) are also expected variables as per the SDTM IG in each SDTM dataset

SDTM Implementation Guide (SDTM IG)

- For guidance when creating SDTMs, there is the SDTM Implementation Guide (IG)
- The SDTM IG provides a full list of all SDTM domains and a breakdown of each variable that can be included in each of the domains
  - To access this, register for an account on the CDISC website, which is free to all clinicians and academic researchers

5.2 Demographics

DM – Description/Overview

A special purpose domain that includes a set of essential standard variables that describe each subject in a clinical study. It is the parent domain for all other observations for human clinical subjects.

DM – Specification

dm.xpt, Demographics — Special Purpose, Version 3.3. One record per subject, Tabulation.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format[1]	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	DM	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating STUDYID-SITEID-SUBJID.	Req
SUBJID	Subject Identifier for the Study	Char		Topic	Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.	Req
BRTHDTC	Date/Time of Birth	Char	ISO 8601	Record Qualifier	Date/time of birth of the subject.	Perm
AGE	Age	Num		Record Qualifier	Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).	Exp
AGEU	Age Units	Char	(AGEU)	Variable Qualifier	Units associated with AGE.	Exp
SEX	Sex	Char	(SEX)	Record Qualifier	Sex of the subject.	Req
RACE	Race	Char	(RACE)	Record Qualifier	Race of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, October, 2016) for guidance regarding the collection of race (https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf) See Assumption below regarding RACE.	Exp
ETHNIC	Ethnicity	Char	(ETHNIC)	Record Qualifier	The ethnicity of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, October, 2016) for guidance regarding the collection of ethnicity (https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf).	Perm

SDTM Controlled Terminology Document

- The SDTM Controlled Terminology Document can be found on the CDISC website in either xls or pdf format
  - It contains standard names and definitions for variables which may be a measurement or a test
- The SDTM IG will specify if a pre-defined codelist from the Controlled Terminology Document is required
  - We give an example of the variable “AGEU” in the SDTM DM domain, which from the SDTM IG we can see had a “Controlled Terms, Codelist or Format” also called “AGEU”
  - This informs us that values of AGEU are CDISC controlled , and we can find the accepted values for this variable in the SDTM Controlled Terminology Document under “AGEU”, as shown on the next slide

5.2 Demographics

DM – Description/Overview

A special purpose domain that includes a set of essential standard variables that describe each subject in a clinical study. It is the parent domain for all other observations for human clinical subjects.

DM – Specification

dm.xpt, Demographics — Special Purpose, Version 3.3. One record per subject, Tabulation.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format[1]	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	DM	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating STUDYID-SITEID-SUBJID.	Req
SUBJID	Subject Identifier for the Study	Char		Topic	Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.	Req
BRTHDTC	Date/Time of Birth	Char	ISO 8601	Record Qualifier	Date/time of birth of the subject.	Perm
AGE	Age	Num		Record Qualifier	Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).	Exp
AGEU	Age Units	Char	(AGEU)	Variable Qualifier	Units associated with AGE.	Exp
SEX	Sex	Char	(SEX)	Record Qualifier	Sex of the subject.	Req
RACE	Race	Char	(RACE)	Record Qualifier	Race of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, October, 2016) for guidance regarding the collection of race (https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf) See Assumption below regarding RACE.	Exp
ETHNIC	Ethnicity	Char	(ETHNIC)	Record Qualifier	The ethnicity of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, October, 2016) for guidance regarding the collection of ethnicity (https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf).	Perm

SDTM Controlled Terminology Document

- This is an example of the “AGEU” codelist in the SDTM Controlled Terminology document
- It demonstrates the CDISC Submission Values for the variable AGEU, which include “DAYS”, “HOURS”, “MONTHS” and “WEEKS”
- Going back to our example SDTM DM dataset from earlier, we have age measured in years so this tells us the CDSIC accepted value for AGEU is “YEARS”

A	B	C	D	E	F	G	H
Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66781		No	Age Unit	AGEU	Age Unit	Those units of time that are routinely used to express the age of a subject.	CDISC SDTM Age Unit Terminology
C25301	C66781		Age Unit	DAYS		A unit of measurement of time equal to 24 hours.	Day
C25529	C66781		Age Unit	HOURS	Hours; h; hr	A unit of measurement of time equal to 60 minutes.	Hour
C29846	C66781		Age Unit	MONTHS	Month	One of the 12 divisions of a year as determined by a calendar. It corresponds to the unit of time of approximately to one cycle of the moon's phases, about 30 days or 4 weeks. (NCI)	Month
C29844	C66781		Age Unit	WEEKS	Week	Any period of seven consecutive days. (NCI)	Week
C29848	C66781		Age Unit	YEARS	Year	The period of time that it takes for Earth to make a complete revolution around the sun, approximately 365 days; a specific one year period. (NCI)	Year

VIEWTABLE: Work.Dm

	STUDYID	DOMAIN	USUBJID	SUBJID	BRTHDTC	AGE	AGEU	SEX
1	STUDY001	DM	000101	000101	1978-08-17	43	YEARS	F
2	STUDY001	DM	000102	000102	1985-04-24	36	YEARS	M

SDTM Controlled Terminology Document

- Other examples where this is not so straightforward include the controlled lists for the variables RACE and ETHNIC, where there are only a small list of CDISC accepted values
- These are not extensible and additional category's have to be in a supplementary domain for demographics called SUPPDM which will be shown in a further example
- An alternative would be to map race as "OTHER"

RACE Codelist in the SDTM Controlled Terminology Document

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C74457		No	Race	RACE	Race	Terminology codelist used to identify the race of an individual within the Clinical Data Interchange Standards Consortium Study Data Tabulation Model.	CDISC SDTM Race Terminology
C41259	C74457		Race	AMERICAN INDIAN OR ALASKA NATIVE		A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. (FDA)	American Indian or Alaska Native
C41260	C74457		Race	ASIAN		A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (FDA)	Asian
C16352	C74457		Race	BLACK OR AFRICAN AMERICAN		A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American." (FDA)	Black or African American
C41219	C74457		Race	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER		Denotes a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. The term covers particularly people who identify themselves as part-Hawaiian, Native Hawaiian, Guamanian or Chamorro, Carolinian, Samoan, Chuu. (FDA)	Native Hawaiian or Other Pacific Islander
C43234	C74457		Race	NOT REPORTED	Not reported	Not provided or available.	Not Reported
C17998	C74457		Race	UNKNOWN	U; UNK; Unknown	Not known, not observed, not recorded, or refused. (NCI)	Unknown
C41261	C74457		Race	WHITE		Denotes a person with European, Middle Eastern, or North African ancestral origin who identifies, or is identified, as White. (FDA)	White

