Paper RX-09-2013

Using SAS® in Clinical Programming Project Management from Excel® Spreadsheets

Jean Crain, InVentiv Health Clinical LLC, Indianapolis, IN

TRACKING YOUR PROJECT

Tracking Clinical Programming Project Management as we do from Microsoft Excel© Spreadsheets can give you a wealth of information about your projects if you have standardized format and method of entry by programming and statistics. From our Excel® spreadsheets, using Base SAS®, SAS® Macros and SAS® PROC REPORT we can retrieve the following data and present valuable information on how long a time period project spans, number of resources, time spent, productivity, and audit checking. The data obtained can provide multiple reports for analysis:

- Number of Programs
- Number of Outputs
- Number of Programs by Type (Analysis Data Set, Graph, Listing, Summary)
- Number of Outputs by Type (Analysis Data Set, Graph, Listing, Summary)
- Number of Programs by Employee (Programming, Validation and Statistical Validation and Review)
- Number of Outputs/Reviews by Employee
- Percent completion on projects by type: Original Programming, Validation Programming, Initial Statistical Review, Senior Statistical Review; number of outputs completed / number of outputs.
- First program completion date
- Last program completion date
- Number of calendar days (last date - first date), Overall and by Type.
- Number of work days (count unique dates), Overall and by Type.
- Audit check: Do Initials for Programming and Statistical Activities have a matching Name on Signature Page?
- Comment Data to analyze for trends: When programs/outputs need correction, does that indicate there are there areas for improvement in Requirements? Do the same issues come up from project to project?

We can use this because we do have a standard for the spreadsheet tracking and entries by programming and statistics staff. Our hours are tracked elsewhere, but this will give you a good accounting of number of days for projections and it is an informative method of conducting process improvement.

THE SPREADSHEETS IN THE WORKBOOK

SIGNATURES AND TRACKING PROGRAMMING BY TYPES: ANALYSIS, GRAPH, LISTING AND SUMMARY

The Programming Review and Validation Form (PRVF) Excel® Workbook tracks every program and every output for a study. It also captures the initials and name of the employee on a Signature Page. The same initials are entered on the spreadsheets for the different types of output: Analysis Data Sets, Graphs, Listings and Summaries as are present on the Signature Page. On each programming spreadsheet in the workbook, an employee will enter their initials to accept a program and track its status on a row on that page identified for that step of the process. The date of completion of each step is entered on that spreadsheet by the employee, along with their initials, and a Status Flag is set when their step is completed, signaling to the next employee when it is ready to start their validation process.

We have double programming – Original Release and Original Validation – from 2 programmers, then review by 2 statisticians; the Initial Statistical Reviewer is also programming some portions to check results, and the Senior Statistical Reviewer. Four lines (one for each step) are entered on the PRVF for every output. It is updated and colored with different 'fill' colors based on the step or time point as completed or needing attention of the Original Programmer or Validation Programmer.

When the project starts, the PRVF contains only program name and output name in columns 1 and 2, and Delivery (an abbreviated description: Adhoc, BDR Part 1, Final, etc.) not shown here due to space, and Timepoint. As the steps proceed, flags are set so employees know whose job it falls to next. This outcome of Original Release is OC (Output Complete). The outcome of Original Validation is VC (Validation Complete) or NF (New Finding). When the original programmer addresses any New Finding, the line is set to IR (Issue Resolved). When the validation programmer agrees it is correct, the flag is set to VC. The same process is followed for Initial Statistical QC and Senior Statistical QC. The statistical QC flags are IQC (Initial Stats QC completed) and FQC (Final Stats QC completed). Below are 3 outputs showing the different stages of completion. 1 data set and 2 reports.

Figure 1. Sample of data from the PRVF used for this paper.

Analysis Spreadsheet for this project has 4 data sets. This example does not display the columns for Comments/Findings (New Findings by Validation programmer or Statistical Review) or Resolution (entered by OP on updating to address findings).

