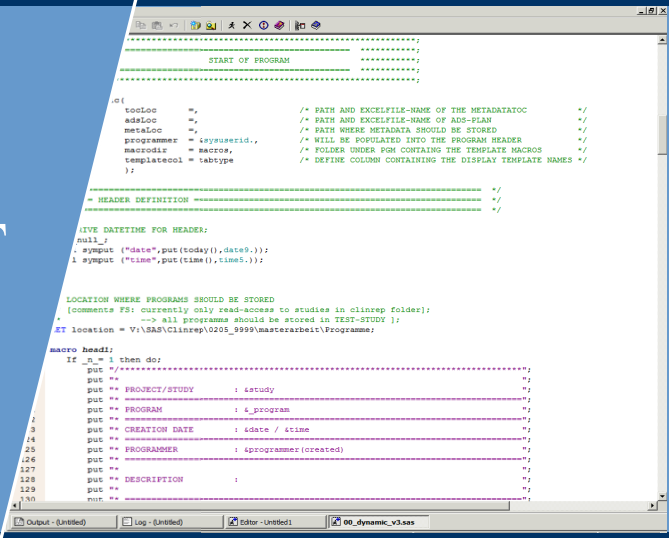
University of Bremen

METADATA DRIVEN PROGRAMMING OF CLINICAL STATISTICAL RESULT DISPLAYS



Master Thesis Presentation

by Florian Storch

16.05.2017

Boehringer Ingelheim

Contents

Boehringer Ingelheim

Introduction - Background

Boehringer Ingelheim

Definition metadata

- Classic definition: „data about data“
- Can be used to describe a collection of related entities (e.g. data, directory structure, …)
- Find them for example in SAS® procedure PROC CONTENTS (e.g. variable name, variable label, length, type)

Example:
AGEYR
- without metadata → guess that AGEYR represents a person‘s age in years
- with metadata → discover from variable label that age is rounded to nearest whole year

Introduction – Background [2]

Boehringer Ingelheim

Example: Consider work flow for a project

- Specifications of deliverables
- Programming
  - Datasets
  - Reports (Tables, Listings, Figures)
- Deliverables to client
  - Datasets
  - Reports
  - Supporting documentation

How are these typically created?

- Specifications in Word or Excel document
- Transcription / cut & paste of variable names, titles and footnotes, etc. into programs

Problems

- Tedious
- Time consuming
- Error-prone

Solution

Development of metadata-driven SAS® programs by identifying repeated items and fixed components in the program code

Contents

Content of the master thesis

Boehringer Ingelheim

Goal

- **To create clinical statistical result displays automatically by using metadata**

Table types

- Descriptive tables (e.g. Demographics)
- Frequency tables (e.g. Medical History)
- Dispostion Tables
- Efficacy tables using ancova

Tasks (To do)

- Identification of the required metadata
- Definition and description of these metadata
- Set-up an environment for the input and storage of these metadata

Table 40: <Endpoint name> [unit] - MMRM results – FAS

				---------- Active vs Placebo -------		
Timepoint	Treatment	N	Adjusted* mean (SE)	Adjusted* mean of difference (SE)	xx% CI	p-value
Week 4	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 8	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 16	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 24	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx

* adjusted for treatment, centre, time, baseline and treatment*time
Common baseline mean (sd at visit 2 = x.xxx (x.xxx)
A spatial power covariance structure is used in the model

Contents

Boehringer Ingelheim

1. Step - Analysis of wanted output table

Table 190: Demographic data – treated set

	Placebo	New Drug 1	New Drug 2	Comparator	Total
Number of patients [N (%)]	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)
Gender [N (%)]					
Male	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Female	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Race [N (%)]					
American Indian / Alaska Native	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Asian	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Black / African American	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Native Hawaiian / Pacific Islander	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
White	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Ethnicity [N (%)]					
Not Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Age [years]					
N	xx	xx	xx	xx	xx
Mean	xx.x	xx.x	xx.x	xx.x	xx.x
SD	xx.x	xx.x	xx.x	xx.x	xx.x
Min	xx	xx	xx	xx	xx
Q1	xx	xx	xx	xx	xx
Median	xx.x	xx.x	xx.x	xx.x	xx.x
Q3	xx	xx	xx	xx	xx
Max	xx	xx	xx	xx	xx