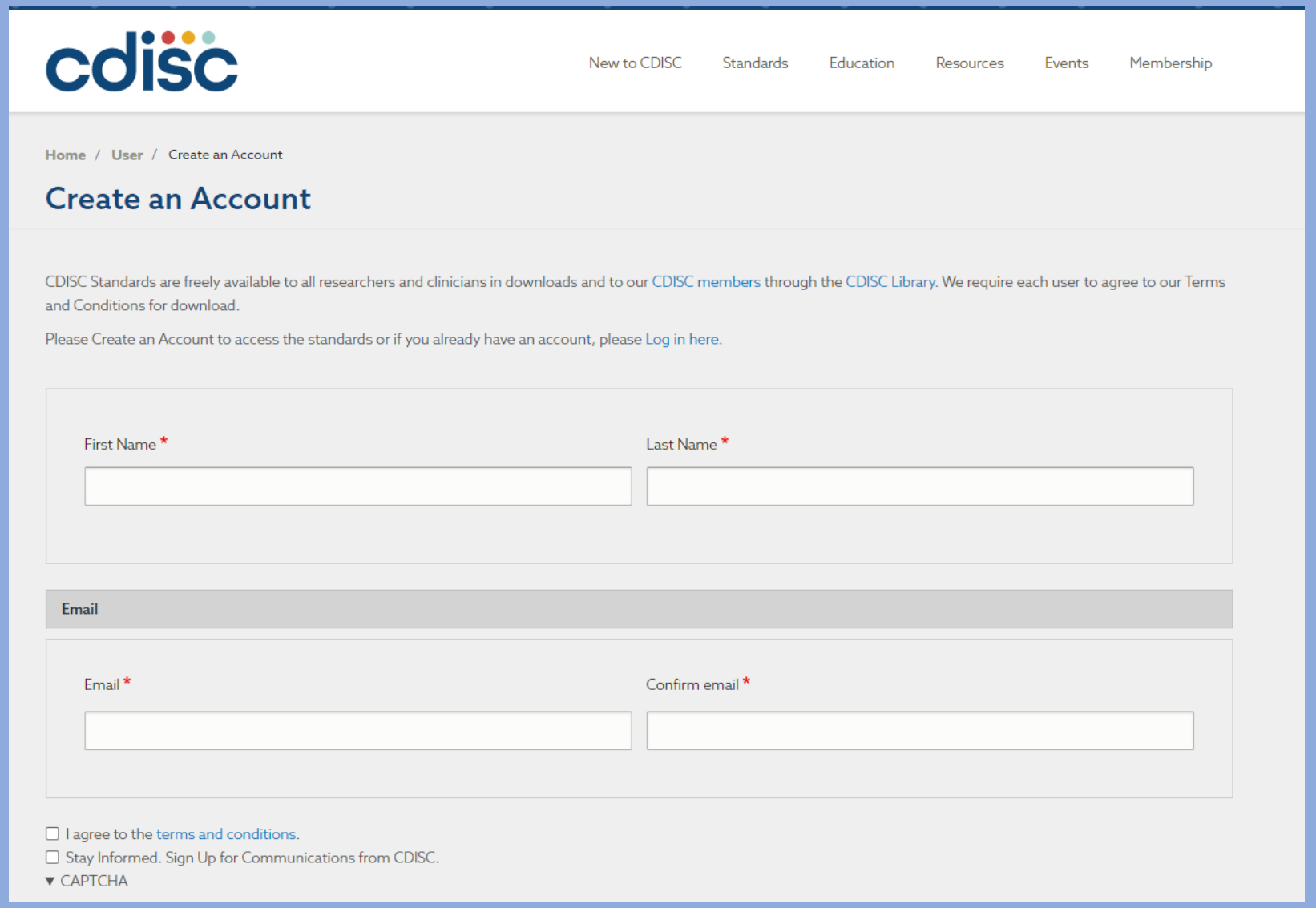
SDTM Controlled Terminology Document

- Here is the codelist for the variable “ETHNIC”, this shows that there are few category’s which are not universally appropriate, therefore SUPPDM would be recommended again here
- This would allow you to describe ethnicity as captured on the CRF by including it within the SUPPDM domain

ETHNIC Codelist in the SDTM Controlled Terminology Document

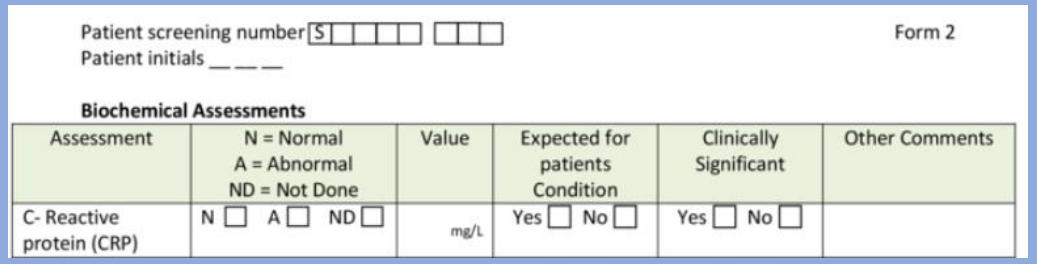
A	B	C	D	E	F	G	H
Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66790		No	Ethnic Group	ETHNIC	Ethnic Group	A social group characterized by a distinctive social and cultural tradition maintained from generation to generation, a common history and origin and a sense of identification with the group; members of the group have distinctive features in their way of life, shared experiences and often a common genetic heritage; these features may be reflected in their experience of health and disease. (NCI)	CDISC SDTM Ethnic Group Terminology
C17459	C66790		Ethnic Group	HISPANIC OR LATINO		A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race. (NCI)	Hispanic or Latino
C41222	C66790		Ethnic Group	NOT HISPANIC OR LATINO		A person not of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. An arbitrary ethnic classification. (NCI)	Not Hispanic or Latino
C43234	C66790		Ethnic Group	NOT REPORTED	Not reported	Not provided or available.	Not Reported
C17998	C66790		Ethnic Group	UNKNOWN	U; UNK; Unknown	Not known, not observed, not recorded, or refused. (NCI)	Unknown

CDSIC Website



Steps when performing an SDTM conversion of clinical trial data – worked example

- To demonstrate how a particular data item would be mapped to an SDTM dataset, we have a worked example highlighting how the laboratory biomarker of C-Reactive Protein would be mapped to a SDTM dataset



Patient screening number S☐☐☐☐ ☐☐☐ Form 2
Patient initials __ __ __

Biochemical Assessments

Assessment	N = Normal A = Abnormal ND = Not Done	Value	Expected for patients Condition	Clinically Significant	Other Comments
C- Reactive protein (CRP)	N ☐ A ☐ ND ☐	mg/L	Yes ☐ No ☐	Yes ☐ No ☐	

  - When converting a data item or variable into SDTM format, first determine which SDTM dataset it belongs to
  - To do this we would first find the appropriate domain from the full list of SDTM domains given on page 18 of the SDTM IG
  - As C-Reactive Protein is a laboratory result, it belongs to the Laboratory Test Results (LB) SDTM dataset
  - Our next step would be to locate the details of the LB domain in the SDTM IG as shown on the next slide

Steps when performing an SDTM conversion of clinical trial data – worked example

LB – Specification

lb.xpt, Laboratory Test Results — Findings, Version 3.3. One record per lab test per time point per visit per subject, Tabulation.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format[1]	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	LB	Identifier	Two-character abbreviation for the domain.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
LBSEQ	Sequence Number	Num		Identifier	Sequence number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req
LBGRPID	Group ID	Char		Identifier	Used to tie together a block of related records in a single domain for a subject.	Perm
LBREFID	Specimen ID	Char		Identifier	Internal or external specimen identifier. Example: Specimen ID.	Perm
LBSPID	Sponsor-Defined Identifier	Char		Identifier	Sponsor-defined reference number. Perhaps preprinted on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on the Lab page.	Perm
LBTESTCD	Lab Test or Examination Short Name.	Char	(LBTESTCD)	Topic	Short name of the measurement, test, or examination described in LBTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in LBTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST" is not valid). LBTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: "ALT", "LDH".	Req
LBTEST	Lab Test or Examination Name	Char	(LBTEST)	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. Note any test normally performed by a clinical laboratory is considered a lab test. The value in LBTEST cannot be longer than 40 characters. Examples: "Alanine Aminotransferase", "Lactate Dehydrogenase".	Req
LBCAT	Category for Lab Test	Char	*	Grouping Qualifier	Used to define a category of related records across subjects. Examples: "HEMATOLOGY", "URINALYSIS", "CHEMISTRY".	Exp
LBSCAT	Subcategory for Lab Test	Char	*	Grouping Qualifier	A further categorization of a test category such as "DIFFERENTIAL", "COAGULATION", "LIVER FUNCTION", "ELECTROLYTES".	Perm
LBORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the measurement or finding as originally received or collected.	Exp
LBORRESU	Original Units	Char	(UNIT)	Variable Qualifier	Original units in which the data were collected. The unit for LBORRES. Example: "g/L".	Exp
LBORNRLO	Reference Range Lower Limit in Orig Unit	Char		Variable Qualifier	Lower end of reference range for continuous measurement in original units. Should be populated only for continuous results.	Exp

