Example Closed Meeting Data Monitoring Committee Report

Frank Harrell

August 1, 2007

Contents

List of Figures

List of Tables

1 Introduction

This is a sample of the part of a closed meeting Data Monitoring Committee report that contains software generated results. Components related to efficacy, study design, data monitoring plan[1], summary of previous closed report, interpretation, protocol changes, screening, eligibility, and waiting time until treatment commencement are not included in this example[2]. This report used a random sample of safety data from a randomized clinical trial. Randomization date, dropouts, and compliance variables were simulated, the latter two not being made consistent with the presence or absence of actual data in the random sample. The date and time that the analysis file used here was last updated was 2002-09-26 22:43:50. Source analysis files were last updated on 2002-07-08 14:59:22. See Section 9 for information about software used.

LaTeX's `hyperref` style was used to produce a `pdf` file with hyperlinks for easy navigation to sections, tables, and graphs using Adobe Acrobat Reader. Internal hyperlinks are shown in blue, and external links to web sites are shown in red.

See the example open meeting report for subject accrual, data availability and completeness, and analyses not stratified by treatment.

2 Baseline Variables

[1] Lan-DeMets monitoring bounds can be plotted using the open source S `ldBands` function in the Hmisc library for Linux and Unix systems.

[2] See Ellenberg, Fleming, and DeMets, *Data Monitoring Committees in Clinical Trials* (Wiley, 2002), pp. 73-74 for recommended components in open and closed data monitoring committee reports.

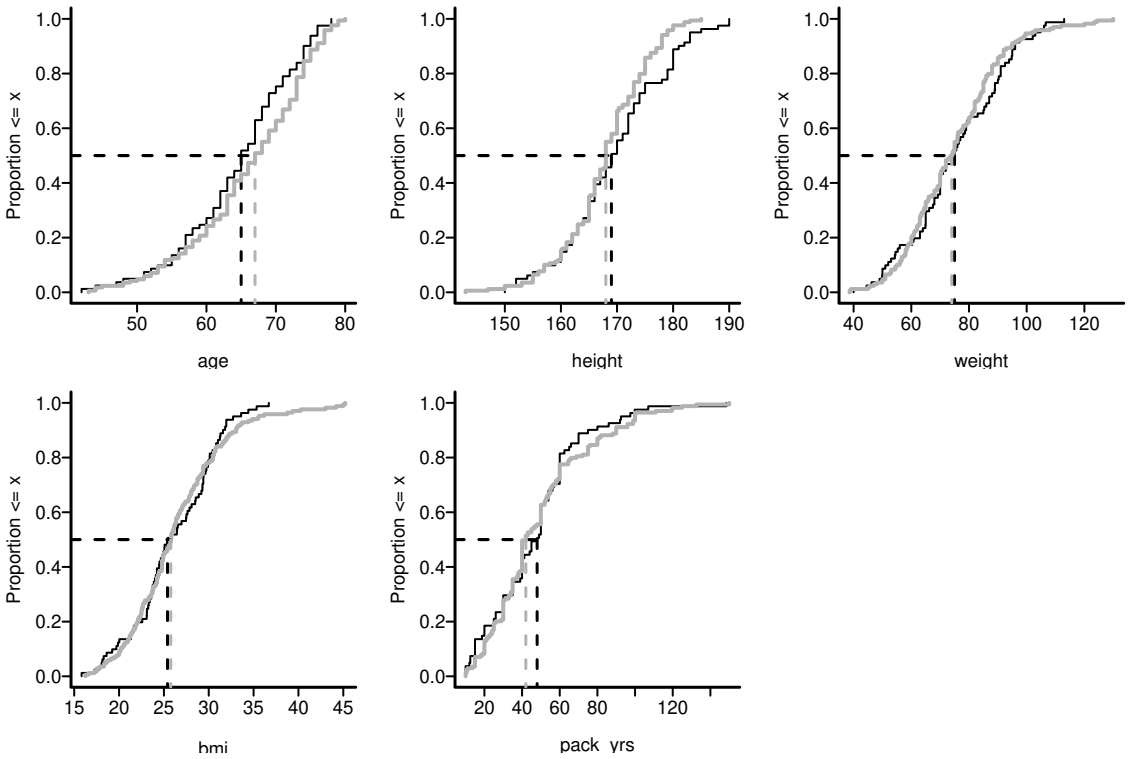

Figure 1: Empirical cumulative distribution plots of the continuous variables in the 'baseline' table . Reference lines are drawn at treatment-specific median values. A:——; B:——.

