NONSTANDARD EDITING TECHNIQUES FOR CLINICAL DATA

Richard A. Enz, Doron Steger, John Sanocki, Hoechst-Roussel Pharmaceuticals, Inc.

ABSTRACT

In clinical research, the volume of data is often large and the types encountered vary widely. Standard editing techniques (i.e. range checks, format checks) will assure that the correct type of data is in the right place. However, minor errors that pass range and format checks can drastically affect results of analysis. To prevent this, we have developed some nonstandard techniques for editing data, using SAS.

The flexibility of SAS allows us to produce specialized listings and scatter plots that point out possible errors and inconsistencies in a fashion facilitating easy review.

Some of the techniques used are:

- listings that indicate the subject's adherence to the visit schedule
- scatter plots that display large amounts of data on a single page
- scatter plots that highlight extreme deviations from the sample population
- dosage listings that indicate whether subjects followed the study protocol
- listings of related subject data from different records combined in an easy-to-read format to simplify checking for inconsistencies.

With these techniques, many hidden errors are found and corrected before the analysis begins. This simplifies the review process and greatly reduces the need to rerun analyses because of subtle errors in the data.

BACKGROUND - STANDARD EDITING TECHNIQUES

Usually the first program employed for reviewing clinical data is a field-by-field editor. Variations of this program abound within the pharmaceutical industry, but most generate the following information:

- formatted listing of the data (often decoded where applicable)
- flags for out-of-range data
- flags for incorrect data types
- flags for missing data
- summary of errors

At Hoechst-Roussel, the clinical system runs off a data dictionary[1,2] that interfaces with many programs, including the editor (Figure 1). The data dictionary is created for each study and is generated prior to data entry. Nontechnical personnel can easily generate edit listings using this system. The output from the editor is in two parts: The first page contains a column-by-column description of each field in the record (Figure 2), and the subsequent pages list each field of a record and applies the editing rules described above to each subject's data (Figure 3). For many years this program was the only one used for editing data. Unfortunately, data errors slipped through, only to be found later, when the final reports had been produced. Weaknesses of the editor included:

- no crossfield editing capabilities
- tedious review of huge studies
- no display of the visit structures
- no flags for gross changes in subject's values

NONSTANDARD EDITING TECHNIQUES

The following SAS programs were developed to overcome these limitations.

Scatter Diagrams

One of the most useful editing techniques at Hoechst-Roussel is the scatter diagram using SAS's PROC PLOT. Hundreds of points can be displayed on one page and the outliers are easily spotted. The larger the study under investigation, the more useful the scatter diagram becomes. Two types are generated:

The first displays a plot of the data (Figure 4). A listing of the raw data and a listing of the subjects having the 'N' highest and lowest data values (Figure 5) are also generated. Typically, 10 high and 10 low values are listed, but this can be varied.

The second scatter diagram plots the difference between a single observation and the mean of all observations for a subject (Figure 6). This graph shows subject trends that may not be detected by the first graph. A subject who had a low value and later had a high value, both within the edit range, would probably be detected by this graphing technique.

A gross change score may indicate that a data point for a subject was in error or that an actual clinical change had occurred. The 'N' highest and lowest change values are also generated (Figure 5).

Generalized Inventories

Our generalized inventories have become a major part of the data review. Decisions on status of subjects are made using this program. Two types are available (Figures 7,8):

The first inventory lists on one page all data relevant to one subject.

- listings of missing forms
- adherence to visit schedule
- multiple observations per visit
- dosage schedule (any interruptions)

The second inventory provides the statisticians with a listing of analyzable data for selected variables. At each visit, the program indicates whether the data is analyzable and if not, provides the reason for exclusion. Totals at the end provide a cross-check against any efficacy and safety output.

Dosage Records

One of our most important clinical trial records is the dosage record. Most of our other programs link to this record in order to determine start of study, end of study, dosage, and visit structures. Because of its importance, an edit program is developed specifically for the dosage record (Figure 9). This program provides the mean, minimum, maximum, and most common daily dose and the number of days on treatment for each week the subject was in the study. It will also show any problems with the dates, discontinuation of treatment in the middle of a study, or dosage that was either too low or too high (in realtion to the study protocol).