- This shows an extract of the LB domain in the SDTM IG, which shows the list of variables for this dataset (column header within the SDTM LB Domain)
- Using the final Column "Core" we can see if the variable is a required "Req", expected "Exp" or permissible "Perm"
- We know CRP is a laboratory test and to locate how to describe that and to populate the column, we have to go to the codelist as indicated(column 4 in the table)
  - If we didn't know it was in the LB domain, then we would search the SDTM Controlled terminology for "CRP" and find it was under LBTESTCD which highlights it belongs in SDTM LB
- LBTESTCD and LBTEST are column headers and the codelist tells us the values which are required. In this case, these are "CRP" and "C Reactive Protein" which you will see on the next slide.
- The patients laboratory results and the laboratory units goes in the "LBORRES" and "LBORRESU" variables

Steps when performing an SDTM conversion of clinical trial data – worked example

- To locate these codelists in the SDTM Controlled Terminology Document search (Ctrl+F) the words “LBTESTCD” (below) and “LBTEST” (next slide)
- This codelist is extensible as indicated in the column “Codelist Extensible (Yes/No)”
  - This means that if we have an unusual laboratory test that we cannot find in the LBTESTCD codelist, we can create our own LBTESTCD
- Within the LBTESTCD codelist, we would now search (Ctrl+F) for the word C-Reactive Protein
- From this, we can see the variable LBTESTCD has a value of “CRP”

An example of the codelist for LBTESTCD in the SDTM Controlled Terminology Document is below

050 C64548

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C65047		Yes	Laboratory Test Code	LBTESTCD	Laboratory Test Code	Terminology used for laboratory test codes of the CDISC Study Data Tabulation Model.	CDISC SDTM Laboratory Test Code Terminology
C100429	C65047		Laboratory Test Code	A1AGLP	Alpha-1 Acid Glycoprotein	A measurement of the alpha-1 acid glycoprotein in a biological specimen.	Alpha-1 Acid Glycoprotein Measurement
C80167	C65047		Laboratory Test Code	A1ANTRYP	Alpha-1 Antitrypsin; Serum Trypsin Inhibitor	A measurement of the alpha-1 antitrypsin in a biological specimen.	Alpha-1 Antitrypsin Measurement
C100462	C65047		Laboratory Test Code	A1MCREAT	Alpha-1 Microglobulin/Creatinine	A relative measurement (ratio or percentage) of the alpha-1 microglobulin to creatinine in a biological specimen.	Alpha-1 Microglobulin to Creatinine Ratio Measurement
C100461	C65047		Laboratory Test Code	A1MICG	Alpha-1 Microglobulin; Protein HC	A measurement of the alpha-1 microglobulin in a biological specimen.	Alpha-1 Microglobulin Measurement
C80168	C65047		Laboratory Test Code	A2MACG	Alpha-2 Macroglobulin	A measurement of the alpha-2 macroglobulin in a biological specimen.	Alpha-2 Macroglobulin Measurement
C172524	C65047		Laboratory Test Code	A73OXC	7-Alpha hydroxy-4-cholesten-3-one; 7-alpha-Hydroxy-4-cholesten-3-one	A measurement of the 7-alpha-hydroxy-4-cholesten-3-one in a biological specimen.	7-alpha-Hydroxy-4-cholesten-3-one Measurement
C154761	C65047		Laboratory Test Code	AAMAPAC	Alpha-Aminoadipate; Alpha-Aminoadipic Acid	A measurement of the alpha-aminoadipic acid in a biological specimen.	Alpha-Aminoadipic Acid Measurement
C154759	C65047		Laboratory Test Code	AAMBTAC	Alpha-Aminobutyric Acid; Alpha-aminobutyrate	A measurement of the alpha-aminobutyric acid in a biological specimen.	Alpha-Aminobutyric Acid Measurement
C100430	C65047		Laboratory Test Code	AAP	Alanine Aminopeptidase	A measurement of the alanine aminopeptidase in a biological specimen.	Alanine Aminopeptidase Measurement
C64548	C65047		Laboratory Test Code	CRP	C Reactive Protein	A measurement of the C reactive protein in a biological specimen.	C-Reactive Protein Measurement
C147324	C65047		Laboratory Test Code	CRTCLRBS	Creatinine Clearance Adjusted for BSA	A measurement of the volume of serum or plasma that would be cleared of creatinine by excretion of urine for a specified unit of time (e.g. one minute), adjusted for body surface area.	Creatinine Clearance Adjusted for BSA
C150847	C65047		Laboratory Test Code	CRTCLRE	Creatinine Clearance, Estimated	An estimate of the volume of serum or plasma that would be cleared of creatinine by excretion of urine for a specified unit of time (e.g. one minute).	Estimated Creatinine Clearance

Steps when performing an SDTM conversion of clinical trial data – worked example

- We would repeat this process to determine the “CDISC Submission Value” for the codelist LBTEST, which we can see below is “C Reactive Protein”

An example of the codelist for LBTEST in the SDTM Controlled Terminology Document

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C67154		Yes	Laboratory Test Name	LBTEST	Laboratory Test Name	Terminology used for laboratory test names of the CDISC Study Data Tabulation Model.	CDISC SDTM Laboratory Test Name Terminology
C165944	C67154		Laboratory Test Name	Bruton's Tyrosine Kinase, Free	Bruton's Tyrosine Kinase, Free	A measurement of the free Bruton's tyrosine kinase in a biological specimen.	Free Bruton's Tyrosine Kinase Measurement
C75352	C67154		Laboratory Test Name	Buprenorphine	Buprenorphine	A measurement of the buprenorphine drug present in a biological specimen.	Buprenorphine Measurement
C74701	C67154		Laboratory Test Name	Burr Cells	Burr Cells; Echinocytes	A measurement of the Burr cells (erythrocytes characterized by the presence of small, blunt projections evenly distributed across the cell surface) in a biological specimen.	Burr Cell Count
C75364	C67154		Laboratory Test Name	Butabarbital	Butabarbital	A measurement of the butabarbital in a biological specimen.	Butabarbital Measurement
C75365	C67154		Laboratory Test Name	Butalbital	Butalbital	A measurement of the butalbital present in a biological specimen.	Butalbital Measurement
C111142	C67154		Laboratory Test Name	Butyrylcholinesterase	Acylcholine Acylhydrolase; Butyrylcholinesterase; Non-neuronal Cholinesterase; Plasma Cholinesterase; Pseudocholinesterase	A measurement of the total butyrylcholinesterase in a biological specimen.	Butyrylcholinesterase Measurement
C64548	C67154		Laboratory Test Name	C Reactive Protein	C Reactive Protein	A measurement of the C reactive protein in a biological specimen.	C-Reactive Protein Measurement
C122103	C67154		Laboratory Test Name	C-C Chemokine Receptor Type 5	C-C Chemokine Receptor Type 5; CD195	A measurement of the CCR5, chemokine (C-C motif) receptor type 5, in a biological specimen.	C-C Chemokine Receptor Type 5 Measurement
C74736	C67154		Laboratory Test Name	C-peptide	C-peptide	A measurement of the C (connecting) peptide of insulin in a biological specimen.	C-peptide Measurement
C150837	C67154		Laboratory Test Name	C-peptide/Creatinine	C-peptide/Creatinine	A relative measurement (ratio or percentage) of the C-peptide to creatinine in a biological specimen.	C-peptide to Creatinine Ratio Measurement
C74702	C67154		Laboratory Test Name	Cabot Rings	Cabot Rings	A measurement of the Cabot rings (red-purple staining, threadlike, ring or figure 8 shaped filaments in an erythrocyte) in a biological specimen.	Cabot Ring Count
C75346	C67154		Laboratory Test Name	Caffeine	Caffeine	A measurement of the caffeine in a biological specimen.	Caffeine Measurement
C125942	C67154		Laboratory Test Name	Calbindin	Calbindin	A measurement of the total calbindin in a biological specimen.	Calbindin Measurement

