

An Extraordinary Reporting System for the Pharmaceutical Industry - the Macro Way

Judy L. Gilbert, Eli Lilly and Company

ABSTRACT

Automated Clinical Data Output, ACDO, is an Eli Lilly and Company project that has been dedicated to developing flexible, sophisticated SAS® macros which are used to create output across all disease classes. SAS macros are developed in a modular design so that multiple reports may use many of the same modules. The primary focus of the reports has been demographic, adverse event, and laboratory data, but some of the modules also analyze or summarize drug efficacy data. Although the listing programs will handle any clinical trial design, including crossovers, the summaries are generally applicable only for parallel designs.

A fundamental and powerful attribute of this reporting system is flexibility. The SAS macro facility is fully utilized. All reports contain multiple nested macros, SAS procedures, and functions. Customer-specified parameters number from 37 in the simplest report to over 200 in a more complex report. All parameters default so that the customer can submit a quick report with ease, or he can customize his report in an increasingly detailed way by changing parameter values. The categories of parameters include: patient population, time restriction, report specific processing, printing, and optional variable specifications.

The reports' functionalities extend from listings to extremely complex summaries that produce multiple different p-Values. Listings allow the ability to print different variables and request different paging schemes. The summaries vary in complexity: some use a simple PROC FREQ while others use the more complex PROC GLM.

The reporting system presently consists of 35 different report drivers. The flexibility of these 35 drivers, however, results in hundreds of variations of reports. There are approximately 200 different macros and about 320 customer parameters in this reporting system. The ultimate goal is to reduce the time to deliver generic and *ad hoc* reports needed for Lilly personnel, Lilly investigators, and outside regulatory agencies.

INTRODUCTION

Timely reporting of clinical trial data to regulatory authorities is an important step in remaining competitive in the pharmaceutical industry. Clinical trial reporting needs are diverse, complex and can change at a moment's notice. They diversify among the data as well as the customers. Adverse events, chronic and historical illnesses, laboratory, demographic, summary, and efficacy data are the major data categories. Data can also be reported from different sources, such as DB2 and traditional SAS data sets. Customers include in-house research scientists (physicians), statisticians, clinical research administrators, management, external local and global regulatory agencies, and external physicians who participate in clinical trials.

Ultimately the clinical trial reporting system must effectively support these diverse and complex requirements or else risk forfeiting or delaying all of the research and development effort that was involved in creating a new drug.

ACDO was originally conceived back in 1985, when a few of our associates saw the benefits of creating some sort of generic reporting system. In 1990 a systems development team was created and in 1991 we started actually coding the system. This development team consisted of four Lilly analysts (one of whom was project administrator) and in January of 1991 three outside contractors were added to the team. This paper focuses on a small fraction of the entire system.

Each report, or "product", of ACDO consists of the following:

- A "cover page" which details the value of each parameter for a specific execution of a report
- The actual report
- Corresponding system documentation
- JCL to execute the actual report
- Exception listings
- Explanation pages (customer focused descriptions of options available).

The cornerstone of ACDO is Quality. Each report was written by one person and double validated from a systems point of view by another analyst and then finally, given to the client for final review and validation. Because of Lilly's commitment to quality, the documentation and validation of these reports was an enormous task. All reports document "overall tests" (tests encompassing running all functions of the reports simultaneously) as well as "specific logic checks" (tests to validate a specific function within a report). Many hours were spent validating and attempting to break the SAS code. All to ensure one fact: accurate information to our customers!

ACDO development was greatly enhanced by a steering committee of associates representing all functional areas involved in achieving worldwide regulatory approval for a compound. They included:

- research scientists
- statisticians
- clinical research administrators
- regulatory
- medical writing
- system analysts.

This multi-faceted team was essential to the success of the project. Steering committee members spanned divergent functional disciplines and therapeutic classes. Early involvement by our associates began the quality process which enveloped the entire project. Additionally, they formed a crucial link for information sharing across the large Medical division within Eli Lilly and Company. The development team could focus more on creating the system, while our associates kept all areas abreast of the status and features of our project: they were our top notch sales representatives!