Prgname	Output	OP	VP	Comt By:	Date Added	Timepoint	Flag	OP	OP-Date	VP	VP-Date	SR	SR-Date
reactgen.sas	Reactgen	LE	AM	LE	3-Jul	Original Release	OC	LE	3-Jul				
reactgen.sas	Reactgen	LE	AM	AM	11-Jul	Original Validation	VC			AM	12-Jul		
reactgen.sas	Reactgen	LE		SE	17-Jul	Initial Stats QC	IQC					SE	17-Jul
reactgen.sas	reactgen	LE	JT	JT	18-Jul	Senior Stats QC	FQC					JT	18-Jul

The Summaries Spreadsheet for this project has rows for 32 programs and 102 outputs (tables). All outputs, whether it is a data set, htm, pdf, or cgm file, have 4 steps and at a minimum 4 rows; one for each step. This example shows a New Finding and an Issue Resolved example (without Comments/Finding or Resolved due to space considerations.)

saf_sae_one.sas	saf_sae_one	SG	CZ	SG	10-Jul	Original Release	OC	SG	10-Jul				
saf_sae_one.sas	saf_sae_one	SG	CZ	CZ	19-Jul	Original Validation	NF						
saf_sae_one.sas	saf_sae_one	SG	CZ	BD	18-Jul	Initial Stats QC							
saf_sae_one.sas	saf_sae_one	SG	CZ			Senior Stats QC							
saf_se_ons.sas	saf_se_ons_grp	SW	KV	SW	9-Jul	Original Release	OC	SW	9-Jul				
saf_se_ons.sas	saf_se_ons_grp	SW	KV	KV	12-Jul	Original Validation	VC	SW	15-Jul	KV	15-Jul		
saf_se_ons.sas	saf_se_ons_grp	SW	KV	KV	15-Jul	Original Validation	VC	SW	15-Jul	KV	15-Jul		
saf_se_ons.sas	saf_se_ons_grp	SW	KV	JT	18-Jul	Initial Stats QC	IR	AN	20-Jul				
saf_se_ons.sas	saf_se_ons_grp	SW	KV			Senior Stats QC							

IMPORT DATA FROM MICROSOFT EXCEL® SPREADSHEETS

There data are imported from Microsoft Excel ® spreadsheets in the PRVF workbook by sheet. I used the mixed=YES option because of date fields. I wanted to read the fields as character and create SAS dates. SAS Options is set as YEARCUTOFF=1920 in this program. Early on in a project, one spreadsheet can have dates and another not and I was getting an ERROR on a date field being both character and numeric when I used mixed=NO.

```
%macro getsheet(in=,out=,sheet=);

PROC IMPORT out = work.&in
            datafile = "&dirpath.\&PRVF_..xls"
            dbms = excel replace;
            usedate = NO;
            mixed = YES;
            sheet = &sheet;
run;
(more statements)
%mend getsheet;
%getsheet(in=sign_,out=sign,sheet=%nrstr('Signatures') );
%getsheet(in=ads_,out=ads,sheet=%nrstr('Analysis') );
%getsheet(in=sum_,out=sum,sheet=%nrstr('Summary') );
%getsheet(in=list_,out=list,sheet=%nrstr('Listing') );
%getsheet(in=graph_,out=graph,sheet=%nrstr('Graph') );
```

DATA SETS CREATED: WANTED (FROM ADS, SUM, LIST, GRAPH) AND INIT (FROM SIGN)

The programming sheets are all read in the data step WANTED. SAS dates are created from the data read in. Data from Signatures is treated differently because its purpose is different from tracking program progress and assignments. At the Final delivery for a project, the page is printed and signed and sent to Central Files.

In the data set INIT, created by reading the spreadsheet Signatures, Department is kept from the header record before Employee Initials and Names so that we have Initials, Name and Department. An Audit report is output for non-matches; either on Signature Page and not on spreadsheets or on spreadsheets and not on the Signature Page. This report is output with every run if observations exist in that match-merge data set.

Figure 2. Signature Page

Programming Review and Validation Sign-Off Form			
Sponsor: Partner Pharmaceuticals		Project Code: 0890001	
Protocol: Anewdrug		Study: A1010	
Folder Path: H:\Data\SAS\Partner\Anewdrug\A1010			
Signing below indicates agreement that the programs/tasks indicated with the corresponding initials on SOP XX-XX-XX Form 1 have been reviewed, output matches the specifications, all discrepancies have been resolved, and validation/QC is complete.			
Project Statistical Programmers			
Initials	Printed Name	Signature	Signature Date (ddmmmyyyy)
AM	Annie Major		
FN	Fatima Namaskar		
LE	Luz Enriquez		
KV	Katya Vollant		
MS	Mrudula Suri		
SG	Swarna Gupta		
TL	Tran Le		
SW	Sophia Wise		
MK	Madhavi Krishnan		
N-S	Nathalie Sanscartier		
JC	Jean Collins		
CE	Chrarles Edward		
BKS	Brook Keller, Sr.		
CZ	Chi Zhou		
ATB	Alex Batchelor		
Project Biostatisticians			
MZ	Ming Zhang		
SE	Sergio Espana		
BD	Brian Dun		
JT	James Thomas		