Special Edits

Occasionally it is necessary to create a special edit for a specific drug under investigation. We try to tailor such an edit to a specific final output. Figure 10 provides an example of this type of edit. This program utilizes the graphics capabilities of the IBM 3800 printer. It creates a file with appropriate graphic control characters that is then printed on the 3800 printer. The program combines related information from different records and displays it in an easy-to-read form, which allows us to verify very quickly that related information is consistent and available. Decisions about status can be easily made by the medical reviewer.

CONCLUSION

The complexity of today's clinical trials with the vast amount of collected data requires new methods of data review. The traditional editing systems are still useful, but advances in computer technology have made possible more sophisticated and discriminating editing techniques.

ACKNOWLEDGEMENTS

Our thanks to Ms. Pamela Sleeper for doing much of the programming, Ms. Helle-Mai Gawrylewski for her careful review and Ms. Betty Casterline for her accurate typing.

REFERENCES

1. Steger, D. and Enz, R.: Data Dictionary Driven Data Analysis System, Proceedings of the Fourth Annual SAS Users Group International Conference. Clearwater, FL. January 29, 1979.

2. Herbel, E.S.: A Data Dictionary/Directory Driven Clinical Management System, Proceedings of the Hewlett Packard General Systems International Users Group, Denver, CO. October 30,1978.

FIGURE 1

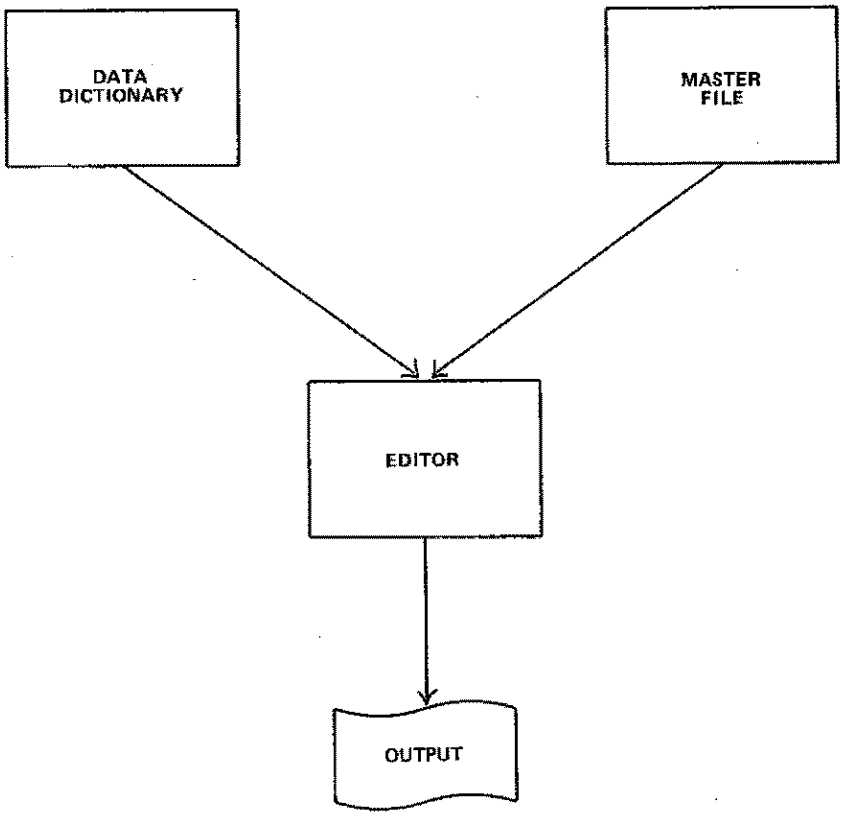

FIGURE 2

PROJECT :B PROTOCOL:304 RECORD# :400

PRINTNAME	LOCATN	DESCRIPTION	*EDIT CRITERIA* MIN	MAX
DATE	21, 6	DATE SAMPLE WAS TAKEN	NO EDIT	
TIME	27, 4	TIME SAMPLE WAS TAKEN	NO EDIT	
HEMATOCR	31, 4	HEMATOCRIT (PERCENT)	37.0000	52.0000
HEMOGLOB	35, 4	HEMOGLOBIN (gm/dl)	12.6000	18.0000
RBC	39, 4	RED BLOOD CELL (Million/cu mm)	4.0000	6.2000
WBC	43, 4	WHITE BLOOD CELL (Thousand/cc mm)	4.5000	11.0000
PLATELET	47, 3	PLATELETS (Thousand/cu mm)	200.0000	500.0000
PLATEST	50, 1	PLATELET ESTIMATE:1=NORM 2=INCR 3=DECR	DECODE:PLATCODE 20	

