

# A Language for Computer Generation of Medical Data Forms Author(s): J. A. Hanley Source: *Biometrics*, Vol. 34, No. 2 (Jun., 1978), pp. 288-297 Published by: International Biometric Society Stable URL: http://www.jstor.org/stable/2530019 Accessed: 26-01-2017 18:37 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms



International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to Biometrics

BIOMETRICS 34, 288–297 June, 1978

# A Language For Computer Generation Of Medical Data Forms

# J. A. HANLEY<sup>1</sup>

Statistical Laboratory, State University of New York at Buffalo, 4230 Ridge Lea Road, Amherst, New York 14226, U.S.A.

#### Summary

In this paper we describe the main features of a small-scale plotting language used to produce medical data forms. It requires no previous knowledge of computers and is easily learned and used. Its use at the Statistical Laboratory at the State University of New York at Buffalo has resulted in considerable time savings, a greater unification of data items and data forms, and an increase in the quality of data collected.

## 1. Introduction

Until recently, applied statisticians have concentrated their efforts on the proper design of experiments and on the methodology for analysis and presentation of results. Less attention has been paid to the formal procedures for the execution of experiments and to proper specification of what data are to be recorded and how they are to be collected. Nevertheless, study data and conclusions derived therefrom are only as good as the method by which data are collected. For this reason, the success of any large-scale research project, particularly when it is a cooperative one, may well depend on the degree of attention paid to an often-neglected aspect of data collection, namely the proper design of data reporting forms. This report shows how the computer can aid the statistician, and those who collaborate with him, in the design and production of forms.

## 2. Background

The application to be described arose from the needs of the Statistical Laboratory in its role as a statistical coordinating center for five national cooperative groups which conduct clinical trials in cancer. In each trial, participating investigators treat patients according to agreed-on procedures set out in a protocol and submit data on a routine basis to the central data-coordinating center, in our case the Statistical Laboratory.

The large number and variety of trials (more than 100, involving more than 7,000 patients per year from some 250 U.S., Canadian and overseas institutions) are constantly monitored and, for most, early dissemination of the study results is required. In the interests of efficiency, it is essential that the data be (a) submitted according to strict reporting criteria and conventions, and (b) in a form which lends itself easily to computer entry (i.e., so that any further abstraction at the central office is avoided) and retrieval. Currently, the data from

<sup>1</sup> Present address: Division of Biostatistics and Epidemiology, Sidney Farber Cancer Center, 44 Binney St., Boston, Mass 02115

Key Words: Clinical trials; Data forms; Computer language.

## COMPUTER GENERATION OF MEDICAL DATA FORMS

each institution are submitted by mail using data-reporting forms. These forms vary considerably in nature depending on the disease being studied, the nature of the treatments being tested and the criteria by which they are to be compared. Typically, for any study, there are several forms including (a) an Initial Registration Form, to identify the patient entry, (b) an On-Study Form, to report important baseline characteristics of a patient and his disease, (c) a Treatment Form, to report treatment administered, reactions encountered and supportive care required, (d) a Follow-up Form, to report periodically on a patient's progress and response to treatment, and (e) a Summary Form, to note the final disposition of a case and the outcome of treatment. The participating institution retains one copy of each completed form. Copies are distributed to a Study Chairman and the Statistical Center.

The most direct way to produce forms is to use a regular typewriter to develop an original which is then mass-produced by photo-offset. However, when large amounts of information are being sought on large numbers of study cases, and especially when forms must be completed in triplicate, this method is inefficient. Instead, both duplication costs and storage space are considerably reduced if forms can be produced on a varitype machine and then collated into multipage no-carbon-required (NCR) sets. However, the varitype process is both lengthy and technical and involves considerable planning and considerable time to execute. Moreover, this method does not lend itself to incorporation of last-minute changes (a common occurrence, when forms have to be sanctioned by several participating centers).

