



PharmaSUG '04
May 23-26 San Diego

Mergers, Consolidations and Proc Compare



Global Research & Development

Fang Dong

Pfizer Global Research & Development,

Michigan Laboratories

Ann Arbor

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or Trademarks of their respective companies.



PharmaSUG '04
May 23-26 San Diego

Overview

- ★ **Introduction**
 - ★ **What is the Issue?**
 - ★ **Why It Is Important**
 - ★ **How We Approached It?**
 - ★ **Conclusions**
-



PharmaSUG '04
May 23-26 San Diego

Introduction

★ **Data Base Consolidation**

- ★ **Ann Arbor → Legacy Pfizer Database**
- ★ **Single Instance**
- ★ **Retain AA Domain**
- ★ **Some Data Standards and Metadata Change**

★ **User Acceptance Testing**

- ★ **Ensure Data Integrity**
 - ★ **Ensure Reporting System Work as Expected**
-



Technology Used

★ Proc Compare

- ★ Flexible to identify data differences

★ SAS Macros

- ★ Powerful to compare a large amount of data



Checking Data Integrity for A Single Data Set

★ Access Data

★ Test Database:

```
Libname testview
```

```
    '/test_view/study_a0001/snapshot1';
```

★ AA Production DB:

```
Libname prodview
```

```
    '/prod_view/study_a0001/snapshot1';
```



PharmaSUG '04
May 23-26 San Diego

Checking Data Integrity for A Single Data Set

★ Example 1:

Examine Demographic Data



Example 1: SAS Code

```
/*Extracting data both databases*/  
data demo_p;  
  set prodview.demo;  
run;  
data demo_t;  
  set testview.demo;  
run;  
  
/*Sort the data by subject number pt*/  
Proc sort data=demo_p;  by pt; run;  
Proc sort data=demo_t;  by pt; run;  
  
/*Compare data sets with default options*/  
proc compare base=demo_p compare= demo_t; run;
```



PharmaSUG '04
May 23-26 San Diego

Example 1: Output

Surprise!



Example 1: Output

Variables with Unequal Values

Variable	Type	Len	Label	Ndif	MaxDif
DOCNUM	CHAR	20	Document Number	94	
INVSITE	CHAR	10	Site	94	
INV	CHAR	10	Investigator	94	

Value Comparison Results for Variables

		Document Number	
		Base Value	Compare Value
Obs		DOCNUM	DOCNUM
1		D346371601	AA_D346371601
2		D221849001	AA_D221849001
3		D221849101	AA_D221849101
		



PharmaSUG '04
May 23-26 San Diego

Checking Data Integrity for A Single Data Set

★ **Example 2:**

Examine Demographic Data

Adjusted for the Expected Data Changes



PharmaSUG '04
May 23-26 San Diego

Example 2: SAS Code

```
/*Adjust for the expected changes*/
```

```
data demo_p;
```

```
set prodview.demo;
```

```
invsite = '000' || invsite;
```

```
docnum = 'AA_' || docnum;
```

```
inv = 'AA_' || inv;
```

```
run;
```

```
Proc sort data=demo_p; by pt; run;
```

```
/*Compare data sets with default options*/
```

```
proc compare base=demo_p compare= demo_t; run;
```



Example 2: Output

Data Set Summary

Dataset	Created	Modified	NVar	NObs
WORK.DEMO_P	30JAN04:15:36:18	30JAN04:15:36:18	28	94
WORK.DEMO_T	30JAN04:13:12:41	30JAN04:13:12:41	27	94

Variables Summary

Number of Variables in Common: 27.

Number of Variables in WORK.DEMO_P but not in WORK.DEMO_T: 1.

Observation Summary

Observation	Base	Compare
First Obs	1	1
Last Obs	94	94

Number of Observations in Common: 94.

....

Number of Observations with Some Compared Variables Unequal: 0.

Number of Observations with All Compared Variables Equal: 94.



Checking Data Integrity for A Single Data Set

- ★ **Example 3:**
 - ★ **Examine Demographic Data**
 - ★ **Adjusted for the Expected Data Changes**
 - ★ **Use option *LISTVAR* to find out which variable is missing**
-



PharmaSUG '04
May 23-26 San Diego

Example 3: SAS Code

```
proc compare base=demo_p compare= demo_t  
  listvar;  
run;
```



Example 3: Output

Variables Summary

Number of Variables in Common: 27.

Number of Variables in WORK.DEMO_P but not in
WORK.DEMO_T: 1.