Output specific

Analysis specific

Display specific

Methods - Identification of the required metadata [2]

Boehringer Ingelheim

2. Step - Programming the wanted output table and check repeated items and/or fixed components in the program code

Contents

Results - Implementation

Boehringer Ingelheim

Documents to enter and maintain the metadata

- **MetadataToC**
- Analysis Dataset (ADS) plan

SAS environment to use the metadata

- **Dynamic Tool**
- Table creating macros
  - Descriptive tables (demographic tables)
  - Frequency tables (medical history tables)
  - Disposition tables
  - Efficacy tables (using ancova)

Results - MetadataToC

Boehringer Ingelheim

1 MetadataToC

2 MetadataToC sheet

STUDY	APPENDIX	NUMBER	…	TABTYPE	DATASET
1234_5678	15.1	1: 1	…	disp	ADSL
1234_5678	15.1	4: 1	…	demo	ADSL

3 StatCat sheet

STATCAT	STATISTICS	FORMAT	ORDER
FA_1	n	4.0	1
FA_1	mean	5.1	2
FA_1	std	5.1	3
FA_1	min	4.0	4

1. Electronic document to enter and maintain the metadata. Consists of two sheets
2. Extension of ToC[1]. Most of the metadata are managed here.
3. Controlling of statistical analysis details (kind of statistic, format, order)

[1] ToC = Table of Contents

Results - ADS plan

Boehringer Ingelheim

1 ADS plan

2 ADSL sheet

Dataset Name	Variable	Variable Label	…	Variable Display Label	CT Name
ADSL	AGE	Age	…	Age [years]	
ADSL	SEXN	Sex	…	Gender [N(%)]	SEXN

