



Lost in Discrepancies

How to facilitate programmatic comparison

Judith Kanz and Fenja Seither



www.mainanalytics.de



PHUSE SDE Heidelberg 2024

Table of Contents



Background



Preparations



Untangle PROC COMPARE output



Follow-up work

Background

Validation of analysis datasets and statistical outputs is of the utmost importance



Automatic comparison of output files is difficult



Idea: Double programming and programmatic compare of datasets used in output creating procedure

Example output

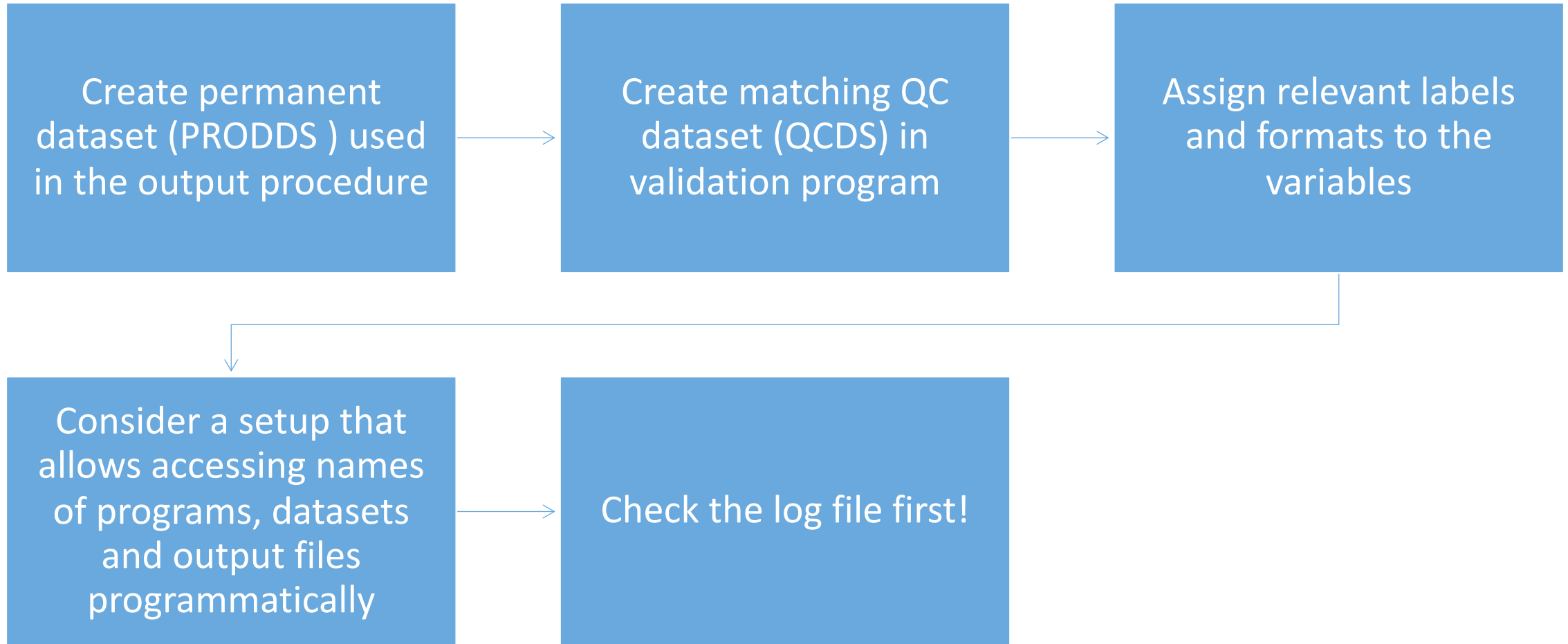
mainanalytics GmbH

Table 6.1: Adverse event frequencies

System organ class/ Preferred term	Placebo					Total Xanomeline				
	N	%	95% CI	Time at risk [patient years]	Incidence rate/100 patient years	N	%	95% CI	Time at risk [patient years]	Incidence rate/100 patient years
Number of patients	86	100.0		35		168	100.0		46	
Number of patients with TEAE	65	75.6	[65.5, 83.4]	15	447	153	91.1	[85.8, 94.5]	12	1242
CARDIAC DISORDERS	12	14.0	[8.2, 22.8]	33	37	28	16.7	[11.8, 23.0]	42	67
ATRIAL FIBRILLATION	1	1.2	[0.2, 6.3]	35	3	4	2.4	[0.9, 6.0]	46	9
ATRIAL FLUTTER	0	0.0	[0.0, 4.3]	35	0	2	1.2	[0.3, 4.2]	46	4
ATRIAL HYPERTROPHY	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
ATRIOVENTRICULAR BLOCK FIRST DEGREE	1	1.2	[0.2, 6.3]	35	3	1	0.6	[0.1, 3.3]	45	2
ATRIOVENTRICULAR BLOCK SECOND DEGREE	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
BRADYCARDIA	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
BUNDLE BRANCH BLOCK LEFT	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
BUNDLE BRANCH BLOCK RIGHT	1	1.2	[0.2, 6.3]	35	3	1	0.6	[0.1, 3.3]	46	2
CARDIAC DISORDER	0	0.0	[0.0, 4.3]	35	0	1	0.6	[0.1, 3.3]	46	2
CARDIAC FAILURE CONGESTIVE	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
MYOCARDIAL INFARCTION	4	4.7	[1.8, 11.4]	35	12	6	3.6	[1.6, 7.6]	44	14
PALPITATIONS	0	0.0	[0.0, 4.3]	35	0	2	1.2	[0.3, 4.2]	46	4
SINUS ARRHYTHMIA	1	1.2	[0.2, 6.3]	35	3	0	0.0	[0.0, 2.2]	46	0
SINUS BRADYCARDIA	2	2.3	[0.6, 8.1]	35	6	15	8.9	[5.5, 14.2]	44	24

Source data: CDISC Pilot

Preparations



Does PRODDS contain ID variables that can be used for PROC COMPARE?