Because of these difficulties the staff at the Statistical Laboratory at Buffalo experimented with computer generation of forms, using an incremental line-plotter. As a result of this, a simple, flexible and easy-to-use language now exists which completely eliminates the varitype process and considerably reduces the time between a form's conception and its availability for use. The main features of this language are described in the remainder of this paper.

#### 3. The Language

#### 3.1 Overview

The language consists of 33 "commands" (see Table 1). Associated with each command are the parameters (e.g., the dimensions of a page) necessary to carry out the prescribed task. Commands can be grouped into 3 categories: (a) those which cause direct pen action, e.g., produce boxes, lines or text, (b) those which change some setting such as letter size, pen location or box size and (c) those which store sets of commands for subsequent use. To produce a form one builds a data file (on cards or on disk) of the commands and parameters required to produce the desired boxes, lines and text which constitute the form. One then runs a computer program called FORMS which reads this data file and produces (i.e., plots) the form.

## 3.2 General Operation and Features

• The PAGE command, normally the first one issued, draws the page outline, establishes the bottom left-hand corner as the origin of the pen's coordinate system, and positions the pen at the top left-hand corner of the page in readiness for the next command.

• The pen can be moved vertically and horizontally by use of the DOWN and AD-VANCE commands (both of which move the pen relative to its current position) or absolutely using the X and Y commands, i.e., by specifying actual coordinates.

• Text strings, lines and boxes, are produced starting at the current pen position. Following their completion, the pen position is updated.

• Sets of boxes, suitable for the recording of dates (i.e., MM DD YY or MM YY) can be produced using the single DATE command.

	I	ne commana	561	
COMMAND	PARAMETER(S)*	INITIAL SETTING	LIMITS	ACTION
PAGE	width height	8.5 11.0	36.0 36.0	Produces a page outline of the (last) specified dimensions. positions the pen at the top left hand corner of the page.
`STRING'			150 characters	Produces the specified text string.
BOX	number of boxes	1	1 to 99	Produces the (last) specified number of boxes
DATE	number of sets	3	2, -2, 3, -3	Produces the (last) specified number of sets (2 boxes) of boxes (use—for MY or MDY).
LEND	where line ends	8.25	-20.0 to 20.0	Draws a horizontal line to the (last) specified ending point.
LLENgth	length	1.0	-20.0 to 20.0	Draws a horizontal line of the (last) specified length.
ADVAnce	amount	0.0	20.0 to 20.0	Moves the pen to the right/ left by the (last) specified positive/negative amount.
DOWN	amount	0.25	-36.0 to 36.0	Moves the pen down/up by the (last) specified positive/ negative amount.
х	position	0.25	-20.0 to 20.0	Moves the pen to the (last) indicated horizontal position.
Y	position	11.00	<b>36</b> .0 to 36.0	Moves the pen to the (last) specified vertical position.
CENTer	left border right border	page width	L/R -20.0 to 20.0	Centers the next rext string relative to the (last) specified borders.
SIZE	BOX-WIDTH BOX-HEIGHT	0.16 0.18	36.0 36.0	Changes the width and height of future boxes
LETTer	LETTER HEIGHT	0.10	0.01 to 10.0	Changes the height of future letters.
FACTor	SCALE FACTOR	1.0	0.1 to 10.0	Enlarges or reduces the entire plot by the specified scale factor.
REPEat	NUMBER OF TIMES	1	1 to 4	Causes future plotting to be "written over" a specified number of times.