Steps when performing an SDTM conversion of clinical trial data – worked example

- As we now know the CDSIC submission value for C-Reactive Protein for the variables LBTEST and LBTESTCD, our next step would be to put this information together and begin creating the SDTM LB domain
- This can be done by:
  - Writing a program to map the model to the data to create an output
    - This could be done by the statistician e.g. in SAS or by a programmer e.g. in SQL
  - The Clinical Data Management System (CDMS) you have been using may have a CDISC compatible module
  - Utilising a CDSIC mapping tool
    - As part of this project a prototype will be demonstrated

Example SAS Code

- Here is an example of SAS Code used to generate the example of the LB Domain we provided earlier
- Note that each section of code preceded by “output” equates to one row, as we created test data this example
- The ordering of the variables in the SDTM domains are important, and they are listed the required order within the SDTM IG. Here this order is specified within the “retain” statement

```
data lb1;
     length domain studyid lbcat lborresu lbtest visit $255. lbtestcd $8.;
     studyid="STUDY001";
     domain="LB";
     usubjid="000101";
     lbseq=1;
     lbtestcd="CRP";
     lbtest="C Reactive Protein";
     lbcat="Haematology";
     lborres=10.2;
     lborresu="mg/L";
     visitnum=01;
     visit="Visit 1: Baseline";
     visitdy=0;
     output;
     studyid="STUDY001";
     domain="LB";
     usubjid="000101";
     lbseq=2;
     lbtestcd="CRP";
     lbtest="C Reactive Protein";
     lbcat="Haematology";
     lborres=9.8;
     lborresu="mg/L";
     visitnum=02;
     visit="Visit 2: 2 Weeks Post Randomisation";
     visitdy=14;
     output;
run;

data lb;
     retain studyid domain usubjid lbseq lbtestcd lbtest lbcat lborres lborresu visitnum visit visitdy;
     set lb1;
run;


libname cdisc "M:\CDISC";

data cdisc.lb;
     set lb;
run;
```

Steps when performing an SDTM conversion of clinical trial data – worked example

- Here is the final result, demonstrating what this would look like in the LB SDTM dataset

VIEWTABLE: Work.Lb

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	VISITNUM	VISIT	VISITDY
1	STUDY001	LB	000101	1	CRP	C Reactive Protein	Haematology	10.2	mg/L	1	Visit 1: Baseline	0
2	STUDY001	LB	000101	2	CRP	C Reactive Protein	Haematology	9.8	mg/L	2	Visit 2: 2 Weeks Post Randomisation	14
3	STUDY001	LB	000101	3	CRP	C Reactive Protein	Haematology	9.3	mg/L	3	Visit 3: 4 Weeks Post Randomisation	28
4	STUDY001	LB	000101	4	CRP	C Reactive Protein	Haematology	8.9	mg/L	4	Visit 4: 6 Month Follow-up	182

- This example shows the laboratory results for C-Reactive Protein for one patient at 4 different time points, shown in the “VISITNUM”, “VISIT” and “VISITDY” variables
- Note that the variable “LBSEQ” is a unique identifier per lab result per patient that has to be assigned

Mapping CRFS to SDTM Datasets through the use of *a*CRFs

- Next, we extend our previous example from considering selected variables to a full Case Report Form (CRF)
- We will create SDTM annotated CRFs (*a*CRFs)
  - Real trial baseline CRF examples from 3 studies are provided
  - Use of colour co-ordinated *a*CRFs grouping data items to SDTM domains
- We will highlight some less straightforward examples and how they may be handled when converting clinical trial data into SDTM format

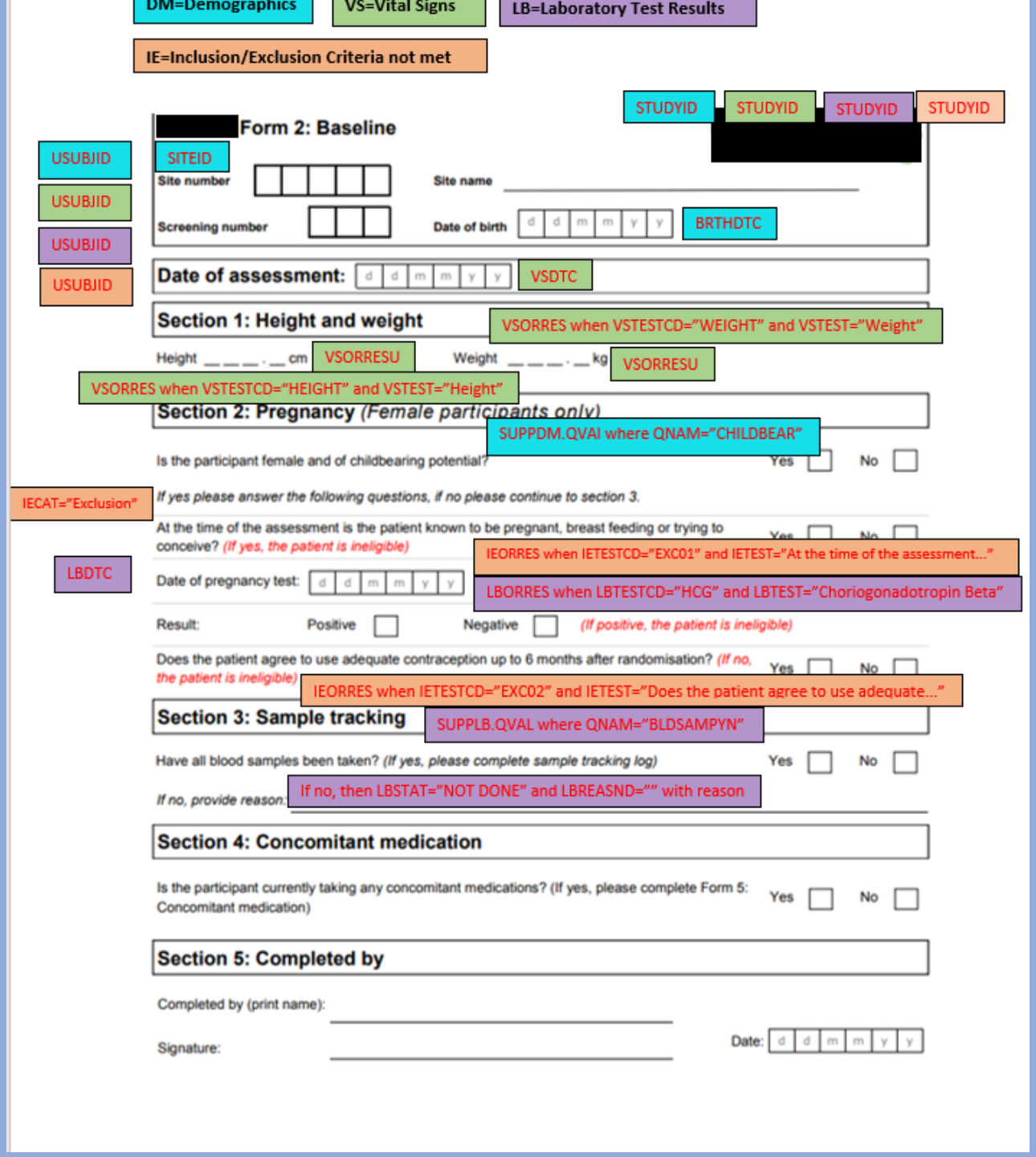
Mapping CRFS to SDTM Datasets through the use of *a*CRFs

- We will cover:
  - Data collected on CRFs which are not needed in SDTMs. These are often admin questions or checkboxes used for CRF completion
  - Trial specific data, such as outcome data which is a struggle to map into an SDTM dataset
  - CRFs which cover multiple SDTM domains, meaning one CRF maps to multiple SDTM domains