Are all variables in PRODDS needed to produce the output?

Are indentations/spaces relevant for PROC COMPARE?

Considerations

How many digits are shown in the output?

Are variable formats used in the output?

Are variable labels displayed in the output?

Supporting macro

```
%macro qc_compare(outid      =      /* Output ID */
                    ,qcds     = out  /* Name of QC dataset */
                    ,key       =      /* Variable names for PROC COMPARE ID (default: none) */
                    ,keep      =      /* Variable names taken into account for PROC COMPARE (default: all) */
                    ,indent    = N    /* Check indentations (N/Y) */
                    ,compress  =      /* Variable names to be compressed before PROC COMPARE (default: none) */
                    ,criterion = 0.000001 /* Value of CRITERION option used for PROC COMPARE */
                    ,labels    = N    /* Check labels (N/Y) */
                    ,formats   = N    /* Check formats (N/Y) */
);
```

- Checks if PRODDS and output file exist
- Performs PROC COMPARE
- Creates supporting temporary datasets
- Creates log file messages in case of discrepancies

Step 1: Align variables

Listing of Variables in WORK.MCCOMP_PRODDS but not in WORK.MCCOMP_QCDS

Variable	Type	Length
----------	------	--------

ORDER	Num	8
-------	-----	---

Listing of Common Variables with Conflicting Types

Variable	Dataset	Type	Length	Label
----------	---------	------	--------	-------

PLACEBO_PERC	WORK.MCCOMP_PRODDS	Num	8	%
	WORK.MCCOMP_QCDS	Char	5	%
XANOMELINE_PERC	WORK.MCCOMP_PRODDS	Num	8	%
	WORK.MCCOMP_QCDS	Char	5	%

Listing of Common Variables with Differing Attributes

Variable	Dataset	Type	Length	Label
----------	---------	------	--------	-------

TEXT	WORK.MCCOMP_PRODDS	Char	200	System organ class/\$	Preferred term
	WORK.MCCOMP_QCDS	Char	250	System Organ Class/\$	Preferred Term

Step 2: Compare ID variables (1)

Comparison Results for Observations

Observation 3 in WORK.MCCOMP_PRODDS not found in WORK.MCCOMP_QCDS:
AESOC=CARDIAC DISORDERS AEDECOD=.

