



QC of the aCRF using SAS

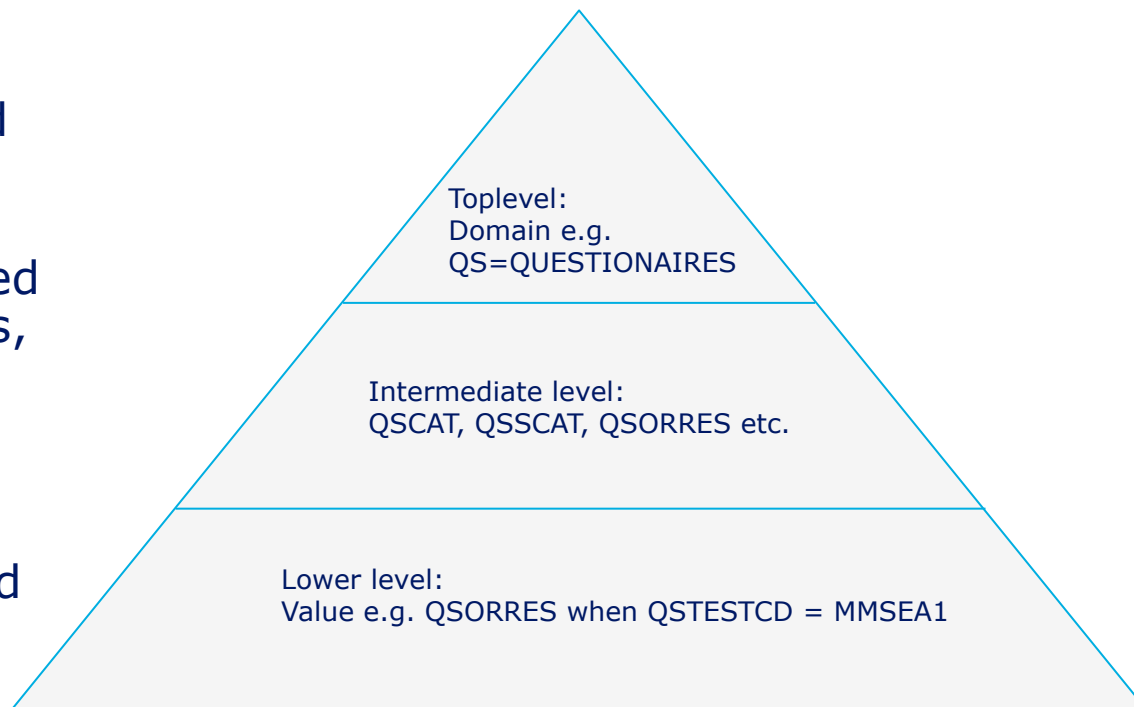
By Senior Consultant Rune Østergaard Pedersen and
Principal Consultant Niels Both

Outline

- The SDTM format
- The aCRF
- QC as a part of documentation for SDTM
 - ways to do QC
- Using the hierarchical structure of SDTM
- Examples using a sample aCRF
- Small examples on how to make the SAS-code
- Output tables
- Improvements

The SDTM format

- Has become the standard for submission of clinical trials.
- The SDTM data are divided into a number of datasets, each with a set of variables, and values within variables.
- The SDTM has a hierarchical structure, and this can be utilized when doing QC.