Example 1 *a*CRF

- This baseline CRF maps to several SDTM domains – DM (Demographics), VS (Vital Signs), LB (Laboratory Test Results) and IE (Inclusion Exclusion Criteria Not Met)

- We have colour coded each SDTM domain and their variables
  - Height and weight are directly mapped to SDTM within the VS domain, (coloured green)
  - Pregnancy test results directly map to the LB domain (coloured purple)
  - Not all collected clinical trial data will need to map to an SDTM dataset (Section 4 and 5)
  - Inclusion/exclusion will be covered on the next slide



Inclusion/Exclusion Criteria Domains

TI Domain

- Part of the trial design domains, usually created prior to data being collected using the study protocol

VIEWTABLE: Work.Ti

	STUDYID	DOMAIN	IETESTCD	IETEST	IECAT	TIVERS
1	STUDY001	TI	INCL01	Those with a confirmed diagnosis of CD	INCLUSION	1
2	STUDY001	TI	INCL02	Identified as having vitamin D deficiency < 50 nmol/L 25(OH)D in the screening study	INCLUSION	1
3	STUDY001	TI	INCL03	Must be at least 18 years of age	INCLUSION	1
4	STUDY001	TI	EXCL01	Currently taking over the counter vitamin D, fish oil or multi-vitamin supplementation and unwilling to stop this to participate in the feasibility trial	EXCLUSION	1
5	STUDY001	TI	EXCL02	Currently receiving Vitamin D containing supplementation prescribed by a healthcare professional	EXCLUSION	1
6	STUDY001	TI	EXCL03	Currently receiving Bisphosphonates	EXCLUSION	1

IE Domain: Inclusion/Exclusion criteria not met

- Based on live patient data indicating which inclusion/exclusion criteria patients failed on
- Here we see the difference between the two variables “VISITDY” and “IEDY” which are both measured in days since randomisation
  - VISITDY captures the scheduled day the visit should have took place on, whereas IEDY captures the day it actually took place on
  - VISITDY remains a consistent variable throughout SDTM domains, whereas variables such as IEDY or LBDY for the LB domain for example, varies depending on the date of the examination being recorded

VIEWTABLE: Work.Ie

	DOMAIN	USUBJID	IESEQ	IETESTCD	IETEST	IECAT	IEORRES	IESTRESC	VISITNUM	VISIT	VISITDY	IEDTC	IEDY
1	TI	0001006	1	INCL01	Those with a confirmed diagnosis of CD	INCLUSION	N	N	1	Visit 1: Baseline	0	2021-06-02	2
2	TI	0001006	2	INCL02	Identified as having vitamin D deficiency < 50 nmol/L 25(OH)D in the screening study	INCLUSION	N	N	1	Visit 1: Baseline	0	2021-06-02	2
3	TI	0001008	1	INCL01	Those with a confirmed diagnosis of CD	INCLUSION	N	N	1	Visit 1: Baseline	0	2021-06-04	4
4	TI	0001008	2	EXCL02	Currently receiving Vitamin D containing supplementation prescribed by a healthcare professional	EXCLUSION	Y	Y	1	Visit 1: Baseline	0	2021-06-04	4
5	TI	0001012	1	INCL03	Must be at least 18 years of age	INCLUSION	N	N	1	Visit 1: Baseline	0	2021-06-03	3
6	TI	0001014	1	EXCL03	Currently receiving Bisphosphonates	EXCLUSION	Y	Y	1	Visit 1: Baseline	0	2021-06-01	0

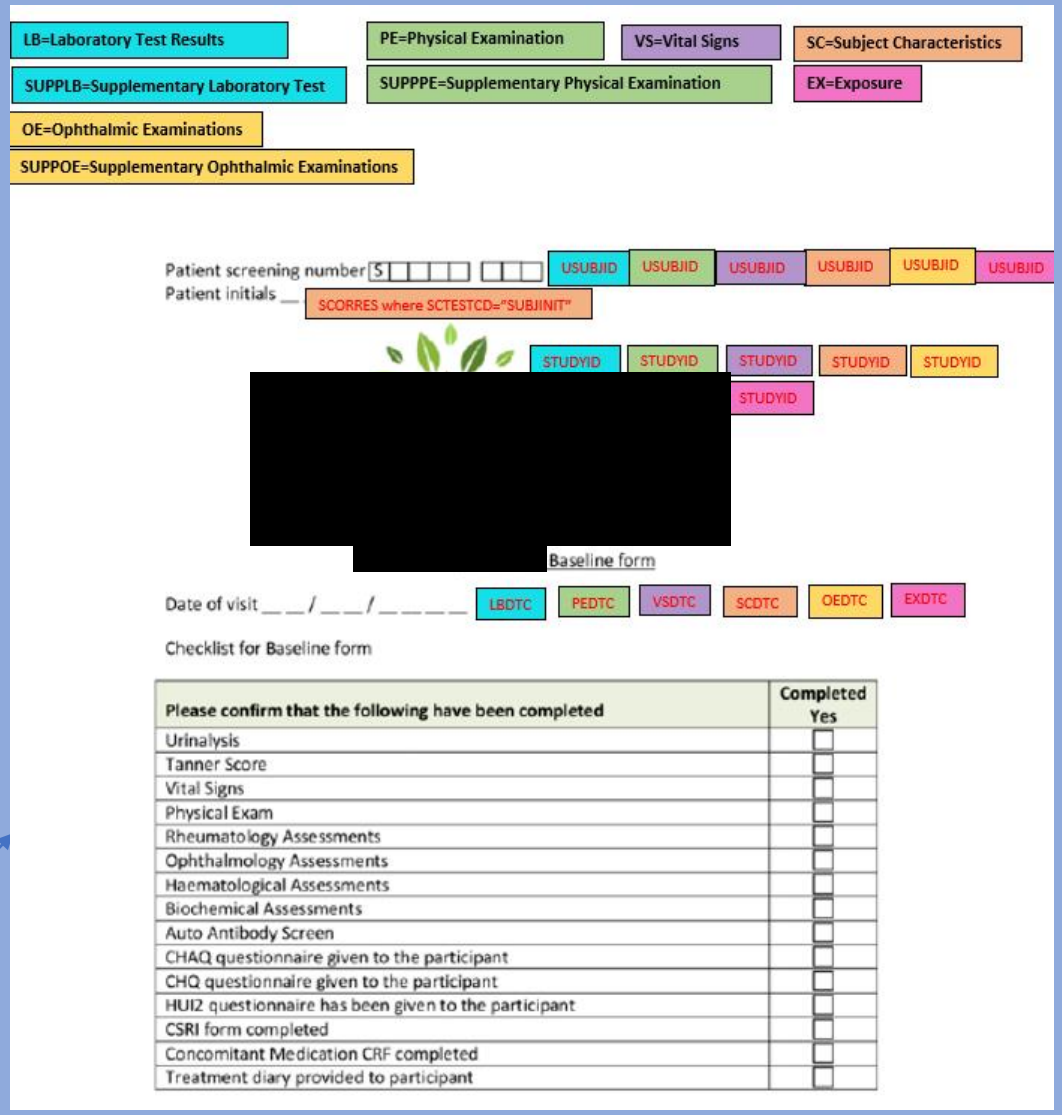
Example 2 aCRF

- This is our second example of an aCRF for a different study. This is a lot more detailed with 8 pages. This example is perhaps more common when capturing baseline clinical trial data

- The cover page only contains key information (subject ID and visit date) rather than trial content

- This example also demonstrates the use of supplementary qualifier SDTM datasets (SUPP--), for data items which don't fit into the original domains