Table 1: Baseline

	N	A $N = 81$	B $N = 169$	P-value
age	250	60 **65** 70	61 **67** 73	0.104[1]
Sex	250			0.453[2]
male		78% (63)	73% (124)	
Race	250			0.189[2]
Black		0% (0)	3% (5)	
Caucasian		98% (79)	96% (163)	
Oriental		1% (1)	0% (0)	
Other		1% (1)	1% (1)	
height	250	164 **169** 175	164 **168** 173	0.185[1]
weight	250	65 **75** 89	63 **74** 85	0.516[1]
bmi	250	23.3 **25.4** 29.6	22.6 **25.8** 29.4	0.996[1]
smoking	250	37% (30)	35% (59)	0.743[2]
pack_yrs	250	30 **48** 60	30 **42** 60	0.862[1]

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Numbers after percents are frequencies. Tests used: [1]Wilcoxon test; [2]Pearson test

3 Interrelationships Among Variables

Variable clustering diagrams are shown in the figures that follow. Variables are grouped according to how they are correlated with one another, as measured by the square of the Spearman ρ rank correlation coefficient computed on all pairs of variables. Variables connected on lower branches are more highly correlated with one another. Variables missing in more than 0.75 of the observations or categorical variables having more than 20 levels are ignored. Categories less than 0.1 prevalent are pooled with other rare categories.