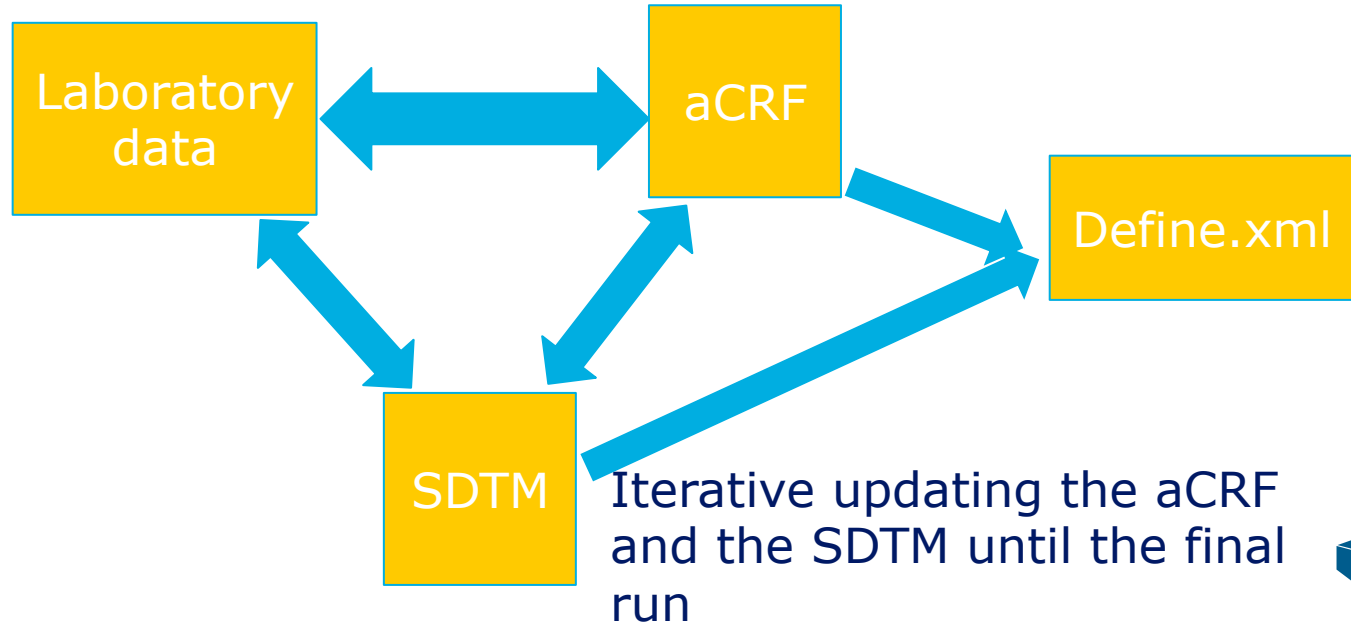


The aCRF

- The raw CRF is annotated to make a link between the SDTM data and the questions on the CRF.
- Often previous trials are applied as templates for annotations.
- The annotations serve as input to the define.xml, and should be correct.
- When designing the aCRF a number of rules may be applied to link the data (domains in SDTM).
 - Colours of background in annotation boxes.
 - Standardized ways of writing notes, supplemental qualifiers, relations and variable names etc.