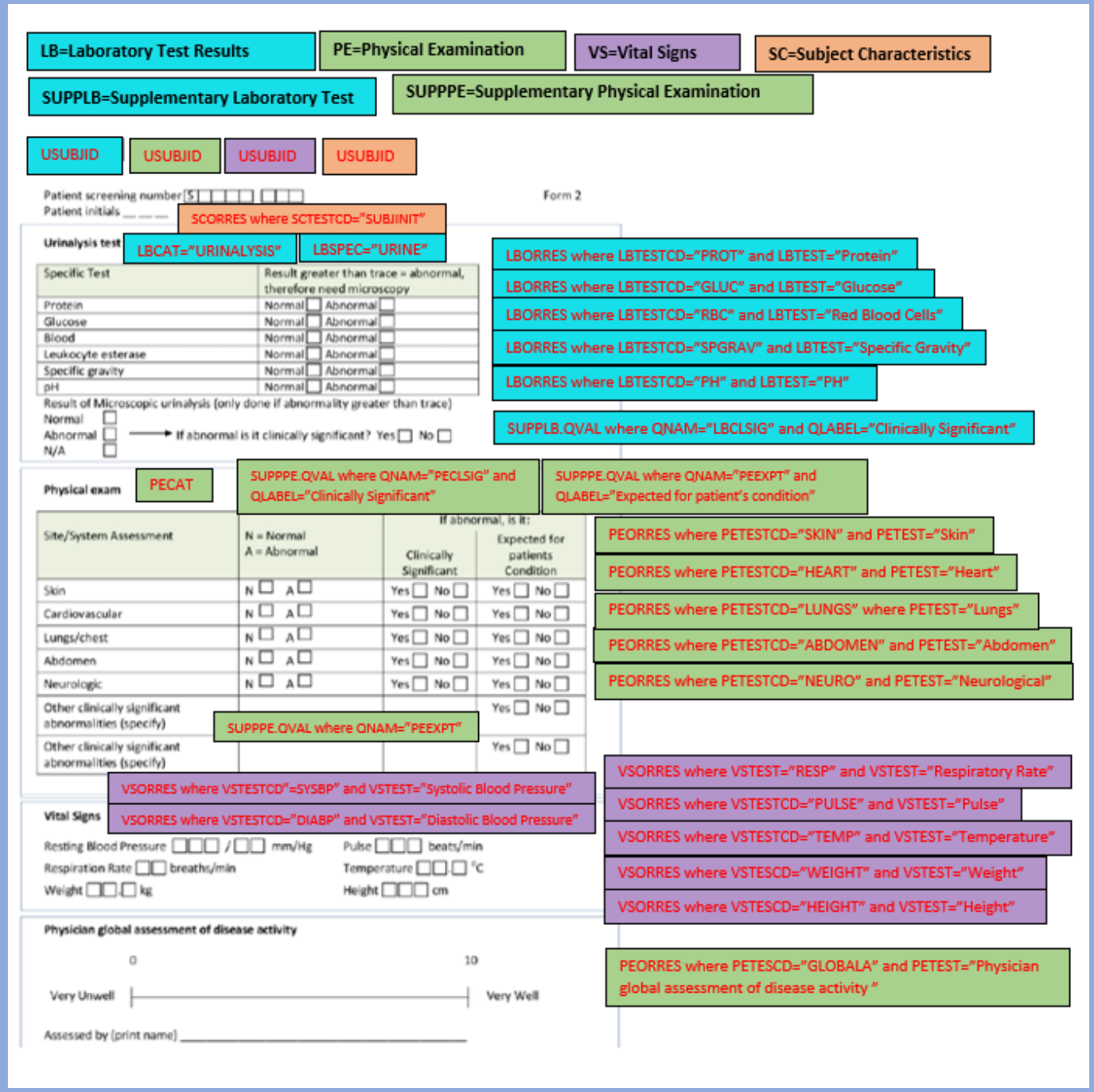
Again, these tick box questions don't need to be mapped to an SDTM dataset



Example 2 aCRF

- XXORRES denotes original result (LBORRES=5.4)
- XXTESTCD denotes test code (LBTESTCD="PROT")
- XXTEST denotes test name (LBTEST="Protein")
- Where XX represents the SDTM domain name

- This page maps completely to either the Laboratory (LB), Physical Examination (PE) or Vital Signs (VS) for test results, or the Subject Characteristics (SC) SDTM domain which contain the subject initials



Supplementary Domains

- Supplementary Domains are used when additional information is collected that doesn't fit
- Examples of supplementary domains include:
  - SUPPLB: Supplementary Laboratory test results domain
  - SUPPDM: Supplementary Demographics results domain
- We provide an example of the SUPPLB domain, capturing whether a laboratory result was clinically significant or not as captured on the CRF
  - The variables "IDVAR" and "IDVARVAL" are used to reference the specific record in the master domain (LB)
- We also provide an example of a SUPPDM domain, which captures the variables Race other ("RACEOTH"), Intention to Treat Flag ("ITTFL") and the Safety Population Flag ("SAFFL")
  - The variable "QORIG" is the origin of the variable, usually populated with "CRF" or "Derived"
  - The variable "RACEOTH" can be used where race is captured differently on the CRF than the accepted CDSIC values for RACE as previously seen

VIEWTABLE: Work.Supplb

	STUDYID	DOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
1	STUDY001	SUPPLB	000101	LBSEQ	1	LBCLSIG	Clinically Significant	Y	CRF
2	STUDY001	SUPPLB	000101	LBSEQ	2	LBCLSIG	Clinically Significant	Y	CRF
3	STUDY001	SUPPLB	000102	LBSEQ	1	LBCLSIG	Clinically Significant	N	CRF
4	STUDY001	SUPPLB	000102	LBSEQ	2	LBCLSIG	Clinically Significant	N	CRF

VIEWTABLE: Work.Suppdm

	STUDYID	DOMAIN	USUBJID	QNAM	QLABEL	QVAL	QORIG
1	STUDY001	SUPPDM	000101	RACEOTH	Race, Other	MAORI	CRF
2	STUDY001	SUPPDM	000101	ITTFL	ITT population flag	Y	Derived
3	STUDY001	SUPPDM	000101	SAFFL	Safety population flag	Y	Derived