3 CT sheet [2]

CT Name	…	Code	Decode
SEXN	…	1	Male
SEXN	…	2	Female

1 Electronic document to enter and maintain metadata related to variables.
Consists of two sheets

2 Extension of ADS plan[1] with additional column "Variable Display Label" (Sheets for other datasets are not necessary)

3 Controlling of the controlled term

[1] ADS plan of Boehringer Ingelheim, ADS = Analysis Dataset

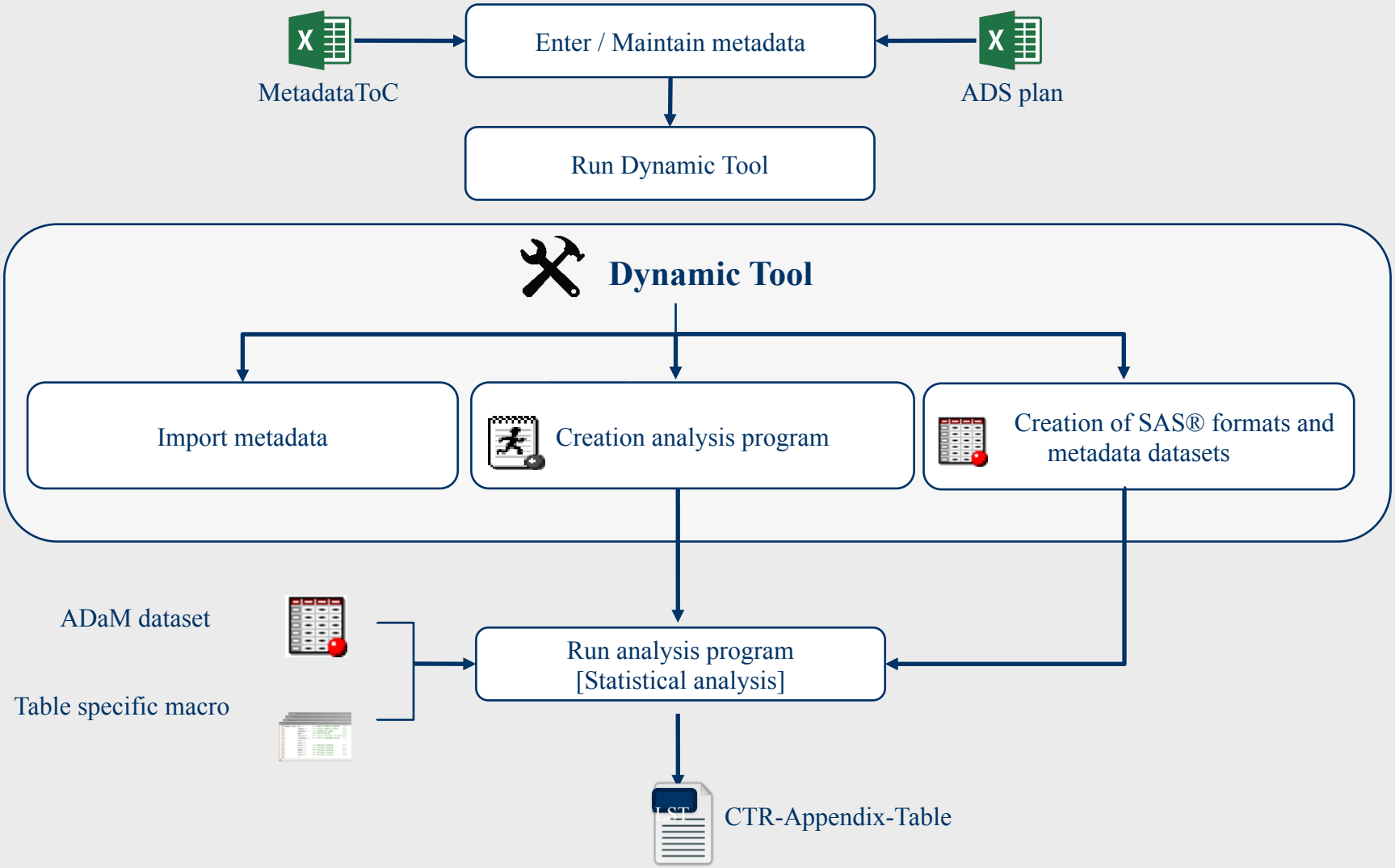
Results - Dynamic Tool

Boehringer Ingelheim

Functionality
- Read in MetadataToC and ADS plan
- Creates permanent SAS® datasets to store metadata
- Building SAS® formats from CT[1] sheet
- **Generation of SAS® code, which contains the required macro call for the creation of the output table defined in MetadataToC**
- **Generated SAS® code will be saved as separate SAS® program per output table (analysis program)**
- Error checks

[1] CT = Controlled Terms

Results - Process Flow Chart



Contents

Example

Boehringer Ingelheim

Specifications for output table

What? (Specification)	How? (Content)
Creation of a demographic table:	containing AGE, SEX and BMI
Specifications **AGE:**	mean 5.2 std 4.1 max 3.0
Specifications **BMI:**	mean 4.1 std 5.2 median 4.1
4 treatment groups:	New Drug 1, New Drug 2, Placebo Comparator
Order **variables:**	SEX – AGE – BMI
Order **treatment:**	Placebo, New Drug 1, New Drug 2, Comparator

Example [2]

Boehringer Ingelheim

1a) Enter metadata into MetadataToC

- Enter metadata in a new row in MetadataToC sheet
- Variables entered in ANALVAR
- ANALVAR also controls the used statistics of a variable (link to StatCat sheet)

A	B	C	F
STUDY	APPENDIX	NUMBER	TITLE 1
1234_5678	15.1	4: 1	Demographic data - treated set

R	V	AF	AG	AI
PROGRAM	OUTNAME	TABTYPE	DATASET	POPUSET
demo_ds.sas	demo_ds1.lst	demo	ADSL	TRTFN

AL	AR
ANALVAR	TREATORD
#SEXN, AGE(Ex_1), BMIBL(Ex_2)	Placebo, New Drug 1, New Drug 2, Comparator

MetadataToC sheet

- Age and BMI are using different statistics and formats
  → define two new categories
- Specify the format for each statistic
- "Order" controls the order of the statistics (1 = displayed on the top
  3 = displayed on the bottom)