FIGURE 3

PROJECT :B PROTOCOL:304 RECORD# :400

INVEST	PATNO	VISIT	OBSERV	DATE	TIME	HEMATOCR	HEMOGLOB	RBC	WBC	PLATELET	PLATEST
				21	27	31	35	39	43	47	50
031	0301	01		011180	1000	38.9	13.4	4.33*	04.0*	198*	NORMAL
031	0301	03		012880	1100	40.0	13.4	4.42*	04.1*	234	NORMAL
031	0301	04		021280	1000	____	____	____	____	____	_
031	0301	05		022680	1000	____	____	____	____	____	_
031	0301	08		041080	1030	39.8	13.5	4.42*	04.2*	200	NORMAL
031	0301	11		052080	0945	39.2	13.1	4.30*	04.0*	233	NORMAL
031	0301	16		060580	0830	37.2	12.9	4.10*	04.0*	194*	NORMAL
031	0301	18		080780	0900	39.1	13.1	4.41*	04.1*	199*	NORMAL
031	0301	44		100380	0930	39.1	13.5	4.45*	03.7*	___	NORMAL
031	0302	01		012580	0930	38.2	12.8	4.62*	04.3*	___	NORMAL
031	0302	03		021580	0915	40.6	13.5	4.93*	04.2*	286	NORMAL
031	0302	05		031380	1015	____	____	____	____	____	_
031	0302	08		042480	0945	37.8	12.7	4.58*	04.3*	271	NORMAL
031	0302	11		060580	0930	38.7	13.0	4.63	04.9	291	NORMAL
031	0302	14		061980	0915	39.4	13.3	4.72*	03.9*	289	NORMAL
031	0303	01		012580	1100	39.7	13.7	4.31	04.6	337	NORMAL
031	0303	03		021580	1015	42.6	14.6	4.64	04.9	300	NORMAL
031	0303	05		031980	0930	____	____	____	____	____	_
031	0303	08		060380	0945	40.5	13.7	4.34	05.1	348	NORMAL
031	0303	11		072180	0900	44.0	14.0	4.59	06.4	337	NORMAL
031	0303	15		081880	1000	41.2	14.2	4.51	04.8	324	NORMAL
031	0304	01		013180	1030	38.7	13.2	4.53	05.1	244	NORMAL
031	0304	03		022080	0930*	36.0*	12.4*	4.26	04.8	275	NORMAL
031	0304	05		031880	1000	____	____	____	____	____	_
031	0304	08		050180	1030*	34.8*	12.1*	4.12	05.9	271	NORMAL
031	0304	11		061380	0930*	35.7*	12.3*	4.17	05.2	239	NORMAL
031	0304	16		062680	0900	40.3	13.2	4.66	05.9	289	NORMAL
031	0304	18		073180	0945	37.8	12.8	4.59	05.8	266	NORMAL
031	0304	44		092680	0945*	35.5*	12.2*	4.24	05.6	245	NORMAL
031	0305	01		060280	0945	49.7	16.2	6.07	04.7	231	NORMAL
031	0305	03		062780	1100	46.0	14.8	5.54*	03.9*	238	NORMAL
031	0305	05		081980	1145	____	____	____	____	____	_
031	0305	08		100980	0930	48.5	15.8	5.99	04.9	247	NORMAL
031	0321	01		020580	1000	41.0	13.6	5.01	08.8	351	NORMAL
031	0321	03		022780	1015	38.6	13.0	4.68	07.4	337	NORMAL
031	0321	05		032680	1050	____	____	____	____	____	_
031	0321	08		050880	0900	39.7	13.6	4.83	08.9	292	NORMAL
031	0321	08		062580	____	____	____	____	____	____	_
031	0321	11		061980	1015	40.3	13.6	4.87	08.5	285	NORMAL
031	0321	13		062680	0930	41.9	13.7	4.96	07.8	333	NORMAL
031	0321	15		080780	0945	40.8	13.8	5.03	07.8	284	NORMAL
031	0321	45		100180	1000	41.6	14.0	5.09	07.4	290	NORMAL