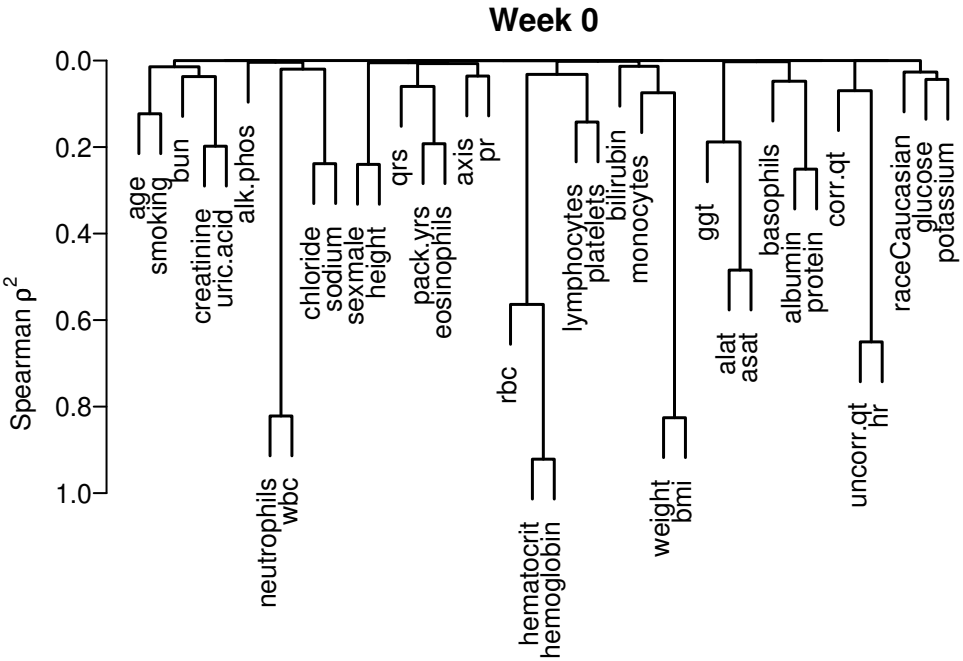

Figure 2: Clustering of variables at Week 0

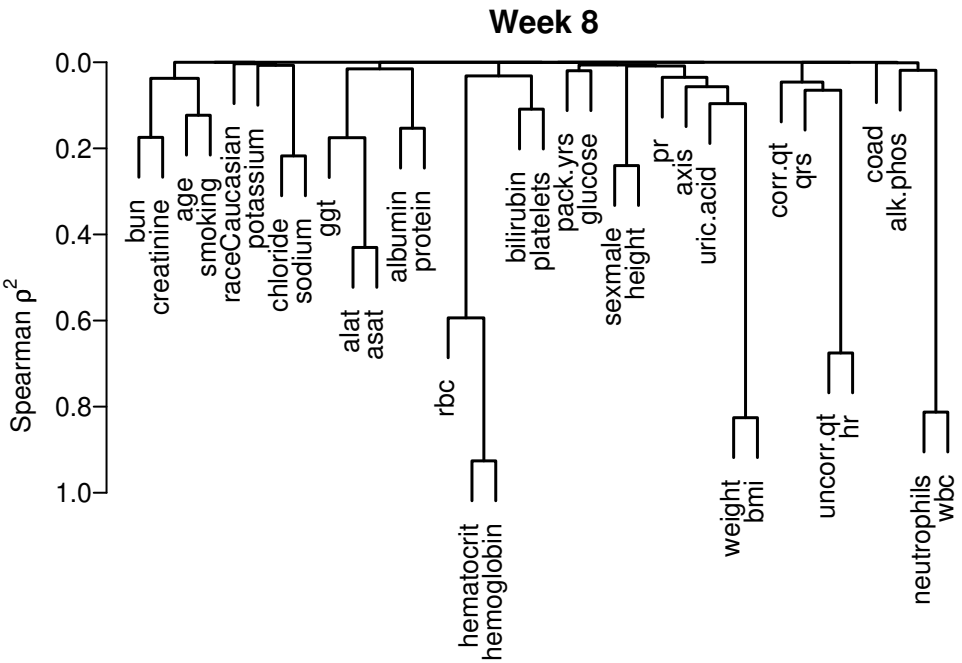

Figure 3: Clustering of variables at Week 8

4 Compliance to Assigned Treatments

Table 2: Compliance by Treatment

	A		**B**	
	N	Compliance	N	Compliance
Week				
2	81	1.00	169	0.99
4	81	1.00	169	0.98
8	81	0.99	169	0.98
12	81	0.93	169	0.96
16	81	0.84	169	0.80
20	81	0.78	169	0.70
Overall				
	486	0.92	1014	0.90