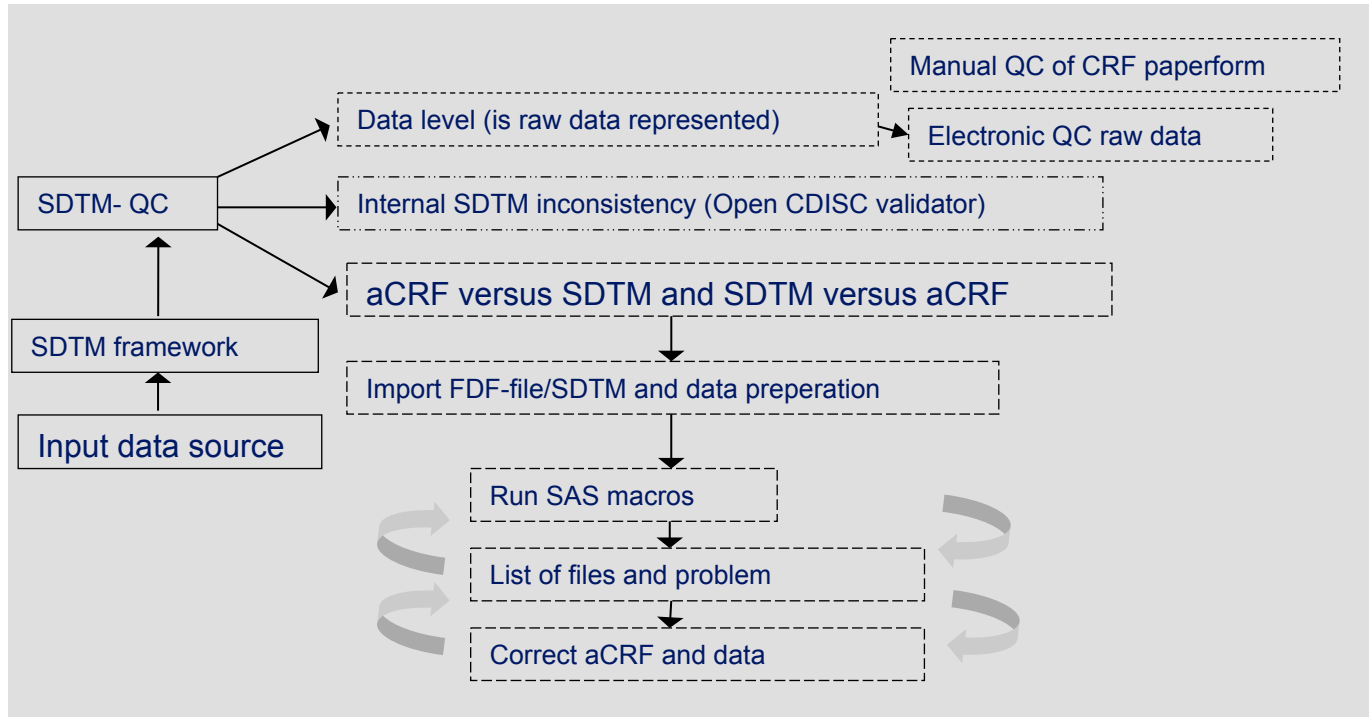
Macro-view of the ideal direct dataflow

Does not origin from aCRF,
only cosmetic like sample
dates



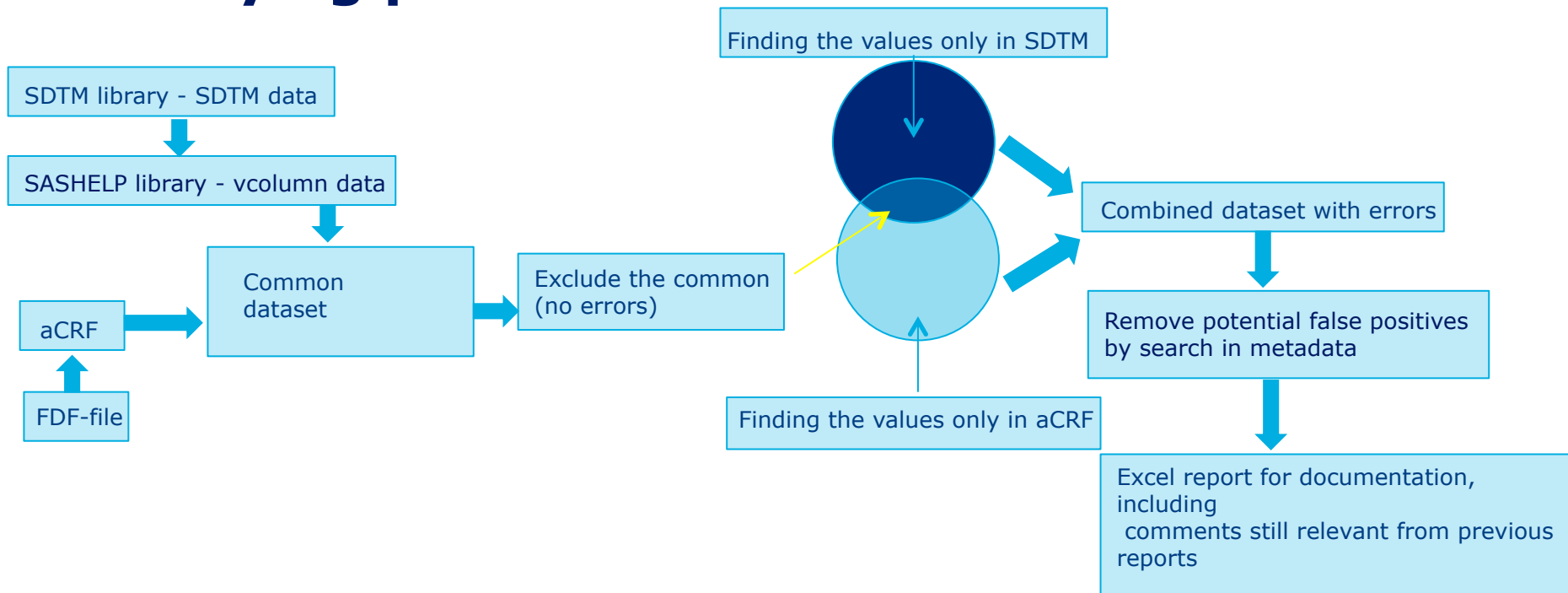
QC as a part of documentation for SDTM and aCRF