ACDO advantages include:

- ACDO is very well documented from both customer and systems viewpoints. Literally hundreds of validation tests were run on each report and each individual macro.
- ACDO allows for quicker analysis. ACDO saves time in output design, producing output, and from database lock to regulatory submission.
- ACDO facilitates design of output. Customers now have a starting point for designing output at time of protocol development.
- ACDO improves business partner relationships. Our associates and systems analysts have increased understanding of terminology and customer needs.
- ACDO increases consistency across studies and products, which allows for reduced review time and increased familiarity and understanding.
- ACDO's repetitive use has led to a high degree of confidence and understanding when reviewing the output reports by both customers and regulatory officials.

ACDO disadvantages include:

- The complexity of the parameters results in a steep learning curve. This is a concern for the new customers of the system.
- System enhancements are more time consuming. Validation tests necessary to complete a report enhancement increases the time to complete the change.
- A natural tendency to restrict the output to only that which ACDO can handle may occur.
- Misinterpretation could potentially occur due to the ability of non-statisticians to obtain reports with p-Values.

Optional variable processing: using the power of SAS macros

Optional variable processing allows the customer to specify data elements at their will. This functionality is used in six of the 35 reports. One summary report allows the customer to specify any 40 optional variables to summarize, and the method of summarizing them. There is also one listing report that allows the customer to specify any 20 optional variables to be listed. The optional variable logic is used consistently across the different reports. Once a requester learns how to specify a type of optional variable, he will know how to do this in any report that allows optional variable processing: with ACDO, customer knowledge is transferable!

This section will discuss the use of optional variables in summary reporting. This processing uses seven parameters to define one variable. They are:

VAR	Defines the 8 character SAS variable.
FILE	Defines the SAS input data set.
LABEL	Defines the column label to attach to the variable.
*FORMAT	Defines the format to print the value of the input variable.
VISIT	Defines the specific point in time to report.
TYPE	Defines what type of variable.
*RANGE	Defines the range.

* Only needed for certain types (defined by parameter TYPE).

The parameter TYPE defines the 'type of' required processing. Two examples follow:

TYPE = F

This directs the logic to do a simple PROC FREQ on the input element. The input elements are categorical in nature. When this is requested, the FORMAT parameter can also be specified. The FORMAT parameter defines the format used to print the values for this element. Please see Exhibit 1 for an example.

TYPE = S

This directs the logic to do simple statistics (count, mean, median, standard deviation, minimum, and maximum) on the input element. The input elements are continuous in nature. Please see Exhibit 1 for an example.

The above processing describes a report solely dedicated to optional variables. There are two other valid values for the TYPE parameter within ACDO presently. However, because of the modular design of ACDO, reports can easily be enhanced to include another type of analysis. SAS macros allow you to do this easily as each of the valid TYPEs is its own macro. Macro capabilities enable great flexibility within our system.

This concept of optional variable processing was again utilized on typical summary reporting. An example of a summary report that allows optional variable processing is the adverse event summary. In this summary, adverse events can be further analyzed by race and gender or by any two variables the requester specifies, including efficacy variables. Please see Exhibit 2 for an example.

Complex summary reporting: for any numeric, continuous variable.

One of the most complex reports available within ACDO is a summary of changes from baseline to endpoint analysis. This summary encompasses many p-Value options. A total of nine parameters direct the logic on how to compute and print the p-Values. Options include what the customer perceives as significant, what model statements within the PROC GLM to use, what error terms to use, how to group the analysis, what type sum of squares analysis to run. Explicit footnotes then detail to the customer which analysis was accomplished. Different analysis for each input variable occurs within a single execution of this report, dependent upon criteria such as the existence of empty cells.

This report can run with any numeric, continuous element: labs, vital signs, and efficacy variables are typical candidates. The customer defines the output format for the median, mean, and standard deviation values. The customer defines the baseline and the endpoint time intervals via visits or particular phases in a clinical trial. A customer can request analysis of the baseline to minimum and maximum value within the endpoint interval. Finally, the customer also has the ability to receive the analysis on the raw data elements or on ranked values (PROC RANK).