FIGURE 4

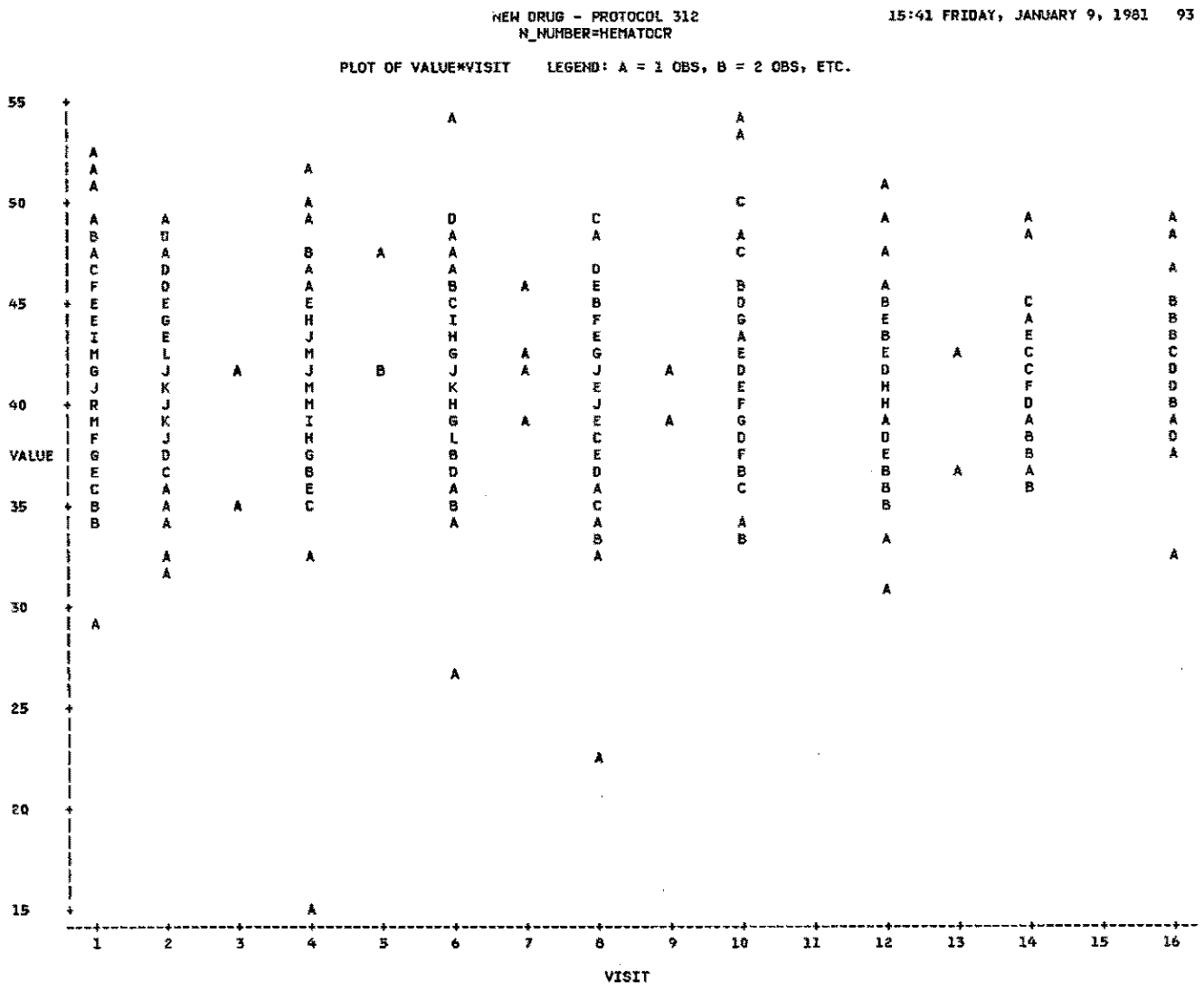
NEW DRUG - PROTOCOL 312
N_NUMBER=HEMATOCR
15:41 FRIDAY, JANUARY 9, 1981 93
PLOT OF VALUE*VISIT LEGEND: A = 1 OBS, B = 2 OBS, ETC.
VALUE
VISIT