5 Dropouts

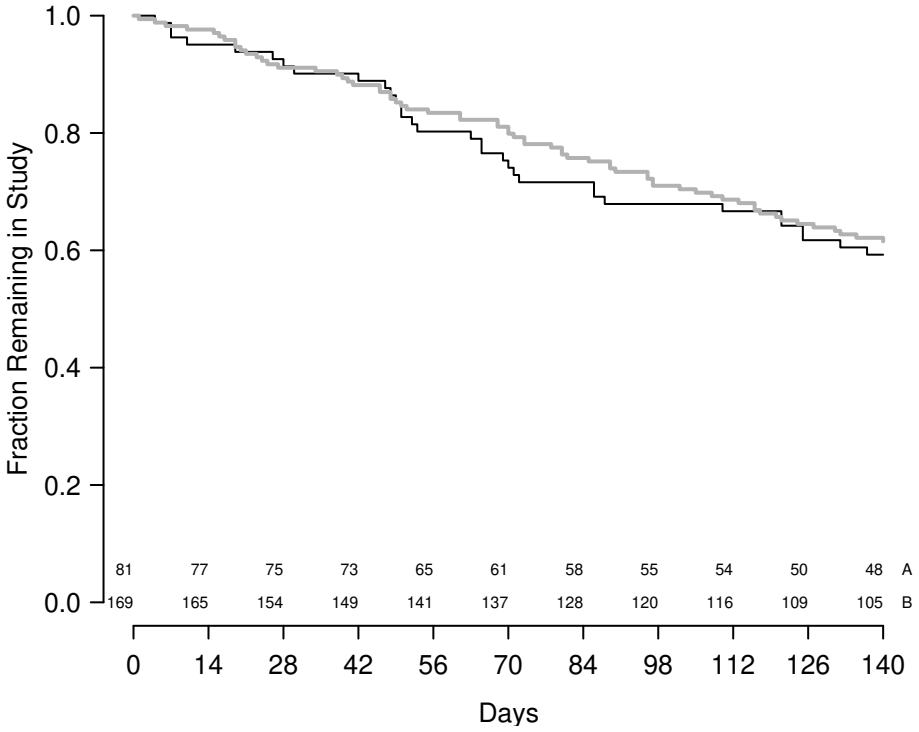

Figure 4: Distribution of time until dropout from study. A:——; B:——.

6 Adverse Events

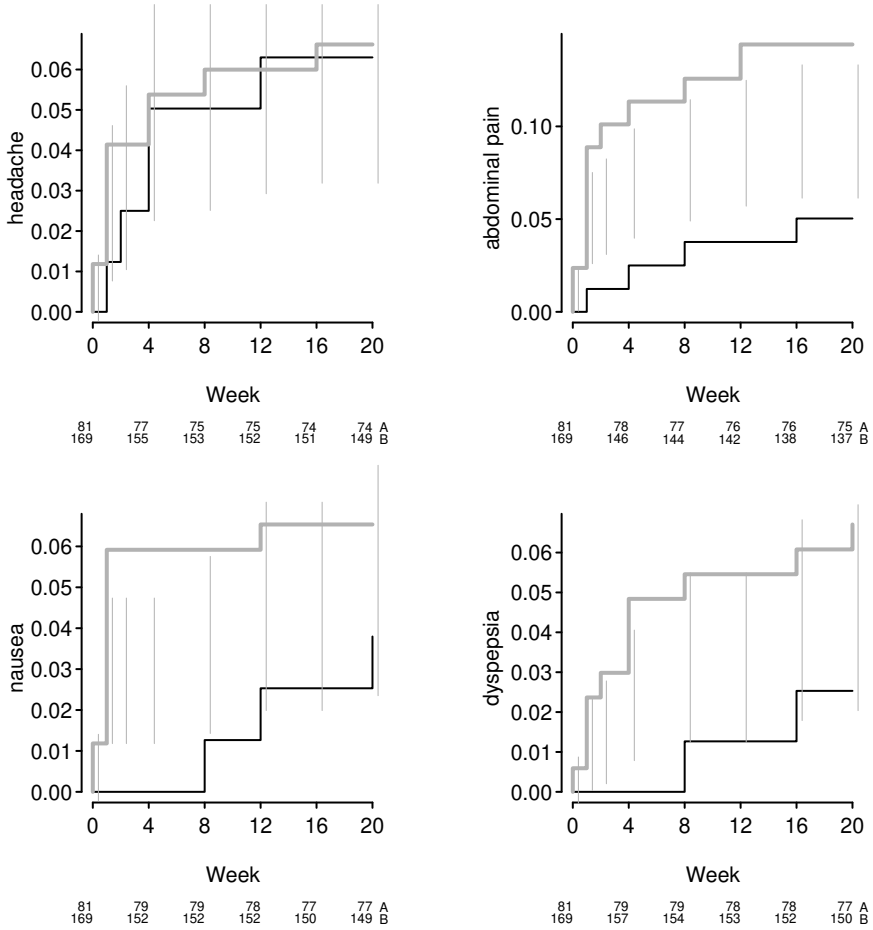

Figure 5: Kaplan-Meier estimates of cumulative probabilities of adverse events by treatment over time. Dotted vertical bars indicate half-widths of approximate 0.95 confidence intervals for differences in probabilities. When the distance between two proportions exceeds the length of the bar, differences are significant at approximately the 0.05 level. A:——; B:——.

In the following tables N is the number of subjects and numbers after percents are frequencies. P-values are from Pearson χ^2 tests.