This report can also be used to analyze crossover study data (i.e., two compounds are given to a patient) if the data are restricted to one of the regimens. In crossover study analysis, the baseline visits could represent the first compound received by the patient and the endpoint visits could represent the second compound given to the patient. Please see Exhibit 3 for an example.

Printing of the report: everything is a macro!

Probably the most complex section of ACDO is the actual report printing. One high level driver macro prints the cover page and then calls each individual report's printing macro. A separate macro prints each item on the report: the titles, the paging variables, the column labels, the printing positions, the actual data lines to print, and the footnotes. Global macro variables define varying column widths, column order, and a maximum number of columns within each report.

The system differentiates between "fixed" versus "varying" columns. For example, if the report is an adverse event summary, the fixed column would be the event classification term and the varying columns would be the number of treatments that existed within the requested paging scheme. Thus, the printing logic creates global macro variables that designate the maximum number of fixed and varying columns for each execution of a report. The printing logic adds two SAS elements to the data set to be printed. For each specific row, one element contains the total number of variable columns to print and the second element contains the total number of columns, varying and fixed, to print. The values of these elements often vary when executing the reports across different paging schemes (i.e., across the entire clinical trial data versus across only a specific physician's data).

Finally, the actual PUT statement to print a line of the actual data looks as follows:

```
PUT &&_PSN&_I &&_COL&_I &_FMT @
```

where _I designates the number of possible columns to be printed for all rows within a specific page grouping,
 &&_PSN&_I designates the print position of the variable &_I,
 &&_COL&_I designates the actual variable to be printed and
 &_FMT designates any format associated with the printed variable.

A hard-coded string value such as "Tab" can be preassigned to the PSN macro variable. Thus, the entire report can contain embedded tab characters for automatic column alignment and can be easily downloaded into a PC word processing file. The resolution of this code is:

```
PUT @1 VISIT @
PUT \Tab VISIT @    (when hard-coded string value
                     is assigned).
```

CONCLUSION

The flexibility and diversification within the ACDO reporting system meets 50 - 90% of the reporting needs for the different research compounds. This system can and has been used for study reports, safety summaries, annual reports, and regulatory submissions. ACDO is available for hypothesis generation and *ad hoc* internal or external requests. In addition to the reporting flexibility, ACDO has been optimized to the extent that in a single day, a customer can run all 35 ACDO reports and variations of them. Historically, one report could take one to six weeks to code and validate!

Client involvement is a crucial component in project success. Our associates defined what the output should look like and what flexibility they wanted. The ACDO system is entirely modular, thanks to the SAS macro language support for parameter passing; ACDO as we know it would not exist without this feature. Flexibility to encompass any request was given top priority. ACDO provides top quality and quick analysis of data. It allows our associates to focus their time and efforts on less mundane tasks and thus greatly increases efficiency. ACDO helps decrease the time to get a new drug to the market -- and does it with elegance!

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Exhibit 1
A generic summary report of optional variables

Variable	Drug 1 (N=153)	Drug 2 (N=169)	Drug 3 (N=319)	Drug 4 (N=329)	Total (N=970)	p-Value
Gender: No.(%)						
Female	89 (58.2)	77 (45.6)	160 (50.2)	141 (42.9)	467 (48.1)	.013
Male	64 (41.8)	92 (54.4)	159 (49.8)	188 (57.1)	503 (51.9)	
Race: No.(%)						
African	37 (24.2)	15 (8.9)	42 (13.2)	51 (15.5)	145 (14.9)	.049
Asian	1 (0.7)	1 (0.6)	3 (0.9)	2 (0.6)	7 (0.7)	
Caucasian	104 (68.0)	144 (85.2)	249 (78.1)	253 (76.9)	750 (77.3)	
Hispanic	9 (5.9)	9 (5.3)	24 (7.5)	21 (6.4)	63 (6.5)	
Other	2 (1.3)	0	1 (0.3)	2 (0.6)	5 (0.5)	
Weight: kg (Visit 1)						
No. Patients	152	167	317	327	963	.041
Mean	73.70	78.49	76.09	77.89	76.74	
Median	71.11	78.93	75.30	77.11	75.75	
Standard Dev.	17.60	16.14	16.82	18.17	17.35	
Minimum	43.55	37.65	40.82	39.01	37.65	
Maximum	141.52	134.26	132.90	188.24	188.24	
Unspecified	1	2	2	2	7	