Example 2 aCRF: Unmappable Data

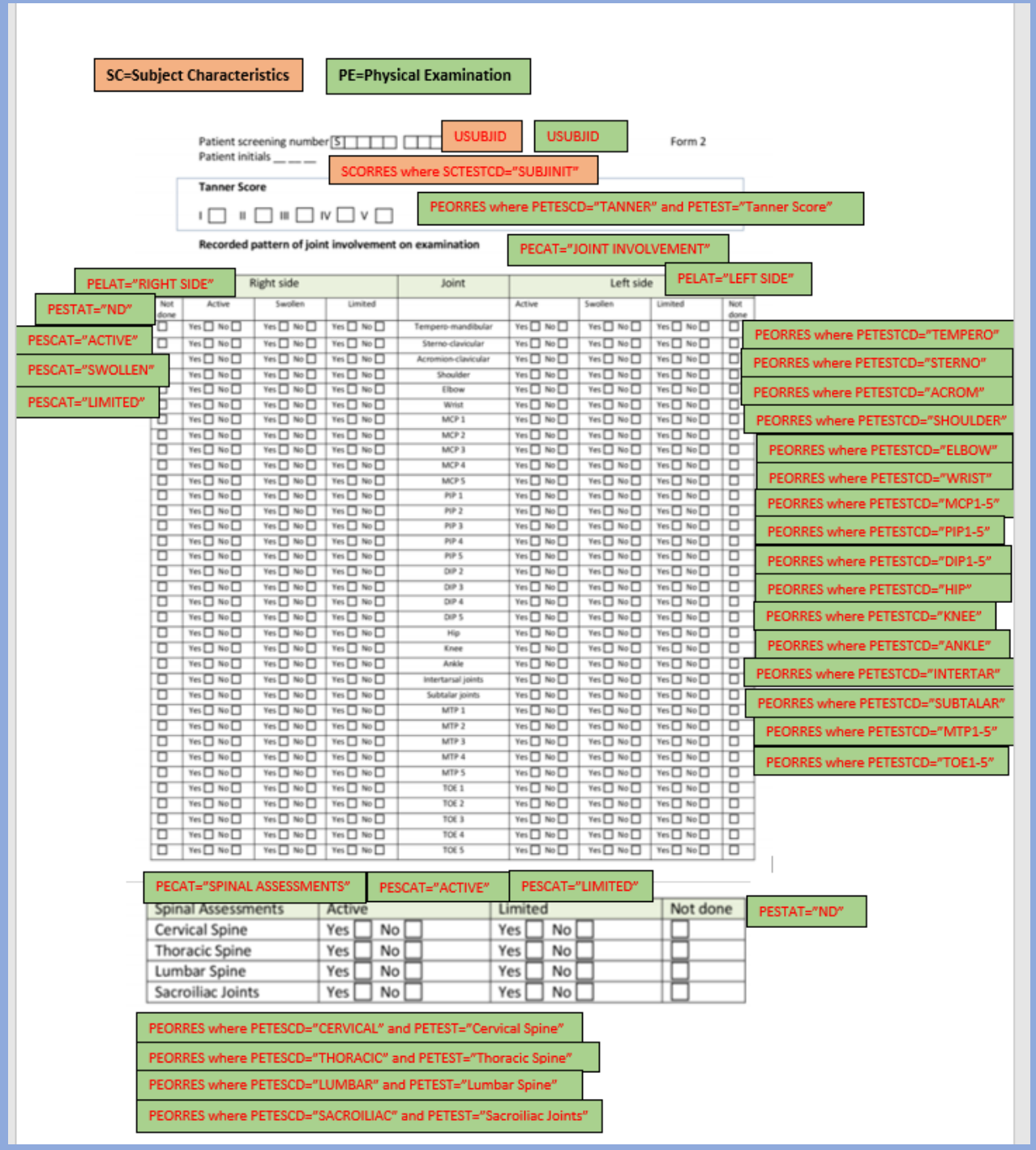
- However, other pages of this aCRF aren't as straightforward and it isn't clear which SDTM dataset the data will map to
- Encountering this may cause confusion to a programmer or statistician where to map the data

Recorded pattern of joint involvement on examination

Right side				Joint	Left side			
Not done	Active	Swollen	Limited		Active	Swollen	Limited	Not done
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Tempero-mandibular	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Sterno-clavicular	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Acromion-clavicular	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Shoulder	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Elbow	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Wrist	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MCP 1	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MCP 2	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MCP 3	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MCP 4	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MCP 5	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	PIP 1	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	PIP 2	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	PIP 3	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	PIP 4	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	PIP 5	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	DIP 2	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	DIP 3	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	DIP 4	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	DIP 5	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Hip	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Knee	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Ankle	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Intertarsal joints	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Subtalar joints	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MTP 1	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MTP 2	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MTP 3	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MTP 4	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	MTP 5	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	TOE 1	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	TOE 2	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	TOE 3	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	TOE 4	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐
☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	TOE 5	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	☐

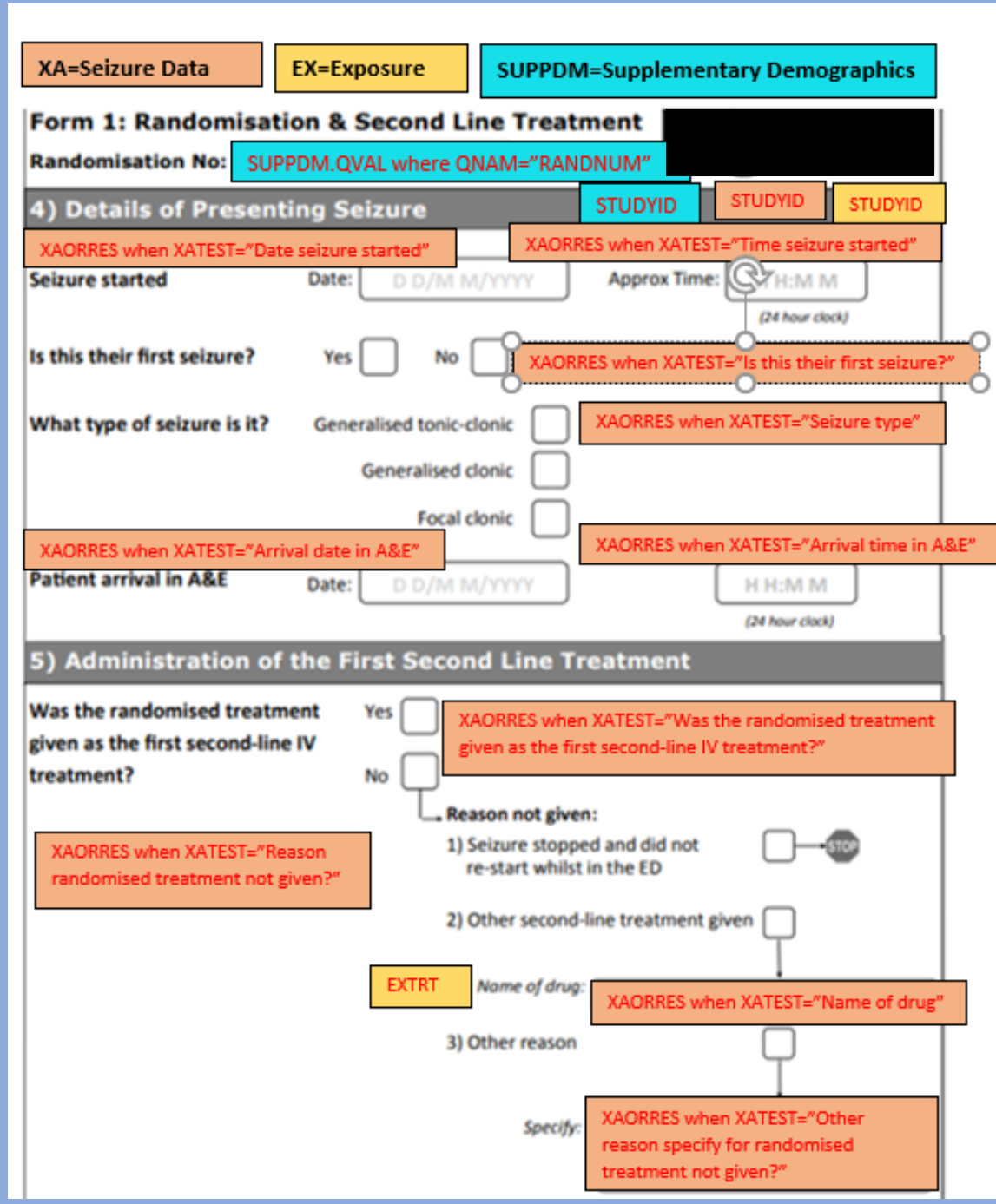
Example 2 aCRF: Unmappable Data & Extending Codelists

- A possible solution would be to extend the controlled terminology for PETESTCD's – Physical Examination Test Codes, as these are physical assessments
- We created these PETESTCDs with a limit of 8 characters, as per your own SOPs
- Another solution would have been to create an additional custom SDTM domain as shown on the next slide



Example 3 aCRF: Demonstrating Custom Domains

- In this example, we have seizure data which doesn't map to an existing SDTM domain
- As a result custom domains can be created as shown here
- These always begin with an X, such as XA, XB, XC.
- Here we define our custom dataset which we have called XA as seizure data, we could have called it XSZ for example which may be more reusable for other trials
  - Use of XSEIZURE may have created problems for longer variable names (e.g. XSEIZURETESTCD)
  - Use of a more descriptive names are helpful when storing in a local domain library



Problems Faced: SDTMs

- For many study specific collected data, it is unclear how these would map to an SDTM domain
- Ultimately, this would be a judgement call for where best to map the data which will likely result in the use of additional custom SDTM domains
- The CRF/CDMS might not be designed with CDISC in mind, meaning not all questions on a CRF will need to be mapped to SDTM domains
- One non-CDISC CRF often maps to several SDTM domains. This is likely to increase programming time when creating SDTM domains
- To complete required SDTM variables, the programmer is often required to impute data such as the Laboratory sequence variable (LBSEQ) and units not collected as a data item e.g. centimetres printed on paper CRF

Mapping tool

Prototype App

- Could we make a prototype of an application that would support the mapping of an existing data dictionary and CRF to SDTM?