Figure 3. Audit Report

Audit Report on Study A1010 run on 03AUG13
Summary for BDR Part 1 of PRVF_MWSUG_2013.xls
Audit Report

Protocol	Initials	Department	Name	Condition
Anewdrug	AN			Name not on Signature Page. On Signature Page. No initials on worksheets.
Anewdrug	MK	Programming	Madhavi Krishnan	

Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

DATA DATES

Reading the data set WANTED, multiple rows are output for each employee action on the program or output in a data step using an array and a do loop. An observation is output for any date not missing with the employee associated with that activity: Original Programmer, Validation Programmer, or Stat Reviewer.

```
DATA dates(keep=&smylevel sheet program output dateof op vp timepoint statflag
comt: upd: stat: init_all );
    set wanted;
    array inits{4} $ comt_by upd_op   upd_vp   statrv ;
    array dates{4}   comt_dt upd_opdt upd_vpdt statrvdt ;

    %do i = 1 %to 4;
          if dates{&i} >. then do;
                dateof   = dates{&i};
                if inits{&i} = 'N/A' then inits{&i} = 'NA';
                init_all = inits{&i};
                format dateof date9.;
                output;
          end;
    %end;
run;
```

In a subsequent data step named EMP_ALL, multiple rows are output for each employee who is on a line in the PRVF where 2 employees performed the same function on the same record. For example, one person was Original Programmer but on their day off, another programmer had to update the program.

PROC SQL FOR COUNTING

PROC SQL will handle multiple observations for the same group variables using the keyword DISTINCT so no subsetting of the data set is required to count only once per group. A very large part of the program is PROC SQL create table statements, particularly the select distinct statements. I believe it is the fastest and easiest way to count numbers without data manipulation to create unique records so they only count as 1 per distinct level selected.

PROC SQL STATEMENTS USED TO COUNT PROGRAMS AND OUTPUTS:

```
  /* The statement below creates Data Set n_days_oa_protocol, if &smylevel is
Protocol, containing 1 observation: Protocol (Anewdrug) and n_days_oa (15) */

     PROC SQL;
      create table n_days_oa_&smylevel as
      select &lev,
             count(distinct dateof) as n_days_oa
      from  emp_all
      group by &lev
      order by &lev;

  /* The statement below creates Data Set n_days_detail_protocol, if &smylevel is
  Protocol, containing 4 observations for Protocol (Anewdrug), step, and
  n_days_detail:    Protocol     step                n_days_detail
                    Anewdrug     Initial Stats QC            1
                    Anewdrug     Original Release           14
                    Anewdrug     Original Validation        11
                    Anewdrug     Senior Stats QC             1   */

     create table n_days_detail_&smylevel as
     select &lev,
            timepoint as step,
            count(distinct dateof) as n_days_detail
     from  emp_all
     group by &lev, step
```

```
    order by &lev, step;
/* The statement below creates Data Set n_people with 1 observation if &smylev
(&lev) is protocol, containing protocol (Anewdrug) and n_emp (19)      */

    create table n_people as
    select &lev,
           count(distinct init) as n_emp
    from emp_all
    group by &lev
    order by &lev;

/* The statement below creates Data Set n_days_detail_any containing 1 observation
for every select statement since I am using union in the select statement, so 5
observations:    protocol     step                n_days_detail
                 Anewdrug     Any Activity              15
                 Anewdrug     Biostatistics              7
                 Anewdrug     Initial Stats QC           7
                 Anewdrug     Programming               14
                 Anewdrug     Senior Stats QC            2           */

    create table n_days_detail_any as
    select &lev,
           'Any Activity' as step length=20,
           count(distinct dateof) as n_days_detail
    from  emp_all
    group by &lev, step
    union
    select &lev,
           'Programming' as step length=20,
           count(distinct dateof) as n_days_detail
    from  emp_all
    where timepoint in ('Original Release', 'Original Validation')
    group by &lev, step
    union
    select &lev,
           'Biostatistics' as step length=20,
           count(distinct dateof) as n_days_detail
    from  emp_all
    where timepoint in ('Initial Stats QC', 'Senior Stats QC')
    group by &lev, step
    order by &lev, step;

    create table emp_days_oa_&smylevel as
    select &lev,
           init,
           count(distinct dateof) as emp_days_oa
    from emp_all
    group by &lev, init
    order by &lev, init;

/* Many statements select data from the Data Set wanted which contains data about
programs and outputs. The statements below create Data Set n_programs containing 1
observation if the &smylevel (&lev) is protocol, as protocol (Anewdrug) and
n_programs (75)   */

    create table n_programs as
    select &lev,
           count(distinct program) as n_programs
    from wanted
    group by &lev
    order by &lev;
```

```
/* Further detail is included over the Data Set n_programs (above) in the statement
below which creates n_prog_type by adding sheet, which contains Analysis, Graph,
Listing, or Summary so it will have 4 observations:
protocol   sheet        n_programs
Anewdrug   Analysis              4
Anewdrug   Graphs                1
Anewdrug   Listings             38
Anewdrug   Summaries            32     */

    create table n_prog_type as
    select &lev,
           sheet,
           count(distinct program) as n_programs
    from wanted
    group by &lev, sheet
    order by &lev, sheet;

/* Expanding over the above level of detail besides sheet, timepoint and statflag
are used which creates Data Set n_prog_type_statflag with values of number of
programs, by sheet (Analysis, Graph, Listing, Summary), by timepoint (Original
Release, Original Validation, Initial Stats QC, Senior Stats QC) by statflag
(showing Completion as OC, VC, IQC, FQC, or NF for New Finding or IR for Issue
Resolved or blank for Not Done. You can see some of that data in Report 2 - Percent
Completed, which also uses pct_compl_prog data also shown below.  */

    create table n_prog_type_statflag as
    select &lev,
           sheet,
           timepoint,
           statflag,
           count(distinct program) as n_programs
    from wanted
    group by &lev, sheet, timepoint, statflag
    order by &lev, sheet, timepoint, statflag;

    create table pct_compl_prog as
    select a.&lev,
           'Programs' as type,
           a.sheet,
           a.timepoint,
           a.statflag,
           a.n_programs as n_items,
           b.n_programs as n_items_compl,
           sum(n_items/n_items_compl)*100 format=8.1 as pct
    from n_prog_type_statflag a,
           n_prog_type b
    where a.&lev = b.&lev
      and a.sheet = b.sheet
    group by a.&lev, type, a.sheet, timepoint, a.statflag
    order by a.&lev, type, a.sheet, timepoint, a.statflag;
quit;
run;
```

