



STREAMLINE SDTM QC

STEPHEN GORMLEY

BIostatistical PROGRAMMING SENIOR MANAGER

PHUSE 2017



- **Background:**

- **Two Main Problems:**

- Duplicate Specifications
- Dual Programming

- **Four Significant Issues**

- **Decision**

- **Three Step Solution:**

- Centralized, Risk Based and Automated.



BACKGROUND

SUBMISSION DATA FILE (“SDF”)

- An in-house built Java ETL tool use to produce SDTM.
- Global Library of Metadata that describes each SDTM.
 - **SDF Global Library Templates** include for example:
 - source to target mapping, target formats, QC instructions, mapping type, functions required, controlled terminology, extended submission metadata (e.g. computational method, origin, core variable status, role).
- Metadata for each SDTM IG version and each SDTM Amgen version.
- Cloned to the study level as **SDF Study Level Templates**.
- SDF Global Library Templates=SDF Study Level Templates with a link maintained to check for compliance.



EXAMPLE SDF TEMPLATE

Column Editor For DataFile: CM_CONMEDS_1

Column Details

Column Attributes

Name: CMDECOD

Data Type: CHAR

Length(Bytes): 200

Label: Standardized Medication Name

Format Name:

Default Mapping Type: Single Mapping

Data Source: CDMS

Default Mapping: CONMEDS_ALL.CMPT

Default Function:

Status: Active

Meta Data: ☐

CDISC Attributes

CDISC Variable Type: CDISC

CDISC Data Type: Text

Display Length: 200

Decimal Precision: 0

Role: Synonym Qualifier

Core Variable Status: Permissible

Case Sensitivity: Mixed

Computational Method:

DES Versioned: ☒

SDTMIG Versioned: ☒

Origin

Origin Document ID Location

Comments:

Notes: Standardized or dictionary-derived text description of CMTRT or CMMODIFY. Equivalent to the generic medication name in WHO Drug. The sponsor should specify the dictionary name and version in the Sponsor Comments column of the Define document. If an intervention term does not have a decode value in the dictionary then CMDECOD will be left blank.

QC Instructions:

SPECIFICATION AND QC

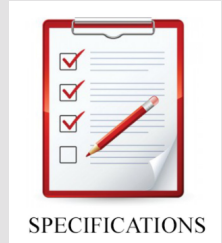


QC

- **Verification:** the SDTM has been programmed correctly (i.e. the SDTM is programmed in accordance with the specification).
- **Validation:** the SDTM is programmed as expected (i.e. the SDTM is programmed in accordance with the expectation of the consumer).

SPECIFICATION

- **Precise:** tell the “programmers” what they need to know.
- **Correct:** describes what is really needed.
- **Clear:** easy to grasp and to understand without error for its intended users.



CCI AND STUDY TEAM

- **The CCI Team:**

- The CDISC Consultancy and Implementation team.
- Produce all SDTM datasets and a number of associated submission deliverables .

- **The Study Team:**

- The clinical study statistical programming team.
- Until the middle of 2016, performed the Quality Control of the SDTM datasets.

TWO MAIN PROBLEMS

PROBLEM ONE

Global Statistical Programming (GSP)

GSP CCI Team (Produce SDTM)

Raw Data: CTDB and Non-CTDB

Specification(s):

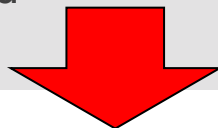
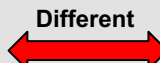
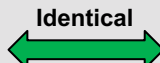
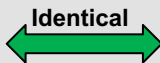
- Protocol, SAP, Annotated CRF
- SDF Global Library and Study Level Templates
- GSP CCI Team SDF Controlled Terminology

GSP Study Team (QC SDTM)

Raw Data: CTDB and Non-CTDB

Specification(s):

- Protocol, SAP, Annotated CRF
- MDDT for SDTM QC in Excel
- GSP Study Team Excel Controlled Terminology



Dual Creation and Maintenance of Specifications

PROBLEM TWO

GSP CCI Team (Produce SDTM)

Raw Data: CTDB and Non-CTDB

Specification(s):

- Protocol, SAP, Annotated CRF
- SDF Global Library and Study Level Templates
- Controlled Terminology: SDF

Code: Generated from SDF (e.g. PROC COMPARE
ae.sas)

Output: ANSDTM



GSP Study Team (QC SDTM)

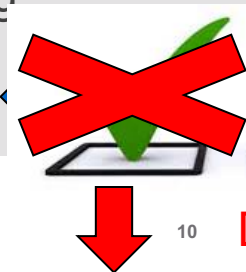
Raw Data: CTDB and Non-CTDB

Specification(s):

- Protocol, SAP, Annotated CRF
- MDDT for SDTM QC in Excel
- Controlled Terminology: Excel

Code: Independent program (e.g. v-
ae.sas)

Output: QCANSDTM



10

Dual Programming

FOUR SIGNIFICANT ISSUES

FOUR SIGNIFICANT ISSUES

1. The requirement on all of GSP (not just The CCI Team) to be experts in SDTM.
2. Ambiguity in the accountability and decision making on SDTM.
3. A large number of false-positives and/or minor differences that are raised when using independent programming in SAS with PROC COMPARE.
4. A large duplication of resource effort.



