

# **A Macro Tool to Reduce Dataset Size for the Regulatory Submissions**

Venkata N Madhira  
Sr. Statistical Programmer  
Consultant at Shionogi Inc, NJ, USA

PhUSE US Connect 2019

# Introduction

- As per the FDA requirements, Submission datasets must be in SAS V5 Transport file format.
- The overall size of the each individual file should not exceed 5 GB.

# Dataset Size is Greater Than 5 GB ?

```
graph TD; A[Dataset Size is Greater Than 5 GB ?] --> B[More Number of Observations]; A --> C[Assigning Extra Length to a Character Variable Unnecessarily];
```

More Number of  
Observations

Assigning Extra Length  
to a Character Variable  
Unnecessarily

# Dataset Size is Greater Than 5 GB ?

```
graph TD; A[Dataset Size is Greater Than 5 GB ?] --> B[More Number of Observations and/or Variables]; A --> C[Assigning Extra Length to a Character Variable Unnecessarily];
```

More Number of  
Observations and/or  
Variables

Assigning Extra Length  
to a Character Variable  
Unnecessarily

As per the study data technical conformance guide:

- It is recommended that the each character variable value length should be as same as the maximum value length of corresponding variable across all domains within the study
- For suppqual datasets, each variable value length should be set to its maximum value length.

- In order to accomplish these complicated tasks, we created a macro tool called ALTVARLEN.
- **Advantages of the Macro Tool ALTVARLEN**
- It is very easy to use (just a couple of parameters to be passed).
- This macro follows the recommendations in study data technical conformance guide.

- This macro resets the character variable length to its maximum value length across the domains within specified library except for suppqual datasets.
- For suppqual dataset variables, resets the variable length to its maximum value length.
- Provides each dataset character variable length summary before and after execution of ALTVARLEN macro in html window where the character variable length had been reassigned.

# Mechanism for **ALTVARLEN**

`%altvarlen (acrossdsn=, lib=);`

| Keyword Parameter | Parameter Value                                                       | Description                                                                                                                                                                                           |
|-------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| lib               | The library name in which your datasets are stored (eg: WORK, Raw...) | <code>%altvarlen (acrossdsn=, lib=WORK);</code>                                                                                                                                                       |
| acrossdsn         | When the value is N<br>(Possible values are Y, N only)                | <code>%altvarlen (acrossdsn=N, lib=);</code><br>For suppqual dataset variables, resets the variable length to its maximum value length.                                                               |
| acrossdsn         | When the value is Y<br>(Possible values are Y, N only)                | <code>%altvarlen (acrossdsn=Y, lib=);</code><br>This macro resets the character variable length to its maximum value length across the domains within specified library except for suppqual datasets. |

- %altvarlen (acrossdsn=Y, lib=WORK);
- The character variable length is reset to its maximum value length across the domains within work library except for suppqual datasets.
- For suppqual dataset variables, resets the variable length to its maximum value length.

Provides each dataset character variable length summary before and after execution of ALTVARLEN macro in html window where the character variable length had been reassigned.

| Domain | Variable Name | Variable Length BEFORE Execution of Macro | Variable Length AFTER Execution of Macro |
|--------|---------------|-------------------------------------------|------------------------------------------|
| AE     | AEACN         | 20                                        | 16                                       |
| AE     | AEBODSYS      | 200                                       | 67                                       |
| AE     | AECONTRT      | 2                                         | 1                                        |
| AE     | AEDECOD       | 200                                       | 47                                       |
| AE     | AEENDTC       | 16                                        | 10                                       |
| AE     | AEHLGT        | 200                                       | 86                                       |
| LB     | LBORRES       | 200                                       | 40                                       |
| LB     | LBORRESU      | 20                                        | 9                                        |
| LB     | LBREASND      | 200                                       | 144                                      |

# Snippet of Code

```
proc contents data=&dsn out=__tempcont (keep=name length type where=(TYPE eq 2 and  
NAME not in (&exc))) noprint;  
run;  
  
proc sql noprint;  
  select 'max (length('||compress(name)||')) as '||compress(name) into: var  
  separated by ', ' from __tempcont;  
quit;  
  
proc sql noprint; create table __max_count as select &var from &dsn; quit;  
  
proc transpose data=__max_count out=__tp;  
run;  
  
proc sql noprint;  
select compress(_name_)||' character ('||strip(put(coll,best.))||')' into:  
variable separated by ', ' from _tp; alter table &dsn modify &variable;  
quit;
```

# CONCLUSION

- Using the macro tool ALTVARLEN, we can reassign the character variable length following study data technical conformance guide, which leads to reduction in dataset size.

# Acknowledgement

- I would like to thank Malla Reddy Boda, Associate Director, Shionogi Inc, for his review and invaluable comments.

Thank You

# CONTACT INFORMATION

Venkata N Madhira

E-mail: [venkatanmadhira@gmail.com](mailto:venkatanmadhira@gmail.com)

Sr. Statistical Programmer, Consultant at  
Shionogi Inc, NJ, USA

Harish Yeluguri

E-mail: [harish.yeluguri@gmail.com](mailto:harish.yeluguri@gmail.com)

Sr. Statistical Programmer, Consultant at  
Shionogi Inc, NJ, USA

Prabhakara Rao Burma

E-mail: [prsas85@gmail.com](mailto:prsas85@gmail.com)

Principal Clinical SAS Analyst, Acerta Pharma,  
CA, USA

Questions ?

# REFERENCES

- Analysis Data Model Implementation Guide Version 1.1
- SAS<sup>®</sup> Certification Prep Guide Advanced Programming for SAS<sup>®</sup> 9
- <https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf>
- SAS<sup>®</sup> and all other SAS<sup>®</sup> Institute Inc. product or service names are registered trademarks or trademarks of SAS<sup>®</sup> Institute Inc. in the USA and other countries. <sup>®</sup> indicates USA registration. Other brand and product names are trademarks of their respective companies.