EXAMPLES COUNTING PROGRAMS AND OUTPUTS BY EMPLOYEE

The summary of Programs and Outputs by Employee is for Completed programs, so it is selected by a Status Flag indicating Completed. Role is created as OP, VP or STATS based on the column that initials are in (OP/OP Update, VP/VP Update, or SR and the value of Status Flag being OC, VC, or IQC/FQC respectively. These PROC SQL statements count programs and output at different summary levels.

```
/* At this level, when protocol is used as &smylevel (&lev), one observation is in
the Data Set n_prog_emp_compl for each employee with initials on a programming
```

```
worksheet with a status flag of Completed (OC, VC, IQC, FQC) containing protocol,
init, n_programs, n_outputs. */

PROC SQL;
    create table n_prog_emp_compl as
    select distinct &lev,
           emp_comp as init,
           count(distinct program) as n_programs,
           count(distinct output) as n_outputs
    from emp_all_compl
    group by &lev, init
    order by &lev, init;

/* One more level is included here, sheet, showing (Analysis, Graph, Listing,
Summary) and the same information as above under each of those categories.  */

    create table n_prog_type_emp_compl as
    select distinct &lev,
           sheet,
           emp_comp as init,
           count(distinct program) as n_programs,
           count(distinct output) as n_outputs
    from emp_all_compl
    group by &lev, init, sheet
    order by &lev, init, sheet;
quit;
run;
```

REPORTS USING PROC REPORT

NUMBERS

The following are excerpts from Reports 1 through 6 which display information about numbers, dates and days. Report 1 has 2 fields, Protocol and Prtdata, created in an intermediate data step after PROC SQL tables were created.

Figure 4. Report 1 – Numbers

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
Dates, Days, Numbers of Employees, Programs and Outputs

Protocol	Numbers
Anewdrug	Number of Employees on this project: 19
	Number of days overall: 15
	Calendar Dates range of 03JUL2013 - 20JUL2013 for Original Release. Number of days: 14
	Calendar Dates range of 05JUL2013 - 20JUL2013 for Original Validation. Number of days: 11
	Calendar Dates range of 15JUL2013 - 21JUL2013 for Initial Stats QC. Number of days: 7
	Calendar Dates range of 17JUL2013 - 18JUL2013 for Senior Stats QC. Number of days: 2
	Calendar Dates range of 03JUL2013 - 20JUL2013 for Programming. Number of days: 14
	Calendar Dates range of 15JUL2013 - 21JUL2013 for Biostatistics. Number of days: 7
	Calendar Dates range of 03JUL2013 - 21JUL2013 for Any Activity. Number of days: 15
	Number of programs: 75
	Number of programs for Analysis: 4
	Number of programs for Graph: 1
	Number of programs for Listing: 38
	Number of programs for Summary: 32
	Number of outputs: 191
	Number of outputs for Analysis: 4
	Number of outputs for Graph: 4
	Number of outputs for Listing: 81
	Number of outputs for Summary: 102