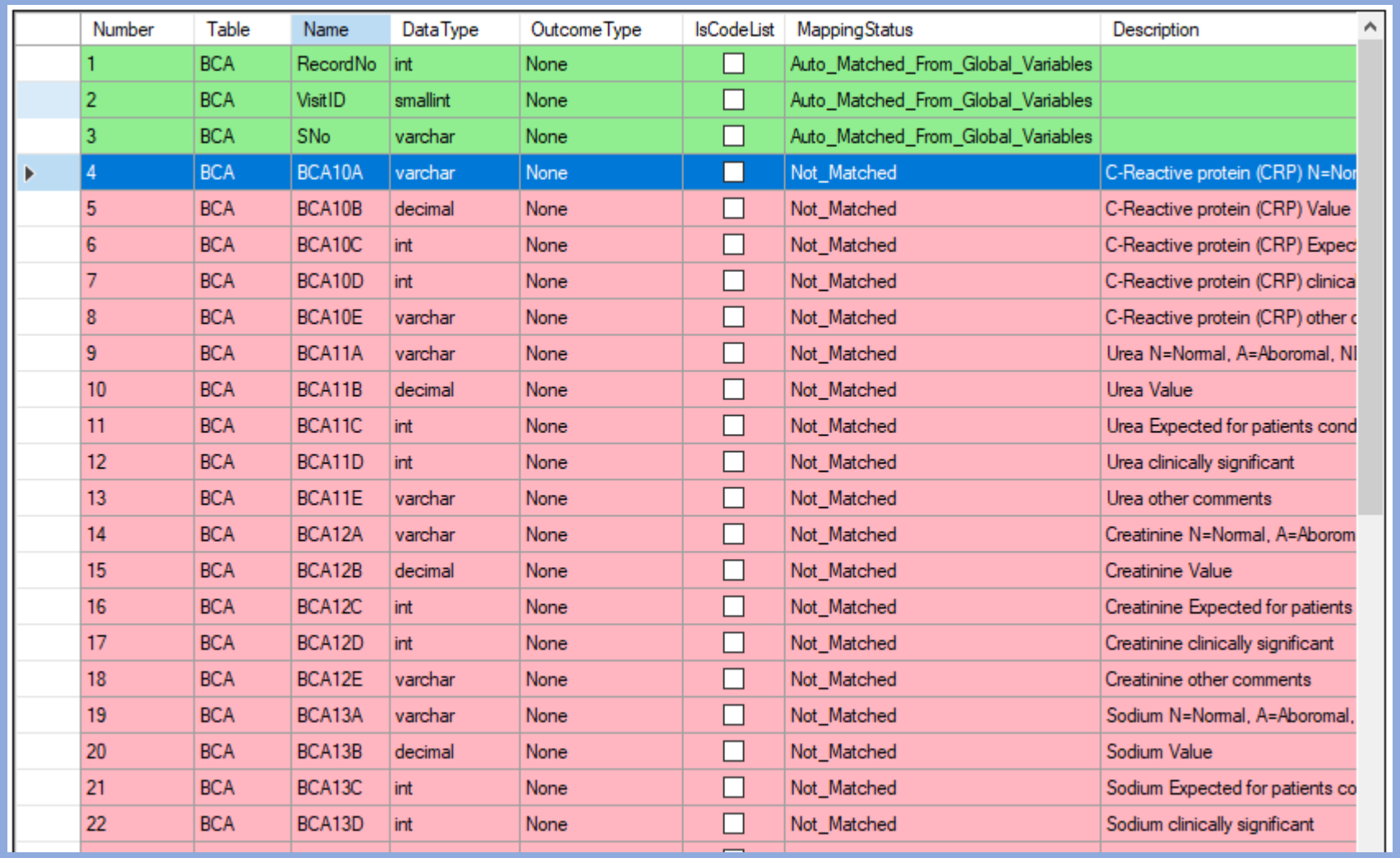
- Basic flow:

- Start with an unmapped data dictionary and blank CRFs
- Allow annotation of the CRFs from the data dictionary
- Map required fields to SDTM

Prototype App – Automatic Mapping

- Imagine removing some of the leg work, and automatically identifying mappings between the data dictionary and the SDTM
- Use the SDTM documents and definition as our semi structured text base, automatically map fields where possible
  - Even though Jonathan said “where possible”, he already had conquering visions of 75%+ auto mapping success – a GUI with a sea of green
- Can’t rely on specific organisational conventions
- “How hard could it be?” – Jonathan Gibb, pre-hair loss

Prototype App – Columbo Reveal



Number	Table	Name	DataType	OutcomeType	IsCodeList	MappingStatus	Description
1	BCA	RecordNo	int	None	☐	Auto_Matched_From_Global_Variables	
2	BCA	VisitID	smallint	None	☐	Auto_Matched_From_Global_Variables	
3	BCA	SNo	varchar	None	☐	Auto_Matched_From_Global_Variables	
4	BCA	BCA10A	varchar	None	☐	Not_Matched	C-Reactive protein (CRP) N=Nor
5	BCA	BCA10B	decimal	None	☐	Not_Matched	C-Reactive protein (CRP) Value
6	BCA	BCA10C	int	None	☐	Not_Matched	C-Reactive protein (CRP) Expec
7	BCA	BCA10D	int	None	☐	Not_Matched	C-Reactive protein (CRP) clinica
8	BCA	BCA10E	varchar	None	☐	Not_Matched	C-Reactive protein (CRP) other c
9	BCA	BCA11A	varchar	None	☐	Not_Matched	Urea N=Normal, A=Aboromal, NI
10	BCA	BCA11B	decimal	None	☐	Not_Matched	Urea Value
11	BCA	BCA11C	int	None	☐	Not_Matched	Urea Expected for patients cond
12	BCA	BCA11D	int	None	☐	Not_Matched	Urea clinically significant
13	BCA	BCA11E	varchar	None	☐	Not_Matched	Urea other comments
14	BCA	BCA12A	varchar	None	☐	Not_Matched	Creatinine N=Normal, A=Aborom
15	BCA	BCA12B	decimal	None	☐	Not_Matched	Creatinine Value
16	BCA	BCA12C	int	None	☐	Not_Matched	Creatinine Expected for patients
17	BCA	BCA12D	int	None	☐	Not_Matched	Creatinine clinically significant
18	BCA	BCA12E	varchar	None	☐	Not_Matched	Creatinine other comments
19	BCA	BCA13A	varchar	None	☐	Not_Matched	Sodium N=Normal, A=Aboromal,
20	BCA	BCA13B	decimal	None	☐	Not_Matched	Sodium Value
21	BCA	BCA13C	int	None	☐	Not_Matched	Sodium Expected for patients co
22	BCA	BCA13D	int	None	☐	Not_Matched	Sodium clinically significant

DEMO

Conclusions and Future projects

- Using CDISC standards at the data collection stage for clinical trials in the future could mean CRFs are designed to make the process more straightforward when mapping data items to SDTM datasets
- Could also lead to an understanding where unnecessary data isn't collected, i.e. where it cannot map to an SDTM dataset
- Potential to increase staff efficiency if CDISC standards were implemented, as the data would always be in the same format and standardised statistical programs could be used
- If data were standardised across clinical trials data sharing would become much more feasible

Thank you for listening, any questions?