FIGURE 5

NEW DRUG - PROTOCOL 312

REC	PROT	INV	VIS	SUBJ	VARIABLE	VALUE	
400	312	36	4	15	HEMATOCR	15.4	*
400	312	32	8	23	HEMATOCR	22.8	*
400	312	32	6	23	HEMATOCR	26.7	*
400	312	34	1	4	HEMATOCR	29.3	*
400	312	34	12	4	HEMATOCR	31	*
400	312	32	2	23	HEMATOCR	32	*
400	312	34	2	4	HEMATOCR	32.2	*
400	312	32	4	23	HEMATOCR	32.3	*
400	312	34	8	19	HEMATOCR	32.4	*
400	312	32	16	23	HEMATOCR	32.8	*
400	312	33	10	4	HEMATOCR	50.1	**
400	312	32	10	14	HEMATOCR	50.2	**
400	312	36	4	4	HEMATOCR	50.2	**
400	312	33	1	7	HEMATOCR	50.6	**
400	312	32	12	6	HEMATOCR	50.8	**
400	312	32	4	6	HEMATOCR	51.8	**
400	312	36	1	24	HEMATOCR	51.8	**
400	312	34	1	18	HEMATOCR	52.7	**
400	312	33	10	3	HEMATOCR	53.3	**
400	312	32	6	6	HEMATOCR	54.4	**
400	312	32	10	6	HEMATOCR	54.4	**
400	312	33	10	4	HEMATOCR	-7.74444	#
400	312	33	10	3	HEMATOCR	-7	#
400	312	34	1	6	HEMATOCR	-6.13333	#
400	312	36	10	15	HEMATOCR	-6.05714	#
400	312	36	8	15	HEMATOCR	-5.25714	#
400	312	36	2	15	HEMATOCR	-5.15714	#
400	312	37	8	5	HEMATOCR	-4.62857	#
400	312	33	10	6	HEMATOCR	-4.75556	#
400	312	33	8	17	HEMATOCR	-4.3625	#
400	312	36	6	15	HEMATOCR	-3.85714	#
400	312	33	2	8	HEMATOCR	3.514286	##
400	312	37	13	9	HEMATOCR	3.6875	##
400	312	34	4	2	HEMATOCR	3.8125	##
400	312	34	8	19	HEMATOCR	4	##
400	312	36	4	24	HEMATOCR	4.05	##
400	312	32	1	8	HEMATOCR	4.54	##
400	312	32	8	1	HEMATOCR	4.6375	##
400	312	36	8	28	HEMATOCR	5.342857	##
400	312	33	16	7	HEMATOCR	6.475	##
400	312	32	8	23	HEMATOCR	7.283333	##
400	312	36	4	15	HEMATOCR	26.24286	##
400	312	32	6	4	HEMOGLOB	1.74	*
400	312	32	8	23	HEMOGLOB	6.8	*
400	312	32	6	23	HEMOGLOB	8	*
400	312	32	2	23	HEMOGLOB	9.8	*
400	312	32	4	23	HEMOGLOB	10.1	*
400	312	34	2	4	HEMOGLOB	10.2	*
400	312	34	12	4	HEMOGLOB	10.2	*
400	312	34	1	4	HEMOGLOB	10.3	*
400	312	32	16	23	HEMOGLOB	10.4	*
400	312	36	2	37	HEMOGLOB	10.4	*