Note: Days is a count of dates signed for completion of a step, not a calculation of calendar days
Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

Report 2 has data with calculation of percent completed. The value under Number of Items is total number of overall outputs for that Protocol and Sheet. Analysis has 4 programs and 4 outputs because the output is a Data Set. Graph has 1 program and 4 outputs. Each item is assessed for unique (meaning 1 output per program) and copies (meaning number of additional macro calls producing output). IR, NF, ND are explained in footnotes and previously in this paper (NF and IR are explained previously on Page 2. ND = Not Done). Number of Programs and Number of Outputs separately determine the Percent Not Completed (Percent ND).

Figure 5. Report 2 – Percent Completed

Report on Study run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
N and % Complete, by Type and Timepoint

							-Completed-		-Not Completed-				
Protocol	Sheet	Timepoint	Type	Number of Items	Unique	Copies	N Compl	Percent Compl	IR	NF	ND	Tot	Percent ND
Anewdrug	Analysis	Original Release	Prog	4			4	100.00				0	0.00
			Output	4	4	0	4	100.00				0	0.00
		Original Validation	Prog	4			4	100.00				0	0.00
			Output	4	4	0	4	100.00				0	0.00
		Initial Stats QC	Prog	4			4	100.00				0	0.00
			Output	4	4	0	4	100.00				0	0.00
		Senior Stats QC	Prog	4			0	0.00				0	0.00
			Output	4	4	0	0	0.00				0	0.00
	Graph	Original Release	Prog	1			1	100.00				0	0.00
			Output	4	1	3	4	100.00				0	0.00
		Original Validation	Prog	1			0	0.00			1	1	100.00
			Output	4	1	3	0	0.00			4	4	100.00
		Initial Stats QC	Prog	1			0	0.00			1	1	100.00
			Output	4	1	3	0	0.00			4	4	100.00
		Senior Stats QC	Prog	1			0	0.00			1	1	100.00
			Output	4	1	3	0	0.00			4	4	100.00
	Listing	Original Release	Prog	38			38	100.00				0	0.00
			Output	81	38	43	81	100.00				0	0.00
		Original Validation	Prog	38			37	97.37		1		1	2.63
			Output	81	38	43	80	98.77		1		1	1.23
		Initial Stats QC	Prog	38			13	34.21	1	1	20	22	57.89
			Output	81	38	43	20	24.69	1	1	56	58	71.60
		Senior Stats QC	Prog	38			0	0.00			38	38	100.00
			Output	81	38	43	0	0.00			81	81	100.00
	Summary	Original Release	Prog	32			31	96.88			1	1	3.13
			Output	102	32	70	101	99.02			1	1	0.98
		Original Validation	Prog	32			25	78.13	3	4	2	9	28.13
			Output	102	32	70	89	87.25	3	4	6	13	12.75
		Initial Stats QC	Prog	32			10	31.25	4	2	14	20	62.50
			Output	102	32	70	51	50.00	4	18	27	49	48.04
		Senior Stats QC	Prog	32			0	0.00			32	32	100.00
			Output	102	32	70	0	0.00			101	101	99.02

Abbreviations: IR=Issue Resolved, NF=New Finding, ND=Not Done
Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

Reports 3 though 6 are Employee level summaries. Report 3 is Days, Dates and Numbers. Note the blank line on row 1. That is because of the erroneous entry of AN rather than AM on a spreadsheet, which does not match up with initials and name on the Signature Page.