Table 1 The Command Set

COMMAND	PARAMETER(S)*	INITIAL SETTING	LIMITS	ACTION
COLOr	COLOR CODE	1	1 to 3	Changes the ink color (1 = black, 2 = red, 3 = blue)
XSTOre	LOCATION NUMBER		1 to 20	Stores the pen's current horizontal position into the specified X location.
XRECall	LOCATION NUMBER		1 to 20	Restores the pen's horizontal position to that stored in the specified X location.
YSTOre	LOCATION NUMBER		1 to 20	Stores the pen's current vertical position into the specified Y location
YRECall	LOCATION NUMBER		1 to 20	Restores the pen's vertical position to that stored in the specified Y location.
SPACe	number of spaces	1	—150 to 150	Causes pen to move right (+ve)/left (-ve) the (last) specified number of spaces.
LSTOre	LOCATION NUMBER		1 to 20	Stores the current letter height into the specified L location
LRECall	LOCATION NUMBER		1 to 20	Restores the letter height to that stored in the specified L location.
XSHI <del>fi</del>	AMOUNT	0.0	-20.0 to 20.0	Causes future plotting to be shifted horizontally by the indicated amount (+ = right).
YSHIft	AMOUNT	0.0	-36.0 to 36.0	Causes future plotting to be shifted vertically by the indicated amount $(+ = up)$ .
NIOL	LOCATION # location #		1 to 20 1 to 20	Draws a vertical line from the current pen position to the position stored in the indicated Y location. If a second vertical location is specified, the line joins the 2 specified Y loca- tions. At completion, pen is restored to its precommand position.
OUTEr	NUMBER		1 to 20	Saves the subsequent com- mands (those preceding END) for later use.
END				Denotes the end of the set of commands to be stored.

Table 1 Continued

COMMAND	PARAMETER(S)*	INITIAL SETTING	LIMITS	ACTION
INNEr	NUMBER		1 to 20	Saves the subsequent com- mands (those preceding END) for later use.
QUOTe	'NEW QUOTE INDICATOR'	,		Changes the special character which is used to enclose text strings.
UPPEr	'NEW UPPER-CASE INDICATOR'	`>'		Changes the special character which, when used in a text string, indicates a switch back to upper-case characters.
LOWEr	'NEW LOWER-CASE INDICATOR'	`<'		Changes the special character which, when used in a text- string, indicates a switch to lower-case characters.
SKIP	'NEW END-OF-LINE INDICATOR'	`&'		Changes the special character which is used to denote the end of each line of input.
OUTEr	NUMBER NUMBER-OF-TIMES		1 to 20 1 to 100	Executes the specified OUTER loop the specified number of times
INNEr	NUMBER NUMBER-OF-TIMES		1 to 20 1 to 100	Executes the specified INNER loop the specified number of times.

Table 1 Continued

\* Parameters in UPPER CASE are MANDATORY, parameters in lower case are optional

• At any stage, the sizes of boxes and letters can be changed as desired.

• The pens's current vertical or horizontal position can be stored for later recall. Similarly letter sizes can be stored and recalled.

• Entire sections can be moved from one part of a form to another using the XSHIFT and YSHIFT commands.

• Parameters need not be repeated unless their values are to be changed; they will instead default to the most recently specified values. This feature makes the sequence X DOWN equivalent to a carriage-return on a typewriter.

• The placement of commands is entirely format free, and multiple commands can be placed on a single line. The only restriction is that each command and each parameter must be separated by at least one blank space.

• One can abbreviate commands (e.g., P is sufficient to signify PAGE since there is no other command beginning with the letter p).

#### 3.3 Special Features

One feature deserves special explanation. Because forms often contain repeated patterns, a special command is included which frees the user from having to repeatedly issue the same

sets of commands and from having to make numerous corrections when revising a form. To take advantage of this feature, one first defines and stores various master patterns. To invoke a pattern at a later juncture one simply specifies the pattern number and the number of times it is to be executed. Often one wishes to vary a text string each time the pattern is executed. To do this one designs and stores a pattern using a null text string; the real text strings are substituted at the time the pattern is executed. Pattern may be nested in that an (OUTER) pattern may contain one or more previously defined (INNER) patterns.

## 4. The Production Process

The user, working from a rough sketch, builds the file of commands and parameters. He then submits this file to be processed by the FORMS program which generates the form(s). In addition to the plotted form(s), the program also produces a print-file which echoes each command. This enables the user to rectify any mistakes quickly or make other changes before resubmitting a job. Once a set of forms is completed, the command file can be stored to await any further revisions. We have found that using a text editor to build and alter a command file, and a CRT graphics terminal to preview each form, we can produce a final copy in less than a day.