Observation 6 in WORK.MCCOMP_QCDS not found in WORK.MCCOMP_PRODDS:
AESOC=CARDIAC DISORDERS AEDECOD=ATRIOVENTRICULAR BLOCK FIRST DEGREE.

Observation 7 in WORK.MCCOMP_PRODDS not found in WORK.MCCOMP_QCDS:
AESOC=CARDIAC DISORDERS AEDECOD=ATRIOVENTRICULAR BLOCK FIRST\$ DEGREE.

Observation 7 in WORK.MCCOMP_QCDS not found in WORK.MCCOMP_PRODDS:
AESOC=CARDIAC DISORDERS AEDECOD=ATRIOVENTRICULAR BLOCK SECOND DEGREE.

Observation 8 in WORK.MCCOMP_PRODDS not found in WORK.MCCOMP_QCDS:
AESOC=CARDIAC DISORDERS AEDECOD=ATRIOVENTRICULAR BLOCK\$ SECOND DEGREE.

Step 2: Compare ID variables (2)



Temporary dataset 1: MCCOMP_CHECKKEY

- ID variables only
- Flag variables PROD and QC
- All records from both datasets

Step 2: Compare ID variables (3)

WORK.MCCOMP_CHECKKEY

View: Column names

Filter: prod ne qc and aeDecod eq ""

Total rows: 296 Total columns: 4 Filtered rows: 46

	AESOC	AEDECOD	PROD	QC
1	Cardiac Disorders		.	1
2	CARDIAC DISORDERS		1	.
3	Congenital, Familial And Genetic Disorders		.	1
4	CONGENITAL, FAMILIAL AND GENETIC DISORDERS		1	.
5	Ear And Labyrinth Disorders		.	1
6	EAR AND LABYRINTH DISORDERS		1	.
7	Eye Disorders		.	1
8	EYE DISORDERS		1	.
9	Gastrointestinal Disorders		.	1
10	GASTROINTESTINAL DISORDERS		1	.
11	General Disorders And Administration Site Conditions		.	1
12	GENERAL DISORDERS AND ADMINISTRATION SITE CONDITIONS		1	.
13	Hepatobiliary Disorders		.	1

Step 2: Compare ID variables (4)

WORK.MCCOMP_CHECKKEY x

View: Column names | Filter: prod ne qc and aeecod ne " " x

Total rows: 171 Total columns: 4 Filtered rows: 26

	AESOC	AEDECOD	PROD	QC
1	CARDIAC DISORDERS	ATRIOVENTRICULAR BLOCK FIRST DEGREE	.	1
2	CARDIAC DISORDERS	ATRIOVENTRICULAR BLOCK FIRST\$ DEGREE	1	.
3	CARDIAC DISORDERS	ATRIOVENTRICULAR BLOCK SECOND DEGREE	.	1
4	CARDIAC DISORDERS	ATRIOVENTRICULAR BLOCK\$ SECOND DEGREE	1	.
5	CARDIAC DISORDERS	SUPRAVENTRICULAR EXTRASYSTOLES	.	1
6	CARDIAC DISORDERS	SUPRAVENTRICULAR\$ EXTRASYSTOLES	1	.
7	CARDIAC DISORDERS	WOLFF-PARKINSON-WHITE SYNDROME	.	1
8	CARDIAC DISORDERS	WOLFF-PARKINSON-WHITE\$ SYNDROME	1	.
9	GASTROINTESTINAL DISORDERS	GASTROOESOPHAGEAL REFLUX DISEASE	.	1
10	GASTROINTESTINAL DISORDERS	GASTROOESOPHAGEAL REFLUX\$ DISEASE	1	.
11	INFECTIONS AND INFESTATIONS	LOWER RESPIRATORY TRACT INFECTION	.	1
12	INFECTIONS AND INFESTATIONS	LOWER RESPIRATORY TRACT\$ INFECTION	1	.