10 LOWEST

10 HIGHEST

10 LOWEST CHANGE SCORES

10 HIGHEST CHANGE SCORES

FIGURE 6

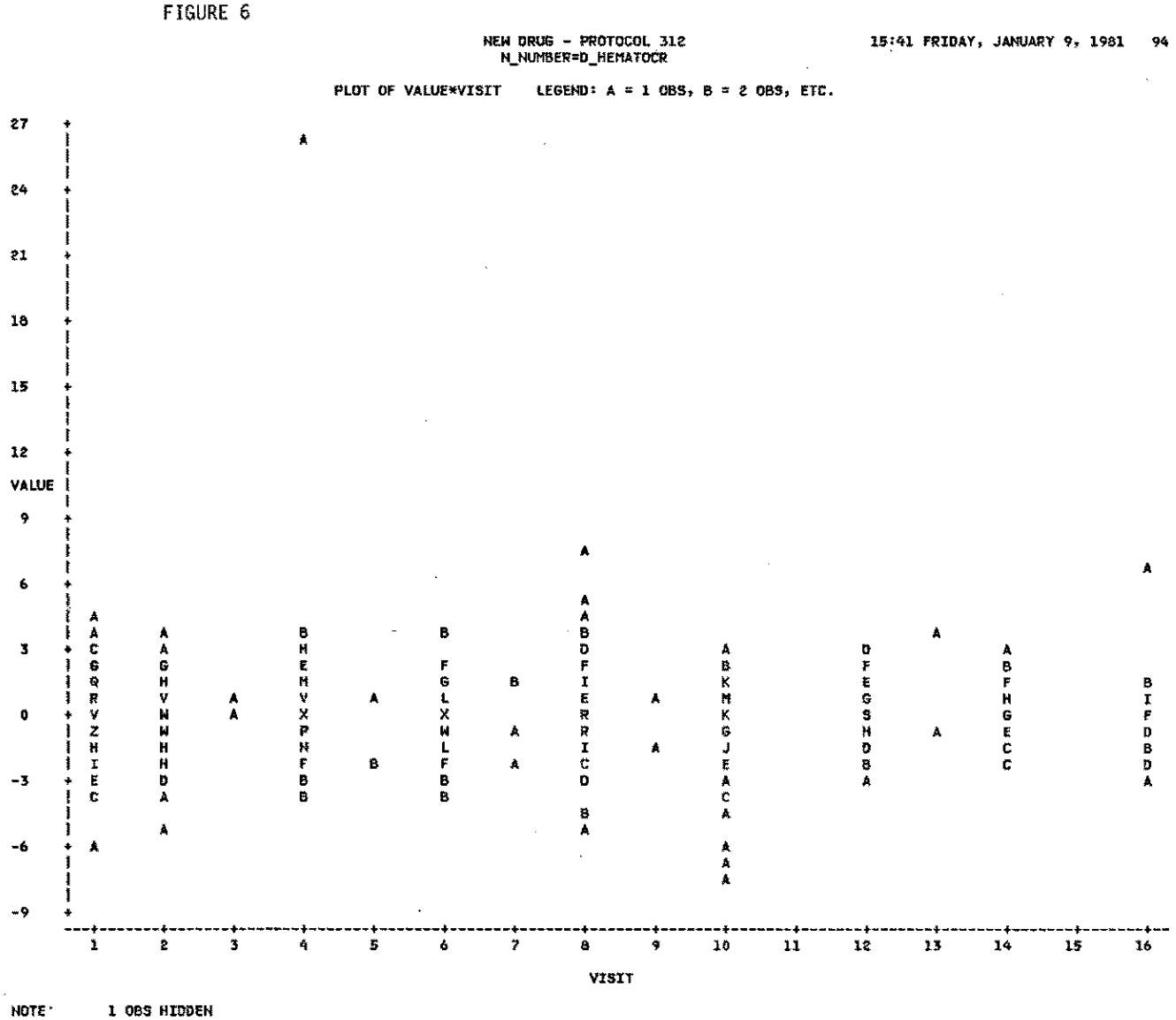
NEW DRUG - PROTOCOL 312
N_NUMBER=D_HEMATOCR
15:41 FRIDAY, JANUARY 9, 1981 94
PLOT OF VALUE*VISIT LEGEND: A = 1 OBS, B = 2 OBS, ETC.
VALUE
VISIT
NOTE: 1 OBS HIDDEN