Once a final version is agreed upon, the pages are plotted in enlarged versions, using highquality paper and liquid ink. Magnification is obtained by inserting a single FACTOR command before the final run. These large-scale versions are then photographically reduced to offset any deficiency caused by poor resolution. At this stage, the pages are ready to be printed.

## 5. An Example

It would be tedious to discuss each command available with FORMS. However, the following simple example illustrates the simplicity and usefulness of the process. The form presented in Figure 1 was produced by the commands shown in Table 2. A sample of a data form which was generated by this process, and which is in current use in cooperative studies, is shown in Figure 2.

# 6. Hardware and Software Considerations

## 6.1 Software

FORMS, used to produce the forms illustrated in Figures 1 and 2, is written in FOR-TRAN, with no machine dependencies other than a) DECODE statements, b) END-OF-FILE checks, and c) assignment of I/O unit numbers. All plotting is achieved by just two plotting subroutines: PLOT(X, Y,SWITCH) which causes the pen to move to the point (X, Y), with the pen on or off the paper and SYMBOL(X, Y,height, text string) which produces the required text strings. Both these subroutines are relatively standard among the major plotting packages to be found at computer installations. As currently written, FORMS will compile and run with 70,000 (octal) words of central memory, under the NOS 1.2 Operating System in use at the CDC CYBER 173 installation at the State University of New York at Buffalo.

The program did necessitate some special features. Since the 64 characters of the CDC character set do not allow for easy input and processing of both upper and lower case characters, two special characters which must be imbedded in the text string are reserved to indicate changes to upper and lower case letters for printing. Also, the absence of lower case

LABORATORY REPORT
INSTRUCTIONS: This form is to be completed and submitted within 1 week of each scheduled test.
PATIENT IDENTIFICATION
LABORATORY DATA
INVESTIGATOR: SEND TOP 3 COPIES TO DATA CENTER
STATLAB 2/77

*Figure 1* The form produced by the commands in Table 2.

#### COMPUTER GENERATION OF MEDICAL DATA FORMS

Table 2The Command Used to Produce Figure 1

```
OUTER 1 DOWN 0.10 X 0.0 LEND 3.0 DOWN .25 XR 1 END
OUTER 2 LS 1 LET 0.09 C 0.0 3.0 '' LR 1 DOWN 0.3 XR 1 END
OUTER 3 BOX 4 '' X 1.4 LEND 2.8 DOWN 0.09 C 1.4 2.8 ''
                                                      XR 1 END
OUTER 4 BOX 2 '.' BOX 1 SP '' XR 1 DOWN 0.22 END
P 3. 5. DO .25 LET .12 C 'LABORATORY REPORT' DO X .2 XST 1
LET .065 'INSTRUCTIONS: T<HIS FORM IS TO COMPLETED AND'
DO .15 X '< SUBMITTED WITHIN 1 WEEK OF EACH SCHEDULED TEST.'
OUTER 1 1 OUTER 2 1 'PATIENT IDENTIFICATION'
OUTER 3 2 ' STUDY''PATIENT NAME' ' CASE NO.' 'INSTITUTION'
OUT 1 1 OUT 2 1 'LABORATORY DATA' DA -3 ' DATE SAMPLE TAKEN'
     X DO .4 B 3 ' PLATELET COUNT (<IN 1,000S)'
     X DO .22 O 4 5 'H<GB GM%' 'H<CT %' 'WBC (<IN 1,000S)'
                  'N<EUTROPHILS %' 'L<YMPHOCYTES %'
OUTER 1 1 C 'INVESTIGATOR: SEND TOP 3 COPIES TO DATA CENTER'
Y .25 X .0 LEND 3.0 X 2.2 Y .15 'STATLAB 2/77'
```

symbols in the CALCOMP plotting package required the writing of a special FORTRAN subroutine to plot the 26 lowercase letters.