Listing of Variables in WORK.DEMO_P but not in
WORK.DEMO_T

Variable	Type	Length	Format	Label
----------	------	--------	--------	-------

LASTDT	Char	8	\$8.	Date Dose Ended
--------	------	---	------	-----------------



Checking Data Integrity for All Data Sets in A Study

- ★ **Example 4: Examine All Data Sets in A Study**
 - ★ **Determine Number of Data Sets and Data Set Names**
 - ★ **Extract Data from Test and AA Production DB**
 - ★ **Sort the Data Set by Key Variables**
 - ★ **Compare Data Sets**



Example 4: Step 1

```
/*Get all data sets in the prodview directory*/  
proc contents data=prodview._all_ out=_dsnlist  
  noprint; run;  
data _dsnall (keep=memname);  
  set _dsnlist;  
  by memname;  
  if first.memname; run;  
  
/*Get data set names and number of data sets*/  
data _null_;  
  set _dsnall end=eof;  
  call symput('dset' || left(_n_),memname);  
  if eof then call symput('totdset',_n_);  
run;
```



PharmaSUG '04
May 23-26 San Diego

Example 4: Step 2

```
%Macro Verify;  
/*extracting data from both databases*/  
%do ss=1 %to &totdset;  
  data &&dset&ss._t;  
  set testview.&&dset&ss;  
run;  
%end;  
%do ss=1 %to &totdset;  
  data &&dset&ss._p;  
  set prodview.&&dset&ss;  
  docnum='AA_'!!docnum; inv='AA_'!!inv;  
  invsite='000'!!invsite;  
run;  
%end;
```



PharmaSUG '04
May 23-26 San Diego

Example 4: Step 3

```
/*Sort the data sets by key variables*/  
  
%do ss=1 %to &totdset;  
  Proc sort data=&&dset&ss._p;  
    by pt;  
  run;  
  
  Proc sort data=&&dset&ss._t;  
    by pt;  
  run;  
%end;
```



Example 4: Step 4

```
/*Compare data sets with default options*/  
  
%do ss=1 %to &totdset;  
title "Comparing &&dset&ss._p and&&dset&ss._t";  
proc compare base=&&dset&ss._p  
compare=&&dset&ss._t;  
run;  
title;  
%end;  
%mend verify;
```



Checking Data Integrity for All Data Sets in Any Studies

- ★ **Example 5: Examine Data in Any Studies**
 - ★ **Why Is It Necessary?**
 - ★ **Expand the Macro in Example 4**
 - ★ **Make *Study* and *Snapshot* Macro Parameters**
 - ★ **Re-set the Libname Statements**
 - ★ **Repeat Steps 1-4 of Example 4**
-



Example 5: Adding Step 0

```
%Macro Verify(study, snapshot);  
  
/*legacy production database view*/  
  
Libname prodview "/prod_view/&study./&snapshot";  
  
/*test database view*/  
Libname testview "/test_view/&study./&snapshot";  
...Ditto ...  
step 1-4 of Example 4  
  
%Mend Verify;
```



PharmaSUG '04
May 23-26 San Diego

Preparing Reporting Systems for the Change

- ★ **Infrastructure Change**
- ★ **Data Standard Change**
- ★ **Metadata Change**
- ★ **Process Change**



Preparing Reporting Systems for the Change

- ★ **Infrastructure Change**
 - ✦ Database access methods,
 - new server,
 - new connection methods
 - new instance name,
 - new accounts.
 - Should be fully tested



PharmaSUG '04
May 23-26 San Diego

Preparing Reporting Systems for the Change

- ★ Infrastructure Change
- ★ **Data Standard Change**
 - ★ Prefix of AA_
 - ★ Process Change



PharmaSUG '04
May 23-26 San Diego

Preparing Reporting Systems for the Change

- ★ Infrastructure Change
- ★ Data Standard Change
- ★ **Metadata Change**
 - ★ **Impact-less: Internal Keys**
 - ★ **Impactful: Domain Name AA_Standard**



PharmaSUG '04
May 23-26 San Diego

Preparing Reporting Systems for the Change

- ★ Infrastructure Change
- ★ Data Standard Change
- ★ Metadata Change
- ★ **Process Change**
 - ★ **Configuration Setting Change → Variable Name Change (NOTDONE → NOT_DONE)**



Lessons Learned

★ **Communication**

- ★ **'Language' Barriers Between Departments**
- ★ **Stakeholders**

★ **Plan and Execute**

- ★ **Detailed, Carefully Thought-out UAT Plan**
- ★ **Resources**

★ **Training of General Users**

- ★ **Info Sharing Sessions**
- ★ **Testers as SMEs**



PharmaSUG '04
May 23-26 San Diego

Conclusions

- ★ **SAS Proc Compare:**
 - ★ **Functionality to Check Data Integrity**

 - ★ **SAS Macro:**
 - ★ **Flexible and Powerful to Check Large Amount of Data**

 - ★ **SAS Programmers: Important Contribution**
 - ★ **Success of the Database Consolidation Initiative**
-



PharmaSUG '04
May 23-26 San Diego

Questions?



PharmaSUG '04
May 23-26 San Diego

Contact Information

Fang Dong

Associate Director

Pfizer Inc.

2800 Plymouth Road

Ann Arbor, MI 48105

(734) 622-1530

fang.dong@pfizer.com