Step 3: Compare non-ID variables (1)

Value Comparison Results for Variables

		95% CI	
		Base Value	Compare Value
AESOC	AEDECOD	PLACEBO_CI	PLACEBO_CI
ANY EVENT		[65.5, 83.4]	[65.5, 83.4]
CARDIAC DISORDERS		[8.2, 22.8]	[8.2, 22.8]
CARDIAC DISORDERS	ATRIAL FIBRILLATION	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	ATRIAL FLUTTER	[0.0, 4.3]	[0.0, 4.3]
CARDIAC DISORDERS	ATRIAL HYPERTROPHY	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	ATRIOVENTRICULAR BLO	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	ATRIOVENTRICULAR BLO	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	BRADYCARDIA	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	BUNDLE BRANCH BLOCK	[0.2, 6.3]	[0.2, 6.3]
CARDIAC DISORDERS	BUNDLE BRANCH BLOCK	[0.2, 6.3]	[0.2, 6.3]

Step 3: Compare non-ID variables (2)

Variables with Unequal Values

Variable	Type	Len	Label	Ndif	MaxDif
XANOMELINE_TRISK	NUM	8	Time at risk [patient years]	13	0.0082
XANOMELINE_RATE	NUM	8	Incidence rate/100 patient years	13	0.827

Value Comparison Results for Variables

AESOC	AEDECOD	Time at risk [patient years]			
		Base		Diff.	% Diff
		XANOMELIN E_TRISK	Compare XANOMELIN E_TRISK		
ANY EVENT		12.3231	12.3313	0.008214	0.0667
GASTROINTESTINAL DIS		38.1875	38.1958	0.008214	0.0215
GASTROINTESTINAL DIS	NAUSEA	44.0356	44.0438	0.008214	0.0187
GASTROINTESTINAL DIS	SALIVARY HYPERSECRET	44.3669	44.3751	0.008214	0.0185
GASTROINTESTINAL DIS	VOMITING	43.1513	43.1595	0.008214	0.0190
NERVOUS SYSTEM DISOR		38.9268	38.9350	0.008214	0.0211
NERVOUS SYSTEM DISOR	PAROSMIA	45.3689	45.3771	0.008214	0.0181
RESPIRATORY, THORACI		42.1081	42.1164	0.008214	0.0195
RESPIRATORY, THORACI	COUGH	43.3265	43.3347	0.008214	0.0190
RESPIRATORY, THORACI	NASAL CONGESTION	44.5120	44.5202	0.008214	0.0185
SKIN AND SUBCUTANEOU		29.1444	29.1526	0.008214	0.0282
SKIN AND SUBCUTANEOU	ERYTHEMA	38.5599	38.5681	0.008214	0.0213
SKIN AND SUBCUTANEOU	PRURITUS	37.2293	37.2375	0.008214	0.0221

Step 3: Compare non-ID variables (3)



Temporary dataset 2: MCCOMP_CHECKVARS

- All records from both datasets
- ID variables
- non-ID variables with prefixes PROD_ and QC_ side by side
- Flag variables PROD and QC
- Flag variable DIFF_PROD_QC

Step 4: Identify source (1)

WORK.MCCOMP_CHECKVARS

View: Column names Filter: diff_prod_qc eq 1 and aeecod ne " "