## 6.2 Hardware

The SUNYAB Computing Center currently employs a CALCOMP 925/936 Plotter, which can produce plots as high as 36 inches and of even greater width. With a resolution of .002 in. (i.e., 500 pen moves/inch), a  $17'' \times 22''$  plot reduced by 1/2 can provide a final product which is more than acceptable for printing. Even if such a large plotter were not available, half pages, produced at twice scale on an 11'' plotter, produce an acceptable resolution.

#### 7. Conclusion

This paper has attempted to describe the main features of a small-scale plotting language suitable for the production of medical data forms. Its use at the Statistical Laboratory has resulted in considerable time savings, a greater unification of the data items and data forms, and an increase in the quality of the data collected. The computer program, together with documentation on how to implement and use it, is available upon request from the Director, Division of Biostatistics and Epidemiology, Sidney Farber Cancer Center, 44 Binney Street, Boston, Massachusetts 02115.

## **BIOMETRICS**, JUNE 1978

STAGE II CARCINOMA OF 1 INSTRUCTIONS: This form is to be completed and submitted /evaluation form upon completion of all protocol-related	THE OVARY ON-STUDY/SURGERY FORM Page 1 of 1 together with a "Stage IJ Carcinoma of the Ovary" follow-up surgery. Unless otherwise specified, use "N/A", "N/D", or "UNKN"			
to indicate "not applicable", "not done", or "unknown" re	respectively. Do not write in shaded boxes.			
Patlent's Name				
<pre>Birth date (m,y) Height (cm) Current body weight (kg) Body weight before onset of disease (kg) Surface area (m<sup>2</sup>) Race (1-white, 2-black, 3-other)</pre>	Previous malignancies other than skin cancer (1-no, 2-yes, 9-unknown) if YES.specify Previous radiotherapy (1-no, 2-yes. 9-unknown) Previous chemotherapy (1-no, 2-yes. 9-unknown)			
SURGERY (as required by protocol) M D Date of initial surgery (m,d,y) Did initial surgery fulfill protocol requirements? (1-no, 2-yes, 9-unknown) Date of supplemental surgery (if per	DISEASE CLASSIFICATION Disease stage (consult protocol for definitions) 10-stage I Ai 15-stage II A 11-stage I Aii 16-stage II B 12-stage I Bi 17-stage II C 13-stage I Bii 18-stage III 14-stage I C 19-stage IV 09-unknown			
Procedures (1-no, 2-yes, 9-unknown)  removal of left ovary and tube removal of right ovary and tube total abdominal hysterectomy omentectomy biopsy of omentum biopsy of all suspicious nodules aspiration of peritoneal fluid examination of all peritoneal fluid	Histologic type l=serous 2-mucinous 3-endometrioid 4-clear cell 5-other.epecity 9-unknown Histologic grade l=vell differentiated 2-moderately differentiated 3-poorly differentiated 9-unknown			
<pre>other.specify</pre>	d vith this form ] 			

Cooperative Group

Investigator (please print)

Date (m.d.y)

Investigator: Keep last copy and send remaining top copies to Operations Office.

-----

RTOG - 3/77

*Figure 2* A form used in clinical trials.

### COMPUTER GENERATION OF MEDICAL DATA FORMS

#### **Acknowledgments**

The author wishes to thank Gregory Pavlov and Stephen Gould for their part in this work, all of which was supported by grant R10CA12721 from the National Cancer Institute. He would also like to thank Colin Begg for his assistance with the manuscript.

# Résumé

Dans cet article, nous décrivons les traits principaux d'un langage graphique de petite échelle, utilisé pour produire les éditions des données médicales. Il n'exige pas de connassiance préalable des ordinateurs, il est facile de l'apprendre et de l'utiliser. Son emploi au laboratoire Statistique de la State University de New-York à Buffalo, a entraîné de considérables économies de temps, une plus grande uniformité des items et des éditions des données, un accroissement de la qualité des données recueillies.

Received April 1977; Revised January 1978