OPTIONS AND DECISION

OPTIONS



Four Options Considered:

- **Option 1:** Keep the status quo, although there are a number of problems Amgen do produce a high quality SDTM package.
- **Option 2:** Outsource SDTM production and QC completely to a CRO.
- **Option 3:** The Study Teams both produce and QC SDTM.
- **Option 4:** The CCI Team both produce and QC SDTM.
 - Why?
 - One Amgen team more efficient, SDF Global Library already exists and already a wealth of CDISC SDTM expertise within the CCI Team.

DECISION

Three key decisions were made:

1. No longer QC using 100% independent programming of each variable in every SDTM.
2. Introduce a more automated and calculated risk associated with SDTM dataset and variable QC.
3. Centralise the process within The CCI Team.



GSP CCI Team

- Produce and QC SDTM
- Maintain one set of SDTM specifications
- Clear owner and makes the final decision

GSP Study Team

- Review default QC methods and specifications
- Advise on criticality of QC
- Advise on SDTM specifications

THREE STEP SOLUTION

STEP 1: CENTRALISE THE SPECIFICATION

- Within the SDF Global and Study Level Templates.
 - For review, an Excel File (an interpretation of the SDF Study Level Templates) is created containing:
 - the full specification, the QC level to be assigned to each variable for review, extended CDISC metadata (and much more).
 - the assigned QC level.
 - a flag to highlight variables that require further study team input
- Note:** Simple source to target mappings and/or default functions would **not** generally require review.
- A review workflow is tracked and issues are documented in SharePoint.

STEP 1: CENTRALISE THE SPECIFICATION

Column Editor For DataFile: CM_CONMEDS_1

Column Details

Column Attributes

Name: CMSTDY

Data Type: NUM

Length(Bytes): 8

Label: Study Day of Start of Medication

Format Name:

Default Mapping Type: Single Mapping

DataSource: Self

Default Mapping: CMSTDTC

Default Function: 3601da.) ELSE . END ...

Status: Active

Meta Data: ☐

CDISC Attributes

CDISC Variable Type: CDISC

CDISC Data Type: Integer

Display Length: 8

Decimal Precision: 0

Role: Timing

Core Variable Status: Permissible

Case Sensitivity:

Computational Method:

DES Versioned: ☒

SDTMIG Versioned: ☒

Origin:

Origin	Document ID	Location

Comments

Derived per SDTMIG assumption 4.1.4.4. If partial date or null date, then null.

Notes

Study day of start of medication relative to the sponsor- defined RFSTDTC.

QC Instructions

```
macro:g04_dtcdy
parm: raw_dset=vcdms.conmeds_all
parm: raw_datvar=CMSTDTC_raw
parm: sdtn_dtvvar=CMSTDTC
parm: filter= DS_SRCE = 'CONMEDS_1'
and CMSTDTC_RAW ne 'UN UNK 0000'
```

Requirement

If pre-IP (<0), then CM.CMSTDTC - DM.RFSTDTC. If on/after IP (>=0), then CM.CMSTDTC - DM.RFSTDTC + 1 if partial date or null date, then null <+overwrite>

STEP 1: CENTRALISE THE SPECIFICATION

Click for SDTM Table of Contents							
Column Order	Variable Name	Variable Label	Origin	Method Type (Document)	Source Derivation	DefineXML Comment	Implementation Notes
46	AETOXGR	Standard Toxicity Grade	CRF		TBA	CTCAE version 4.03	Toxicity grade according to a standard toxicity scale such as Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Sponsor should specify scale and version used in the Comment column of the Define data definition document. Value should contain valid numbers only; datatype will be changed to numeric in future version of SDTM.
50	EPOCH	Epoch			TBA	See CRT Reviewer's Guide for EPOCH derivation details.	
51	AESTDTC	Start Date/Time of Adverse Event	CRF		TBA		
52	AEENDTC	End Date/Time of Adverse Event	CRF		TBA		Date when the event ended, worsened in severity or resulted in death. If continuing or unknown, then date is null.
53	AESTDY	Study Day of Start of Adverse Event	Derived	Computation	TBA		Study day of start of adverse event relative to the sponsor-defined RFSTDTC.
54	AEENDY	Study Day of End of Adverse Event	Derived	Computation	TBA		Study day of end of event relative to the sponsor-defined RFSTDTC.
55	AESEV	Severity of Adverse Event	CRF		TBA		Severity of adverse event according to the sponsor-defined scale.

