Spreadsheet from Hell

	A	B	C	D	E	F	G	H	I	J	K
1	Comparison of Drug A and Drug B										
2	Drug A	Age of	Patient	Height	Weight	24hrhct	Blood pressure	Tumor	Race	Date	Complications
3		Patient	Gender	(inches)	(pound)			stage		enrolled	
4											
5	1	25	Male	61”	>350	38%	120/80	2-3	Hispanic	1/15/99	no
6	2	65+	female	5’8”	161	32	140/90	II	White	2/05/1999	yes
7	3	?	Male	120cm		12	>160/110	IV	Black	Jan 98	yes, pneumonia
8	4	31	m	5’6”	obese	40	140 sys 105 dias	?	African-American	?	
9	5	42	f	>6 ft	normal	39	missing	=>2	W	Feb 99	
10	6	45	f	5.7	160	29	80/120	NA	B	last fall	n
11	7	unknown	?	6	145	35	normal	1	W	2/30/99	n
12	8	55	m	72	161.45	12/39	120/95	4	African-American	6-15-00	y
13	9	6 months	f	66	174	38	160/110	3	Asian	14/12/00	y
14	10	21	f	5’							
15											
16	Drug B										
17	1	55	m	61	145	normal	120/80 120/90	IV	Native American	6/20/	3
18	2	45	f	4”11	166	?	135/95	2b	none	7/14/99	n
19	3	32	male	5’13”	171	38	140/80	not staged	NA	8/30/99	n
20	4	44	na	65	?	40	120/80	2	?	09/01/00	n
21	5	66	fem	71	0	41	140/90	4	w	Sep 14th	y, sepsis
22	6	71	unknown	172	199	38	>160/110	3	b	unknown	y, died
23	7	45	m	?	204	32	140 sys 105 dias	1	b	12/25/00	n
24	8	34	m	NA	145	36	130	3	w	July 97	n
25	9	13	m	66	161	39	166/115	2a	w	06/06/99	n
26	10	66	m	68	176	41	1120/80	3	w	01/21/58	n
27											
28	Average	45		65	155	38					

FIGURE 14.3. Spreadsheet from hell. An example of the improper way to enter data for a medical research project.

Ten Data Entry Commandments

1. Enter all, or most, of the data as numbers. Avoid entering letters, words, string variables (e.g., NA, 22%, <3.6), or anything that resembles a cartoon curse word grawlix (*&#%!@?!,). In Excel, all columns, with the exception of names and text comments, should be formatted as numbers or dates (not as general or text).

2. Give each column a unique, simple, one-word name, eight characters or less with no spaces, beginning with a letter, and place this name in the first row.

3. Put only one variable in a column. Do not combine variables in the same column.

4. Enter each patient (or unit of analysis) on a separate line, beginning on the second line/row.

5. Give each research participant or patient a unique case number (1, 2, 3, etc.) in the first column. Delete patient name, SS#, MR#, and any identifying information before sending it to a biostatistician. Always save the spreadsheet with a password.

6. Enter cases and controls in the same spreadsheet. Use one variable to define the control group (TREATED: 0 = no, 1 = yes).

7. Quantify. Enter continuous measurements when possible.

8. Create a simple guide (or key) using a word processor to explain variable abbreviations, value coding, and how missing values were entered. Be consistent.

9. Think through the analysis before collecting any data.

10. Have a biostatistician or other methodologist review the coding before data entry and again after the first 10 patients have been entered.

FIGURE 14.2. Ten data entry commandments. Guidelines for data entry.

Spreadsheet from Heaven

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CASE	GROUP	AGE	SEX	HT	WT	HCT	BPSYS	BPDIAS	STAGE	RACE	DATE1	COMPLIC
2	1	1	25	1	61	350	38	120	80	3	3	1/15/1999	0
3	2	1	65	2	68	161	32	140	90	2	1	2/5/1999	1
4	3	1	25	1	47	150	38	160	110	4	2	1/15/1998	1
5	4	1	31	1	66	161	40	140	105	2	2	4/1/1999	0
6	5	1	42	2	72	177	39	130	70	2	1	2/15/1999	0
7	6	1	45	2	67	160	29	120	80	1	2	3/6/1999	0
8	7	1	44	1	72	145	35	120	80	1	1	2/28/1999	0
9	8	1	55	1	72	161	39	120	95	4	2	6/15/2000	1
10	9	1	0.5	2	66	174	38	160	110	3	4	12/14/2000	1
11	10	1	21	2	60	155	40	190	120	2	2	11/14/2000	0
12	11	2	55	1	61	145	41	120	80	4	5	6/20/1999	1
13	12	2	45	2	59	166	39	135	95	2	1	7/14/1999	0
14	13	2	32	1	73	171	38	140	80	1	1	8/30/1999	0
15	14	2	44	2	65	155	40	120	80	2	2	9/1/2000	0
16	15	2	66	2	71	145	41	140	90	4	1	9/14/1999	1
17	16	2	71	1	68	199	38	160	110	3	2	1/14/1999	1
18	17	2	45	1	69	204	32	140	105	1	2	12/25/2000	0
19	18	2	34	1	66	145	36	130	75	3	1	7/15/1997	0
20	19	2	13	1	66	161	39	166	115	2	1	6/6/1999	0
21	20	2	66	1	68	176	41	120	80	3	1	1/21/1998	0

FIGURE 14.4. Spreadsheet from heaven. An example of the proper way to enter data for a medical research project.