StatCat	Statistics	Format	Order
Ex_1	mean	5.2	1
Ex_1	std	4.1	2
Ex_1	max	3.0	3
Ex_2	mean	4.1	1
Ex_2	std	5.2	2
Ex_2	median	4.1	3

StatCat sheet

Example [3]

Boehringer Ingelheim

1b) Enter metadata into ADS plan

- ADSL sheet contains new column “Variable Display Label”
- Use “Variable Display Label” to control how a variable is labeled in the output table
- Use “CT Name” to link to CT sheet

A	B	D	E	F	J
Dataset Name	***Used by Study***	***Variable Name***	***Variable Label***	***Variable Display Label***	***CT Name***
ADSL	1234_5678	AGE	Age	Age [years]	
ADSL	1234_5678	SEXN	Sex (N)	Gender [N (%)]	SEXN
ADSL	1234_5678	BMIBL	BMI	Body mass index [kg/m²]	

ADSL sheet

- Define the Controlled Terms (CT) as usual

A	D	E	F	G
CT Name	***Type***	***Sort***	***Code***	***Decode***
SEXN	integer	1	1	Male
SEXN	integer	2	2	Female

CT sheet

Example [4]

Boehringer Ingelheim

2) Use Dynamic Tool

```
*   ------------------------------------------------------------------------------  *;
*   USED LIBRARIES                                                                  *;
*   ------------------------------------------------------------------------------  *;
libname metadata "C:\masterarbeit\Daten\Metadata";

*   ------------------------------------------------------------------------------  *;
*   INCLUDE MACRO "DYNAMIC"                                                         *;
*   ------------------------------------------------------------------------------  *;
%include "C:\masterarbeit\Programme\Dynamic Tool\00_dynamic.sas";


/*  ============================================================================== */
/*  ===== MACRO CALL ============================================================= */
/*  ============================================================================== */

%dynamic(tocLoc  = C:\masterarbeit\Dokumente\Archiv\MetadataToC.xlsm,
         adsLoc  = C:\masterarbeit\Dokumente\ADS_Plan.xlsm,
         metaLoc = C:\masterarbeit\Daten\Metadata);
```

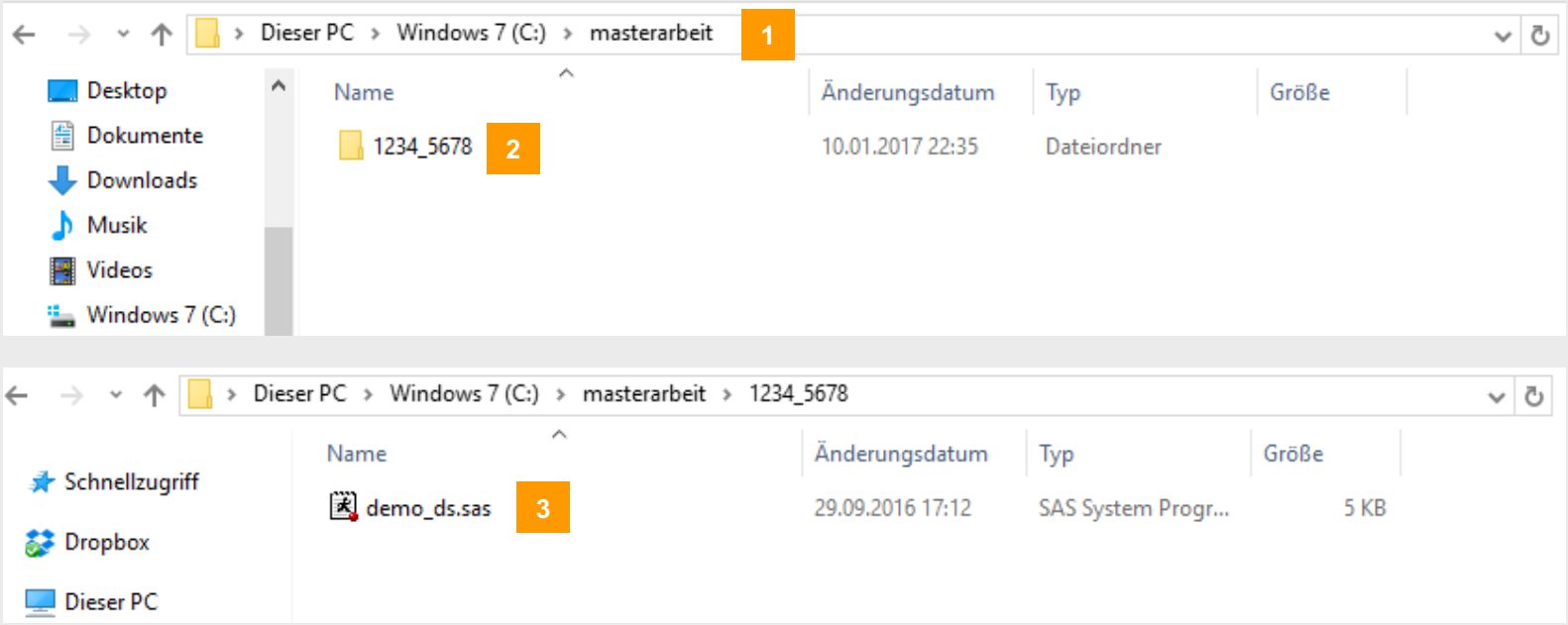
1. Define LIBNAME statement for the metadata (path has to be the same as path for metaLoc)
2. Include macro “00_dynamic”
3. Define parameters in macro call