FIGURE 7

HOE 984

12:38 WEDNESDAY, JANUARY 7, 1981 65

PROT= 320 CASE= 66 PAT_NO= 66 DAYS IN STUDY=36 INV=85 TREATMENT=
STATUS=

FORM	VIS 0 <= 3	VIS 1 4-10	VIS 2 11-17	VIS 3 18-24	VIS 4 25-31	VIS 5 >= 32
DOSAGE	1	8	11	18 21	30	
DOSAGE (END)	8	11	18	21 30	36	
APDI	1					
PREV.PSYCH.MED.	*** -32 -32					
OTHER PREV.MED.	.					
PHYSICAL HISTORY	.					
DEPR. STATUS	1					
BECK INVENTORY	1					
SPMSQ	1					
COVI-RASKIN	1					
BPRS		8		21		36
CGI		8	15	21	30	36
HAMILTON DEP.	1	8		18 21	30	36
HSCL	1	8		18 21	30	36
VITAL	1	8		18 21	30	36
PHYS.	1			21		36
ECG	1			18		36
HEMATOLOGY	1			18		36
URINALYSIS	1			18		36
SERUM CHEMISTRY	1			18		36
CON_MED		5				
SIDE EFFECT (500)					27	
FINAL						36

MISSING RECORDS ARE:
PSYCHMOTOR
PLUTCHICK

FIGURE 8

(HP 9) PROTOCOL 310
INV: STRAUSS M (040)

APPENDIX

INVENTORY OF SUBJECTS IN ANALYSIS

TREATMENT GROUP: PLACEBO

LUMBAR SPINE VARIABLES

PRINCIPALLY AFFECTED JOINT	SUBJ	STATUS	FLEXION							EXTENSION							REASON FOR EXCLUSION
			B	1	2	3	4	E	M	B	1	2	3	4	E	M	
LUMBAR SPINE	344	EFFIC	X	X	-	X	X	X	X	X	X	-	X	X	X	X	
	345	EFFIC	X	X	-	X	X	X	X	X	X	-	X	X	X	X	
	351	EFFIC	X	-	X	X	X	X	X	X	-	X	X	X	X	X	
	352	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	368	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	405	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	416	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	417	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	420	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	435	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	439	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	441	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	444	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	446	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	471	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	475	SAFETY	0	-	0	-	-	-	-	0	-	0	-	-	-	-	2- OFF DRUG FOR 3 DAYS
	478	EFFIC	X	-	X	-	X	X	X	X	-	X	-	X	X	X	
	486	EFFIC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
NO. OF SUBJ			18	18	18	18	18	18	18	18	18	18	18	18	18	18	
INCLUDED			17	15	15	16	17	17	17	17	15	15	16	17	17	17	
EXCLUDED			1	3	3	2	1	1	1	1	3	3	2	1	1	1	

B = BASELINE (DAYS -1,0,1)
1 = WEEK 1 (DAYS 5-11)
2 = WEEK 2 (DAYS 12-18)
3 = WEEK 3 (DAYS 19-25)
4 = WEEK 4 (DAYS 26-36)
E = END POINT
M = MULTIVARIATE
X = ANALYZABLE DATA
0 = UNANALYZABLE DATA
- = MISSING DATA

FIGURE 9

(HP) PROTOCOL 310
INV: BYRNE H (038)

APPENDIX

DOSAGE INFORMATION BY TABLETS

TREATMENT GROUP: PLACEBO

PRINCIPALLY AFFECTED JOINT	SUBJ	WEEK	NO. OF DAYS ON THIS DOSAGE	MEAN DAILY DOSE	MINIMUM DAILY DOSE	MAXIMUM DAILY DOSE	MOST COMMON DAILY DOSE	DOSE DAYS
LUMBAR SPINE	168	1	7	3.6	3.0	4.0	4.0	4
		2	7	5.0	5.0	5.0	5.0	7
		3	7	5.9	5.5	6.0	6.0	6
		4	7	6.0	6.0	6.0	6.0	7
		5	4	6.0	6.0	6.0	6.0	4
	174	1	7	3.1	3.0	4.0	3.0	6
		2	7	5.1	5.0	5.5	5.0	6
		3	7	6.0	6.0	6.0	6.0	7
		4	7	6.0	6.0	6.0	6.0	7
	178	1	7	3.0	3.0	3.0	3.0	7
		2	7	4.9	4.0	5.0	5.0	6
		3	7	5.9	5.5	6.0	6.0	6
		4	7	6.0	6.0	6.0	6.0	7
		5	1	6.0	6.0	6.0	6.0	1
CERVICAL SPINE	101	1	7	3.1	3.0	4.0	3.0	6
		2	7	4.0	4.0	4.0	4.0	7
		3	7	4.0	4.0	4.0	4.0	7
		4	7	5.3	4.0	6.0	6.0	4
		5	7	6.0	6.0	6.0	6.0	7
		6	7	6.0	6.0	6.0	6.0	7
		7	3	6.0	6.0	6.0	6.0	3
	169	1	7	3.0	3.0	3.0	3.0	7
		2	7	3.9	3.5	4.0	4.0	6
		3	7	4.9	4.5	5.0	5.0	6
		4	7	5.9	5.5	6.0	6.0	6
		5	1	6.0	6.0	6.0	6.0	1