Figure 6. Report 3 – Employee Data: Days, Dates and Numbers

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
By Employee Dates, Days, Numbers of Programs and Outputs

Protocol	Department	Name	Dates and Days[a]
Anewdrug			Calendar Dates range of 20JUL2013 - 20JUL2013. Nuumber of days: 1
	Biostatistics	Brian Dun	Calendar Dates range of 18JUL2013 - 19JUL2013. Nuumber of days: 2
		James Thomas	Calendar Dates range of 18JUL2013 - 19JUL2013. Nuumber of days: 2
		Ming Zhang	Calendar Dates range of 16JUL2013 - 19JUL2013. Nuumber of days: 2
		Sergio Espana	Calendar Dates range of 17JUL2013 - 21JUL2013. Nuumber of days: 4
	Programming	Annie Major	Calendar Dates range of 05JUL2013 - 20JUL2013. Nuumber of days: 11
		Alex Batchelor	Calendar Dates range of 16JUL2013 - 16JUL2013. Nuumber of days: 1
		Brook Keller, Sr.	Calendar Dates range of 09JUL2013 - 20JUL2013. Nuumber of days: 7
		Chrarles Edward	Calendar Dates range of 16JUL2013 - 20JUL2013. Nuumber of days: 5
		Chi Zhou	Calendar Dates range of 09JUL2013 - 19JUL2013. Nuumber of days: 8
		Fatima Namaskar	Calendar Dates range of 09JUL2013 - 19JUL2013. Nuumber of days: 8
		Jean Collins	Calendar Dates range of 16JUL2013 - 19JUL2013. Nuumber of days: 4
		Katya Vollant	Calendar Dates range of 05JUL2013 - 19JUL2013. Nuumber of days: 11
		Luz Enriquez	Calendar Dates range of 03JUL2013 - 19JUL2013. Nuumber of days: 8
		Mrudula Suri	Calendar Dates range of 08JUL2013 - 19JUL2013. Nuumber of days: 9
		Nathalie Sanscartier	Calendar Dates range of 11JUL2013 - 19JUL2013. Nuumber of days: 6
		Swarna Gupta	Calendar Dates range of 09JUL2013 - 19JUL2013. Nuumber of days: 8
		Sophia Wise	Calendar Dates range of 09JUL2013 - 20JUL2013. Nuumber of days: 9
		Tran Le	Calendar Dates range of 17JUL2013 - 19JUL2013. Nuumber of days: 3

a. Days is a count of dates signed for completion of a step, not a calculation of calendar days or actual days worked.
Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

Report 4 is Output Completed (N and %) by Employee. Percent in this case is percent of overall outputs in the study. This is good to know for Performance Reviews at year-end.

Figure 7. Report 4 – Output Completed (N and %) by Employee.

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
Output Completed (N and %) by Employee

Protocol	Department	Name	Programs	Outputs	Total Outputs	Unique	Copies	Percent of Total
Anewdrug	Biostatistics	Brian Dun	10	27	191	10	17	14.14
		James Thomas	7	35	191	7	28	18.32
		Ming Zhang	3	3	191	3	0	1.57
		Sergio Espana	11	14	191	11	3	7.33
	Programming	Annie Major	18	42	191	18	24	21.99
		Alex Batchelor	1	1	191	1	0	0.52
		Brook Keller, Sr.	12	16	191	12	4	8.38
		Chrarles Edward	7	7	191	7	0	3.66
		Chi Zhou	6	58	191	6	52	30.37
		Fatima Namaskar	13	18	191	13	5	9.42
		Jean Collins	3	3	191	3	0	1.57
		Katya Vollant	17	55	191	17	38	28.80
		Luz Enriquez	9	43	191	9	34	22.51
		Mrudula Suri	17	31	191	17	14	16.23
		Nathalie Sanscartier	12	13	191	12	1	6.81
		Swarna Gupta	11	28	191	11	17	14.66
		Sophia Wise	14	48	191	14	34	25.13
		Tran Le	2	2	191	2	0	1.05

Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

Report 5 is a Summary of Programs Completed, by Employee and Type. It is summarized by Employee and Type (Analysis, Graph, Summary, Listing) but does not include percent. It just includes numbers.

Figure 8. Report 5 – Programs Completed by Employee, by Type

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
Summary of Programs Completed[a] by Employee, by Type

Protocol	Department	Name	Type	Programs	Outputs	Unique	Copies
Anewdrug	Biostatistics	Brian Dun	Listing	6	10	6	4
			Summary	4	17	4	13
		James Thomas	Analysis	1	1	1	0
			Summary	6	34	6	28
		Ming Zhang	Analysis	3	3	3	0
		Sergio Espana	Analysis	4	4	4	0
			Listing	7	10	7	3
	Programming	Annie Major	Analysis	3	3	3	0
			Listing	9	12	9	3
			Summary	6	27	6	21
		Brook Keller, Sr.	Listing	11	15	11	4
			Summary	1	1	1	0
		Chi Zhou	Listing	1	35	1	34
			Summary	5	23	5	18
		Fatima Namaskar	Listing	11	16	11	5
			Summary	2	2	2	0
		Katya Vollant	Analysis	2	2	2	0
			Listing	6	6	6	0
			Summary	9	47	9	38
		Luz Enriquez	Analysis	2	2	2	0
			Listing	2	36	2	34
			Summary	5	5	5	0

a: Completed is Programs with status of Completed on which the employee worked.
Difficulty levels are not assigned at present.
Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