Total rows: 255 Total columns: 29 Filtered rows: 8

Columns	AEDECOD	PROD_XANOMELINE_FREQ	PROD_XANOMELINE_TRISK	QC_XANOMELINE_TRISK	PROD	QC	DIFF_PROD_QC
Select all							
☐ AESOC	1 NAUSEA	9	44.03559206	44.043805613	1	1	1
☒ AEDECOD	2 SALIVARY HYPERSECRETION	4	44.366872005	44.375085558	1	1	1
☐ PROD_ORDER	3 VOMITING	10	43.151266256	43.159479808	1	1	1
☐ QC_ORDER	4 PAROSMIA	1	45.368925394	45.377138946	1	1	1
☐ PROD_TEXT	5 COUGH	10	43.326488706	43.334702259	1	1	1
☐ QC_TEXT	6 NASAL CONGESTION	4	44.511978097	44.52019165	1	1	1
☐ PROD_PLACEBO_FREQ	7 ERYTHEMA	28	38.559890486	38.568104038	1	1	1
☐ QC_PLACEBO_FREQ	8 PRURITUS	47	37.229295003	37.237508556	1	1	1
☒ PROD_XANOMELINE_FREQ							
☐ QC_XANOMELINE_FREQ							
☐ PROD_PLACEBO_PERC							
☐ QC_PLACEBO_PERC							
☐ PROD_XANOMELINE_PERC							
☐ QC_XANOMELINE_PERC							
☐ PROD_PLACEBO_CI							

Step 4: Identify source (2)



Use common approach

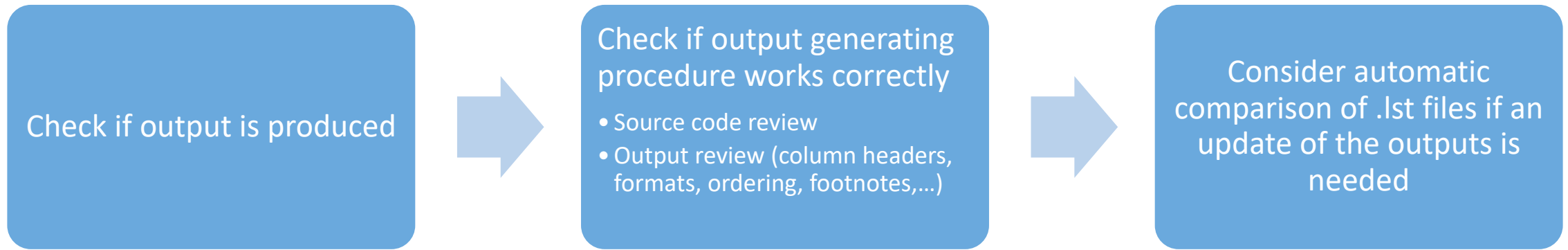
- Source code review
- Log file review using options
MPRINT/MLOGIC/SYMBOLGEN
- Check datasets from interactive run
- Ask main programmer to provide a dataset
with calculations on patient level

Log file messages

- All issues are also shown in the log file as error or warning messages
- For reruns checking all PROC COMPARE outputs can be replaced by log file scans

```
ERROR: output outtlf/t_001_aefreq.lst does not exist
WARNING: QC of DSTLF.T_001_AEFREQ - variable ORDER in PROD dataset only
WARNING: QC of DSTLF.T_001_AEFREQ - variable PLACEBO_PERC with conflicting types
WARNING: QC of DSTLF.T_001_AEFREQ - variable TEXT with different labels
WARNING: QC of DSTLF.T_001_AEFREQ - variable XANOMELINE_PERC with conflicting types
WARNING: QC of DSTLF.T_001_AEFREQ - values for key variables differ
WARNING: QC of DSTLF.T_001_AEFREQ - values differ
```

Follow-up work



Thank you for your attention

fenja.seither@mainanalytics.de
judith.kanz@mainanalytics.de