Notes: The 'Drug 1', 'Drug 2', etc. elements for the pharmaceutical industry are typically the different treatments being analyzed. The customer can label each of these elements in any way that can be displayed within 8 characters. In this example, 'Drug 1', 'Drug 2', etc. was used. Also, the customer can specify the order of the columns. The 'Total' column is always optional and can also be requested to be the only column printed. The 'p-Value' column is also optional. The 'Gender' and 'Race' variables are examples of TYPE=F processing. Also note that these are the labels defined by parameter LABEL. The 'Weight' variable is an example of TYPE=S processing. Also note Visit 1 was defined by parameter VISIT. 'Unspecified' indicates no data was available for this number of patients.

Exhibit 2
A typical adverse event summary enhanced with optional variable processing

Origin Column	Sex	Drug 1 N n (%)	Drug 2 N n (%)	Drug 3 N n (%)	Drug 4 N n (%)	Total N n (%)
Abdominal Pain						
African Descent	Female	0 0	0 0	1 0	2 0	3 0
	Male	3 0	0 0	3 0	0 0	6 0
Caucasian	Female	7 1 (14.3)	8 0	8 0	9 0	32 1 (3.1)
	Male	7 0	11 2 (18.2)	20 2 (10.0)	24 4 (16.7)	62 8 (12.9)
Hispanic	Female	0 0	0 0	1 0	0 0	1 0
	Male	2 0	0 0	3 0	0 0	5 0
Other	Female	0 0	0 0	1 0	0 0	1 0
Western Asian	Male	0 0	0 0	0 0	1 0	1 0
African Descent		3 0	0 0	4 0	2 0	9 0
Caucasian		14 1 (7.1)	19 2 (10.5)	28 2 (7.1)	33 4 (12.1)	94 9 (9.6)
Hispanic		2 0	0 0	4 0	0 0	6 0
Other		0 0	0 0	1 0	0 0	1 0
Western Asian		0 0	0 0	0 0	1 0	1 0
	Female	7 1 (14.3)	8 0	10 0	11 0	36 1 (2.8)
	Male	12 0	11 2 (18.2)	27 2 (7.4)	25 4 (16.0)	75 8 (10.7)
Total		19 1 (5.3)	19 2 (10.5)	37 2 (5.4)	36 4 (11.1)	111 9 (8.1)

Exhibit 3
Summary of Laboratory Changes from Baseline

Lab Test: Hematocrit

Lab Unit: Actual Number/Decimal

No.	Therapy	N	n	Baseline			Endpoint			Change		
				Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
1)	Drug 3	152	135	0.427	0.420	0.048	0.425	0.420	0.043	-0.002	0.000	0.035
2)	Drug 1	315	282	0.431	0.430	0.046	0.435	0.430	0.044	0.004	0.005	0.033
3)	Drug 2	323	281	0.433	0.430	0.044	0.433	0.430	0.041	0.000	0.000	0.031

----- p-Values ("Endpoint") -----
- Pairwise*3-

No.	Therapy	N	n	Within	Inter-		
				Group*1	action*2	Overall*2	vs.(2) vs.(3)
1)	Drug 3	152	135	.383	.495	.020	.006 .098
2)	Drug 1	315	282	.091			.141
3)	Drug 2	323	281	.717			

*1 The significance of a location shift from zero of the change from baseline within a treatment group is tested by the Wilcoxon Signed Rank procedure.

*2 Type III Sums of Squares from an analysis of variance (ANOVA) on the ranks: PROC GLM model=investigator and treatment for the overall p-Value and model=investigator, treatment, and interaction for the interaction p-Value.

*3 Least-squares mean option in PROC GLM from the ANOVA on the ranks using the mean square for error.