- Overview of the process



Level	Rule of inclusion	Rule of exclusion (data driven)
Domain	Grouping using colors of domain variable	
	On each page the domain should be mentioned e.g. AE=ADVERSE EVENTS	
Variable	The variable names are left of equal signs. In case of no equal sign the text is a variable. QNAM variables are separated with "-" and numbered successively like QNAM=AEACN1-AEACN6	Origin in the data that is not CRF e.g. EDC, could be adjudication data.
Value	Separation of values is done with ",". Certain words are used to identify variables. For example LBORRES when LBTESTCD=TOT_CARB. The "when" and "=" are identifiers.	
Notes and other labels	None	Notes are always initiated with NOTE:, this allows to remove the notes from the SAS dataset originating from the FDF-file. "Not Submitted" labels are excluded.
Replicated forms (paper forms)	Replications of the same forms on the aCRF are written AS PAGE x (where x indicates a number).	
RELREC	The domains in RELREC are identified on the aCRF as RELREC AE,DS. In case of more than two domains the list is expanded as RELREC ZZ,YY,XX	None

Identifying potential errors



- The SAS code starts by checking on the toplevel, moving to the intermediate level and finally the low level. Each level may be exemplified by a pseudocode. In general, terms a logical condition can be made in three loops:
- 1)%If not %sysfunc(exist(sdtm.var)) %then %do;
- %end
- 2)%else %do; /Intermediate level*/
- /*Making a macro variable containing the intermediate level variable*/
- /*If macro variable not in common dataset then output*/
- 3)/*else do*/
- /*Testing a variable on the low level*/
- /*If low level variable do not exist then output error in
- report*/
- /*end*/
- /end*/
- %end

Testing the toplevel

Testing the toplevel:

We make count if the domain exist on the aCRF

```
%if not %sysfunc(exist(sdtm.RELREC)) %then
  %do;

    data RELREConCRF_notDATA;
      LENGTH DOMAIN $ 20 problem $ 500 checkdesc $200;
      domain="RELREC";
      problem="Missing SDTM-RELREC domain";
      checkno=5;
      checkdesc="Find if RELREC domain is missing from SDTM data";
    run;

  %end;
```

Testing inside the low level/intermediate loops

```
Data mytest; /*Datacleaning step*/  
Set rawdata;  
Where myvar1="wish" and myvar2 ="" etc;  
Run;
```

```
Proc Sql noprint; /*Testing step*/  
Create table problems as  
Select b.var as var 1 "Problem" as problem  
From mytest where (a.key not in data_on_crf); quit;
```