ONE TABLET OF = 100 MG

TEST DRUG: NEW DRUG PROTOCOL: XXX
INVESTIGATOR: 276
TREATMENT: 7

FIGURE 10

FINAL DIAGNOSIS: PERITONITIS
STATUS: PATIENTS SUPPORTIVE FOR EFFICACY AND SAFETY CLAIMS

APPENDIX IV
PATIENT SUMMARY

				BEFORE				DURING				AFTER						
INV/PT	AGE	SEX	DYS ON RX	SOURCE	DAY	ORGANISM	S/R	SOURCE	DAY	ORGANISM	S/R	SOURCE	DAY	ORGANISM	S/R	BACT. RESP.	CLIN. RESP.	OVER ALL RESP
276/003	67	M	16	PERIT FLUID	1	FUSOBACT NUCL	S									CURE	SAT	PART
				PERIT FLUID	1	NONE												
												PERFORATED GASTRIC ULCER						
276/010	50	F	6	PERIT FLUID	1	B FRAGILIS	S									CURE	UNSAT	PART
				PERIT FLUID	1	E COLI	S											
				PERIT FLUID	1	E CLOACAE	S											
				PERIT FLUID	1	KLEB. PNEUMON	S											
												APPENDICEAL ABSCESS						
276/013	35	M	8	PERIT FLUID	1	PS MICROS			2	NONE			+ 0	NONE		CURE	SAT	EFF
				PERIT FLUID	1	E COLI	S	SPUTUM	2	E COLI	S	JEJUNOSTOMY	+ 0	E COLI	S			
								SPUTUM	2	H INFLUENZAE		JEJUNOSTOMY	+ 0	ENTEROCOCCUS	S			
												GANGRENOUS SMALL BOWEL						
276/027	56	F	7	BLOOD	1	NONE		WOUND EXUDA	2	NONE						SAT	SAT	EFF
				BLOOD	1	NONE		WOUND EXUDA	2	E COLI	S							
												TUBO-OVARIAN ABSCESS						
276/045	31	M	11	BLOOD	-5	NONE		BLOOD	4	NONE						CURE	SAT	EFF
				BLOOD	-5	NONE		BLOOD	4	NONE								
				WOUND EXUDA	-2	B FRAGILIS	S	WOUND EXUDA	5	B FRAGILIS	S							
				WOUND EXUDA	-2	S EPIDERMIDIS		WOUND EXUDA	5	B FRAGILIS	S							
				BLOOD	1	NONE		URINE	5	NONE								
				WOUND EXUDA	1	NONE		WOUND EXUDA	5	S EPIDERMIDIS								
				BLOOD	1	NONE		URINE	6	NONE								
				WOUND EXUDA	1	E AEROGENES	S	WOUND EXUDA	7	NONE								
				WOUND EXUDA	1	E COLI	S	WOUND EXUDA	7	E COLI	S							
				WOUND EXUDA	1	ENTEROCOCCUS	S	WOUND EXUDA	7	ENTEROCOCCUS	S							
				WOUND EXUDA	1	KLEB. PNEUMON	S	WOUND EXUDA	7	KLEB. PNEUMON	S							
								WOUND EXUDA	7	E AEROGENES	S							
FOLLOW-UP DONE>3 DAYS POST												INTRA-ABDOMINAL ABSCESS						