STEP 2: A MORE RISK BASED APPROACH TO QC



Four Types of QC

1. **Basic:** Metadata Checks

- **Variable Types:** Low risk, direct mappings from source to target, standard function conversions.
- **Example Checks:** Study Level Templates vs SDF Global Library Templates, Submission Deliverables, Analysis Level Checks.

2. **Basic:** Data Checks

- **Variable Types:** Low risk, direct mappings from source to target, standard function conversions.
- **Example Checks:** Log checks, CT, uppercase, ranges, value level metadata, date errors, blank values (including Pinnacle 21)

STEP 2: A MORE RISK BASED APPROACH TO QC



Four Types of QC

3. **Complex and Generic:** SAS with PROC COMPARE

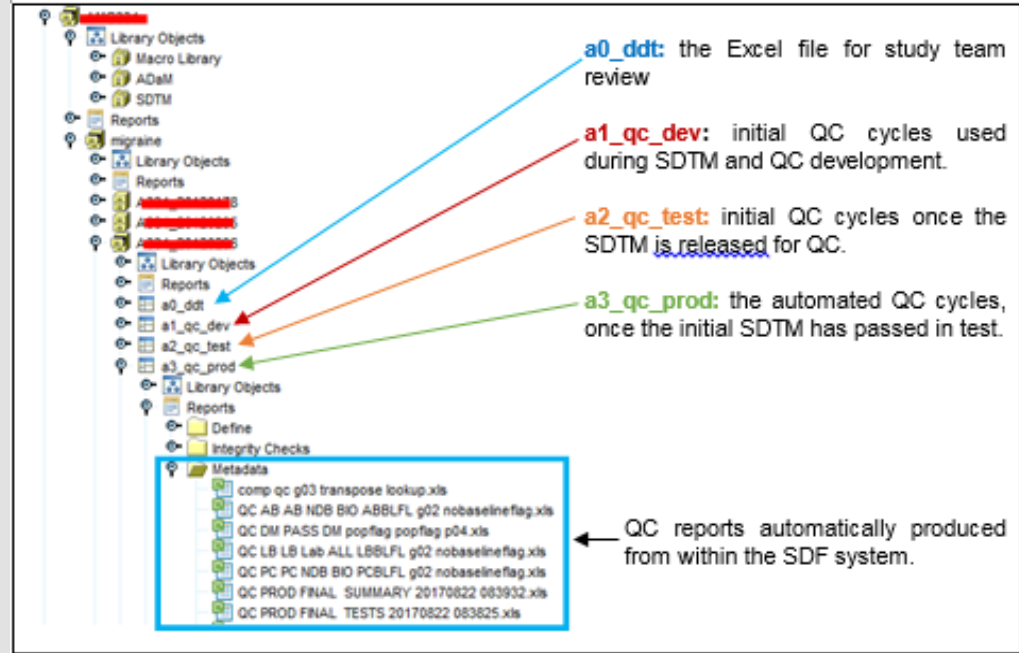
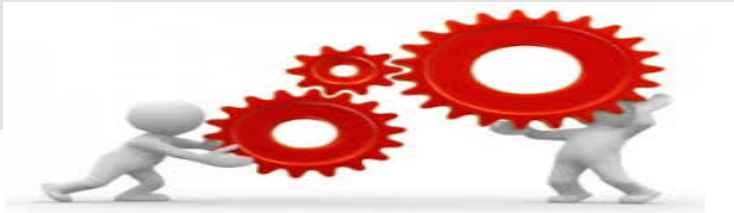
- **Variable Type:** High Risk, Critical and/or Complex.
- **Example Variables :** General SDTM algorithms, e.g. XXBLFL, RFSTDTC, EPOCH, transpose, observation checking.

4. **Complex and Study Specific:** SAS with PROC COMPARE

- **Variable Type:** High Risk, Critical and/or Complex.
- **Example Variables:** Study specific independent programs, e.g. EXDOSE, EXTRT, ARMCD, EPOCH, POPFLAGS, XXSTRESC (where=primary endpoint)

STEP 3: AUTOMATE THE QC

- Automation within SDF.
- QC scheduler in a3_qc_prod
 - RAW > SDTM > QC
- QC reports automatically generated
- QC documentation created from the compliance reports.
- + Pinnacle 21 checks.



CONCLUSION

SUMMARY

- **Background:**
 - CCI and Study Team, SDF, Specification, QC: Verification or Validation
- **Two Main Problems:**
 - Duplicate Specifications
 - Dual Programming
- **Four Significant Issues**
- **Decision**
- **Three Step Solution:**
 - Centralised, Risk Based and Automated.



QUESTIONS?

4	=	21
5	=	27
6	=	36
7	=	49
8	=	57
9	=	76
10	=	?



*"Yes, I think I have good people skills.
What kind of idiot question is that?"*