Examples using a sample aCRF from CDISC

IE=Inclusion/Exclusion

CDISC Study			VISIT	Screening
CDISC01			Assessment Date:	IEDTC
ELIGIBILITY CRITERIA				
INCLUSION CRITERIA	IECAT	Should be IECAT=INCLUSION CRITERIA	Yes	No
Check the appropriate response				
IETEST				
1. Is age 18 - 85.		Will give false positive	<input type="checkbox"/>	<input type="checkbox"/>
2. Has Xyz disease of at least 10 weeks duration confirmed by biopsy			<input type="checkbox"/>	<input type="checkbox"/>
3. Did not respond to a standard course of medication ABC.			<input type="checkbox"/>	<input type="checkbox"/>

Wrong colour

IEORRES when IETESTCD = INCL01

IEORRES when IETESTCD = INCL02

IEORRES when IETESTCD = INCL03

DM=Demographics		Screening
CDISC Study: CDISC01		Assessment I
STUDYID		SCDTC ____/____/____
DEMOGRAPHY		
Date of Birth: ____/____/____		BRTHDTC
SEX		
Gender: <input type="checkbox"/> Male <input type="checkbox"/> Female		
ETHNIC		
Ethnicity: <input type="checkbox"/> Hispanic or Latino <input type="checkbox"/> Not Hispanic or Latino		
RACE		
Race: Check all that apply		
<input type="checkbox"/> White		
<input type="checkbox"/> American Indian or Alaska Native		RACE1-RACE5 in SUPPDM
<input type="checkbox"/> Black or African American		
<input type="checkbox"/> Native Hawaiian or Other Pacific Islander		
<input type="checkbox"/> Asian		
<input type="checkbox"/> Other: __		RACEOTH in SUPPDM

To help the reader we would add:
 NOTE: RACE, when more than one selected, RACE=MULTIPLE and individual responses are RACE1, RACE2, etc. in SUPPDM

False positive error, only white, black or african american and Asian exist.

Missing domain annotation on top of continuing pages will give multiple errors in the program

Should be MHCAT=MEDICAL

CDISC Study CDISC01	SCREENING Assessment Date: MHDTC		
MEDICAL AND SURGICAL HISTORY MHCAT			
Does the subject have any significant medical or surgical history? [NOT SUBMITTED] <input type="checkbox"/> Yes, list the condition(s) below <input type="checkbox"/> No	Year MHSTDTC	“√” if RESOLVED MHENRF = BEFORE	“√” if ONGOING MHENRF = DURING/AFTER
MHTERM	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		

False error due to noone using the supp-domain

EGTESTCD = INTP

OVERALL INTERPRETATION (Please check one):

1 = Normal (do not comment) **EGORRES = NORMAL**

2 = Abnormal, not clinically significant (do not comment) **EGORRES = ABNORMAL**

EGCLSIG=N in SUPPEG

3 = Abnormal, clinically significant. Specify and comment: **EGORRES = ABNORMAL**
Comments [char(200)]

EGCLSIG=Y in SUPPEG

EGCLSP in SUPPEG

Potential error in EGCLSP needs investigation in raw data.

DS=Disposition

CDISC Study CDISC01	Assessment Date: ____/____/____ [NOT SUBMITTED]
------------------------	-----------------------------------------------------------

TERMINATION

Did patient complete the study? Yes No

DSDECOD / DSTERM = COMPLETED when Yes

If patient did not complete the study, indicate the date of termination and check one primary reason to indicate why:

Date of Termination: / / **DSSTDTC**

DSDECOD Patient did not meet Inclusion/Exclusion Criteria at Screening or baseline (specify): **DSTERM**

Discontinued due to lack of Therapeutic Response **DSDECOD**

Discontinued due to Adverse Event
Adverse Event No. _____ (Enter the number from the ADVERSE EVENTS Form)

Missing RELREC: AE,DS

Example of a report (modified from Excel)