PROC REPORT USING SUM IN A COMPUTE BLOCK

Report 6 is a Listing of Programs Completed by Employee and Type. The use of a listing as supplemental information over the numbers reported above, was to determine complexity of programs as well as type programs (Analysis, Graphs, Listings and Summaries). Management is planning to assign a difficulty level on reports, which would allow weighting on numbers of programs. Proc Report was used to sum the number of outputs per Employee in the code below. Following the code is excerpts from the report to show summary lines.

```
      %let fileout=&dirpath.\&outname._EMP_COMPL_LIST.lst;
      %put fileout=&fileout;

      ods listing close;
      ods markup tagset = html.standard body="&fileout..htm";

      title4 "List of Programs Completed by Employee, by Type";
      title5 " ";
      footnote1 "Note: Completed is Programs with status of Completed on which the
  employee worked.";
      footnote2 "Difficulty levels are not assigned at present.";
      footnote3 "Program name: &dirpath.\&pgmname. ";
```

```
    PROC REPORT DATA = emp_data_rpt4 NOWINDOWS split = "~" HEADINE HEADSKIP
SPACING=1 MISSING;
          column &smylevel dept init name sheet program n_outputs role;
          define &smylevel  / order group "&psmylev."    width=10;
          define dept     / order group "Department"     width=20;
          define init     / order noprint ;
          define name     / order group "Name"           width=30;
          define sheet    / group "Type"                       width=10;
          define program    / group "Program"                  width=10;
          define n_outputs  / analysis sum "Outputs"     width=10;
          define role       / group "Role"                            width=10;

    compute after name;
          line 'Total Outputs' n_outputs.SUM ;
    endcomp;
    run;

%end;
title3;
```

Figure 9. Report 6 – List of Programs Completed by Employee.

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
List of Programs Completed by Employee, by Type

Protocol	Department	Name	Type	Program	Outputs
Anewdrug	Programming	Annie Major	Analysis	assay.sas	1 OP
				popula.sas	1 OP
				reactgen.sas	1 VP
			Listing	cdl_auto_imm.sas	1 OP
				cdl_excl.sas	4 OP
				cdl_hosp.sas	1 OP
				cdl_imm_flag.sas	1 OP
				cdl_mh.sas	1 OP
				cdl_neuro.sas	1 OP
				cdl_new_ill.sas	1 OP
				cdl_noncomp.sas	1 OP
				cdl_saf_flag.sas	1 OP
			Summary	cs_mh.sas	1 VP
				saf_ae.sas	6 VP
				saf_ae_sae_comp.sas	1 VP
				saf_fev_meds.sas	1 VP
				saf_lr_maxsev_prod.sas	9 VP
				saf_se_maxsev_prod.sas	9 VP
					Total Outputs 42
		Brook Keller, Sr.	Listing	cdl_bld_delay.sas	1 OP
				cdl_death.sas	1 VP
				cdl_delay.sas	1 OP
				cdl_enroll_site.sas	1 VP
				cdl_excl.sas	4 VP
				cdl_lr_unsch.sas	1 OP
				cdl_no_bld.sas	1 VP
				cdl_pregnancy.sas	1 VP
				cdl_se.sas	2 VP
				cdl_tpc.sas	1 VP
				cdl_vax_norand.sas	1 VP
			Summary	saf_lr_onset.sas	1 VP
					Total Outputs 16

Note: Completed is Programs with status of Completed on which the employee worked.
Difficulty levels are not assigned at present.
Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

REPORT OF COMMENTS/FINDINGS

Another table of comments/findings is created, not including comments not pertinent to Findings and created by timepoint (Original Release, Original Validation, Initial Stats QC, Senior Stats QC) in order to see where we could impove our processes, i.e. where things are being overlooked. Perhaps it is unclear requirements or perhaps it is just an oversight but it is helpful to review and it only takes a few minutes when it is nicely displayed.

```
   PROC SQL;
       create table comtfind as
       select distinct &lev,
                    timepoint as norder format=$stepord.,
                    timepoint,
                    comtfind
       from wanted
       where upcase(comtfind) not in ('OK', 'OK.', 'NO ISSUE', 'NO ISSUES',
       'NO ISSUES DETECTED', 'NO FINDINGS', 'NUMBERS MATCH', 'AS ABOVE', ' ' )
       group by &lev, norder, timepoint, comtfind
       order by &lev, norder, timepoint, comtfind;
   quit;
run;
```

Figure 10. Report 7 – Listing of Comments/Findings by Timepoint.