Parameter
tocLoc = path and excel file name of the MetadataToC
adsLoc = path and excel file name of the ADS plan
metaLoc = path where metadata should be stored

Example [5]

Boehringer Ingelheim

Dynamic Tool creates new study folder in path defined in MetadataToC if this folder doesn't exist



1. Path as defined in MetadataToC
2. Study folder in defined path
3. SAS program created by Dynamic tool

Example [6]

Boehringer Ingelheim

3) Create output table

1. Path where the metadata are stored
2. Path where the macro demo is stored
3. Specific header/comment related to output table
4. Macro call (Result of Dynamic Tool)

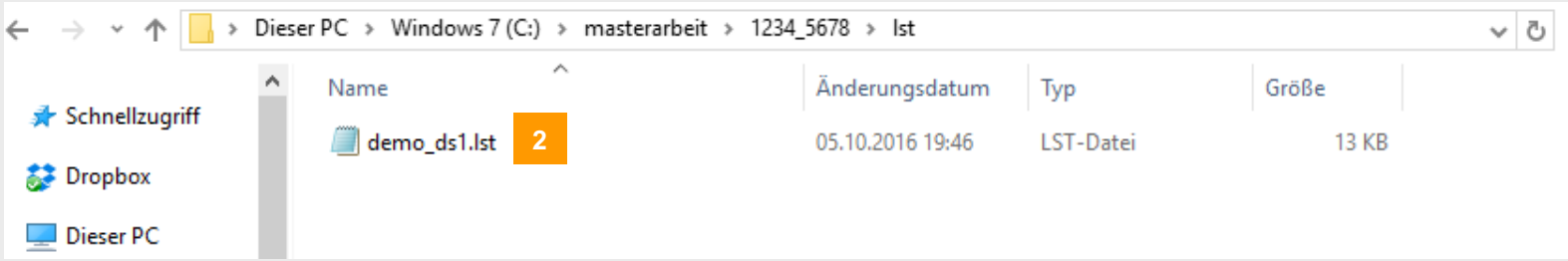
Run analysis program to create the output table

```
*=====================================================================
* Used libraries
*=====================================================================

1  LIBNAME metadata "C:\masterarbeit\Daten\Metadata";
   OPTIONS FMTSEARCH = (metadata);

*=====================================================================
* Include the template macro to be used.
*=====================================================================

2  %INCLUDE "C:\masterarbeit\Programme\macros\demo.sas";

*=====================================================================
* Repeatedly call the macro to produce the various summary outputs.
*=====================================================================



**********************************************************************;
***********  ==============================================  ***********;
***********                START OF PROGRAM                ***********;
***********  ==============================================  ***********;
**********************************************************************;


/*********************************************************************
* 15.1  4: 1  Table
3 * Demographic data - treated set
*
*********************************************************************/


4  %demo(
         ds        = ADSL
       , number    = 4: 1
       , appendix  = 15.1
       , popu      = TRTFN
       , varlist   = (#SEXN, AGE(EX_1), #BMIBL(Ex_2) )
       , treatment = (Placebo, New Drug 1, New Drug 2, Comparator )
       );
```