Check No	Check Description	Category	Problem	Annotation	Domain	Page	Color	Variable	Value	Metadata	Comment
1	Find if a variable has been annotated on a page, but with no matching domain annotation. This could be caused by color differences.	aCRF Annotation Problem	No domain annotated on aCRF	IEDTC		4	C[0.749 023 1.0 1.0]	IEDTC			A color difference has been removed
2	Find if a variable with value (eg XXTESTCD=TEST) is annotated on the aCRF, but that value does not exist in the SDTM data		IETESTCD present but no value of INCL01 exists in data	IEORRES when IETESTCD=INCL01	IE	4		IETESTCD	INCL01		The inclusion criteria was never violated, hence the value is not in SDTM

Example of a report continued.

Check No	Check Description	Category	Problem	Annotation	Domain	Page	Colour	Variable	Value	Metadatas	Comment
3	Find if a supplemental qualifier variable has been annotated on the aCRF, but is not present in data	Annotated on aCRF, but not in SDTM data	QNAM annotated on aCRF, but not present in supp qual data	RACE1 - RACE5 in SUPPDM	SUPPDM	6		QNAM	RACE4		The Native Hawaiian or Other Pacific Islander is never found in the raw data
4	Find if a variable has been annotated on a page, but with no matching domain annotation. This could be caused by color differences.	aCRF Annotation Problem	No domain annotated on aCRF	MHDT C		7	C[0.75 1 1]				The error is due to missing domain annotation

Example of a report continued.

Check No	Check Description	Category	Problem	Annotation	Domain	Page	Colour	Variable	Value	Metadata	Comment
5	Find if a RELREC domain combination that exists in the RELREC SDTM domain has not been annotated on the aCRF	In SDTM data, but not annotated on aCRF	No annotation on aCRF for domains linked in RELREC		RELREC			DS,AE			The annotation RELREC: AE.DS has been added to the aCRF
6	Find if a supplemental qualifier variable has been annotated on the aCRF, but is not present in data	Annotated on aCRF, but not in SDTM data	QNAM annotated on aCRF, but not present in supp qual data	EGCLSP in SUPPEG	SUPPEG	12		EGCLSP			After a look in raw data it looks like noone entered a comment, hence OK

Explanations of the output variables

- **Check No** – Refers to the list of checks used by the SAS macros
- **Check Description** – A short description of the problem/error. For example: Find if a variable has been annotated on a page, but with no matching domain annotation. This could be caused by colour differences.
- **Category** – The program uses one of the categories:
 - 1) aCRF Annotation Problem.
 - 2) Annotated on aCRF, but not in SDTM data.
 - 3) In SDTM data, but not annotated on aCRF.
- **Problem** – this is a more fine description of the problem. For example: No domain annotated on aCRF or QNAM annotated on aCRF, but not present in supp qual data.
- **Page** – the pagenumber of the aCRF that the annotation originates from

:

- **Annotation** – This is a variable containing the original annotation on the aCRF. Is only relevant if the problem is missing data for annotations present on the aCRF.
- **Domain** – the two letter abbreviation or supplemental qualifier abbreviation. E.g. LB and SUPPLB that the data/annotation is linking to.
- **Colour** – this is only relevant for colour issues on the aCRF and refers to the colour code that are under investigation.
- **Variable** – the variable in SDTM. However, in supplemental qualifier it may refer to a qnam.
- **Value** – this is a value of a SDTM variable e.g. a testcd for lbtestcd.
- **Comment** – these are comments used to justify the error, and may be applied for documentation of the QC-process of SDTM. Comments not changed, but still relevant are saved from one iteration of the program to the next.

Conclusion

- The QC program has proven as an important tool, enabling identification of many errors that cannot easily be found manually.
- The complexity of the dataflow often gives a number of false positives, which makes manual QC needed.
- The QC program can easily be applied as a part of the documentation for authorities.
- The way of creating annotation that can be applied to QC on PC have the positive side-effect of being more systematic.
- New employees often benefit from the errors detected by the PC, making their SDTM learning curve less steep.