Report on Study A1010 run on 03AUG13
Subset for BDR Part 1 of PRVF_MWSUG_2013.xls
List of Comments/Findings

Protocol	Timepoint	Comment/Finding
Anewdrug	Original Release	Log saved. Requires StatXact.
	Original Validation	uninitialized variable in the LOG
		1) Table numbers don't correspond to outputs; 2) Age label should be: Age at randomization (years) 3) Warning in the log: WARNING: Apparent symbolic reference GE not resolved.
		1. Delete the space between # and : into the var for randomization group 2. Randomization Groups in one column 3. 'Dose Visit' instead of 'Injection Visit' header for the last 3 columns
		1. No pvalue requested on row 1 "Dosed". 2. Pvalue missing on row 2 "Discontinued". I have patients from each group discontinuing. This is just comp_s ne 1. 3. Specs show just Preferred Term in Column 1, no SOC. 4. Sort order within Events should be Preferred Term.
		1. Please remove 'Dose (as Administered)' from the second column header 3. Please add '0 (0.0) instead of empty
		1. Please use format for the Day colum same as for lr tables: Day1, Day2, day3 etc. 2. combine n and (%) in to one column.
		1. first column should be 'dose' 2. Numbers for Min-max should be Min, max
		1.Please fixed Randomized row's super script, please remove column header 'Dose (as Administered)' as per spec. Please remove extra part '</SUP' td sup?< Randomizedb
		Add footnote for Subject substitution of ATC Code and Text for missing. See specs.
		You need a summarized row for each ATC4 Name. See the indentation of PTEXT (SENAME) on specsum and see the output in Study 1012
		data inconsistant- query should be sent to client
	Initial Stats QC	1 - please check order - subject 10201021 has two fatigues out of chronological order
		I have 9 subjects from ITT population have OMP records from Data Sets: CONDRUG/CONDRIG_D. Please check.
		Need to add summary after "Any dose".
		Numbers ok can we remove NE in footnotes as it does not appear in table
		match numbers however, 1 - the word phase is misspelled in footnote 1
		match numbers however: 1 - please uppercase word dose for the first 3 periods {dose 1, dose 2, dose 3}

Program name: U:\Data\PM\MWSUG\Report_PRVF_MWSUG.sas

CONCLUSION

This paper presents some summaries and listings of the data in our PRVF spreadsheets that can help future projects by examination of current projects. Taking a bit of time to look at the Employees, days, numbers and comments in this manner can yield some insights into Project Management and into Management of Employees.

For example, are some employees only taking the same things study after study and not growing their capabilities? Are they only taking Validation Programmer role and never Original Programmer role (which requires more knowledge of PROC REPORT with an increased emphasis on the look and format of reports as well as html code)? Are some employees only completing listings and not summaries? Have they ever done graphs? You can help develop your employee by encouraging them to set goals to do different types of programs or more difficult programs and learn new skills.

The report by employee on numbers of programs and types of programs can help employees write their accomplishments at year-end Review. They can also help Management compare productivity between all employees in a group to help determine ranking for bonuses, although there are many variables that go into that which can't be captured here; outside projects, initiatives, Team Lead or Supervisory duties will of course decrease an employee's output but increase their value to the company.

I hope this will give you ideas about data that is already available at your company and can be used for many other purposes, or give you an idea about tracking your own projects in Microsoft Excel ® and reporting it using SAS ®.

ACKNOWLEDGMENTS

I would like to acknowledge my Supervisor, Annapurna "Anna" Ravi, for urging me to submit a paper to MWSUG and for providing feedback on the program and my Manager, Ramesh Mouly Satyavarapu, for his valuable input on this program.

RECOMMENDED READING

- Lex Jansen's website: http://www.lexjansen.com/
- SAS® Knowledge Base: http://support.sas.com/resources/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. This program is still a work-in-progress and any feedback and thoughts would be appreciated. I will be happy to share the code with you if you request it. Contact the author at:

Name:	Jean Crain
Enterprise:	InVentiv Health Clinical, LLC
City, State:	Indianapolis, Indiana
E-mail:	jean.crain@inventivhealth.com
Company Website:	http://www.inVentivHealthclinical.com