Example [7]

1 New folder created

2 lst.-file contains output table



Example [8]

Boehringer Ingelheim

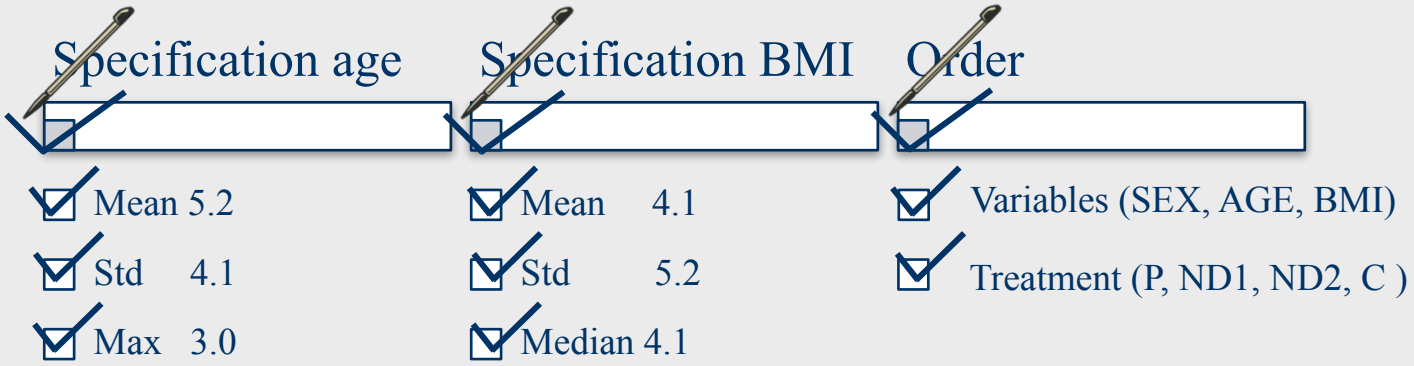


Table 15.1.4: 1 Demographic data - treated set

	Placebo	New Drug 1	New Drug 2	Comparator	Total
Number of patients [N (%)]	258 (100.0)	283 (100.0)	242 (100.0)	287 (100.0)	1070 (100.0)
Gender [N (%)]					
Male	118 (45.7)	141 (49.8)	117 (48.3)	152 (53.0)	528 (49.3)
Female	140 (54.3)	142 (50.2)	125 (51.7)	135 (47.0)	542 (50.7)
Age [years]					
Mean	43.53	43.72	42.96	43.17	43.86
Std	13.1	12.6	12.5	13.4	12.9
Max	76	75	72	74	76
Body mass index [kg/m²]					
Mean	26.7	27.0	27.0	26.7	26.9
Std	6.25	5.94	6.04	6.30	6.13
Median	25.5	25.6	26.1	25.4	25.6

Contents

Summary

Boehringer Ingelheim

- **Creation of following table types:**
  descriptive tables, frequency tables (medical history), disposition tables, efficacy tables (ancova)

- Documents to enter metadata: MetadataToC, ADS plan

- Import MetadataToC and ADS plan by using Dynamic Tool

- Dynamic Tool generates analysis program

- Run analysis program for creation of output table

- Output table will be stored as lst-file

Boehringer Ingelheim

THANK YOU TO

Boehringer Ingelheim Pharma GmbH & Co. KG
Statistical Programming Europe
Dr. Michael Knößl
Martin Schmidt

Universität Bremen
Fachbereich 3 – Mathematik und Informatik
Dr. Werner Wosniok