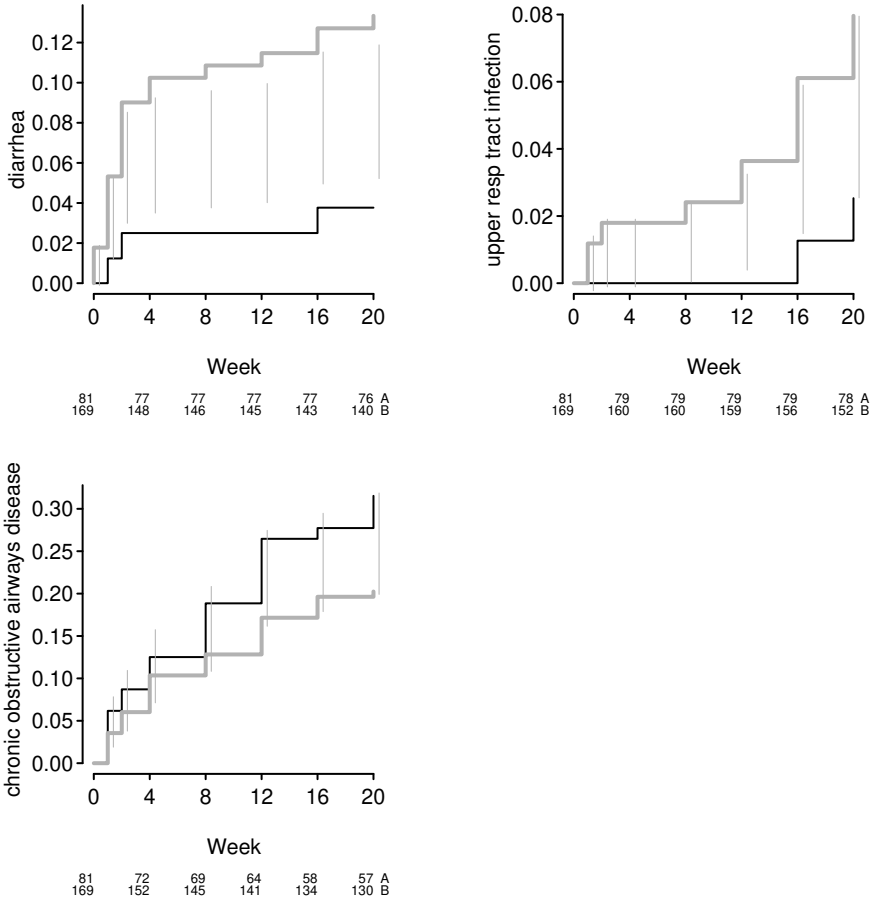

Figure 6: Kaplan-Meier estimates of cumulative probabilities of adverse events by treatment over time (continued)

Table 3: Adverse Events at Week 4

	N	A	B	P-value
		$N = 81$	$N = 169$	
headache	246	2% (2)	2% (4)	0.983
abdominal pain	246	1% (1)	2% (4)	0.534
nausea	246	0% (0)	1% (2)	0.32
dyspepsia	246	0% (0)	2% (3)	0.222
diarrhea	246	0% (0)	2% (3)	0.222
upper resp tract infection	246	0% (0)	0% (0)	
chronic obstructive airways disease	246	6% (5)	6% (10)	0.972

Table 4: Adverse Events at Week 12

	N	A	B	P-value
		$N = 81$	$N = 169$	
headache	241	1% (1)	0% (0)	0.151
abdominal pain	241	1% (1)	2% (3)	0.738
nausea	241	1% (1)	1% (1)	0.602
dyspepsia	241	0% (0)	1% (1)	0.484
diarrhea	241	0% (0)	1% (1)	0.484
upper resp tract infection	241	0% (0)	1% (2)	0.321
chronic obstructive airways disease	241	9% (7)	6% (10)	0.444

Table 5: Adverse Events at Any Time

	A	B	P-value
	$N = 81$	$N = 169$	
headache	6% (5)	7% (11)	0.919
abdominal pain	5% (4)	15% (25)	0.023
nausea	4% (3)	7% (12)	0.29
dyspepsia	2% (2)	7% (11)	0.178
diarrhea	4% (3)	14% (23)	0.016
upper resp tract infection	2% (2)	8% (13)	0.104
chronic obstructive airways disease	32% (26)	20% (34)	0.038

7 EKG Data

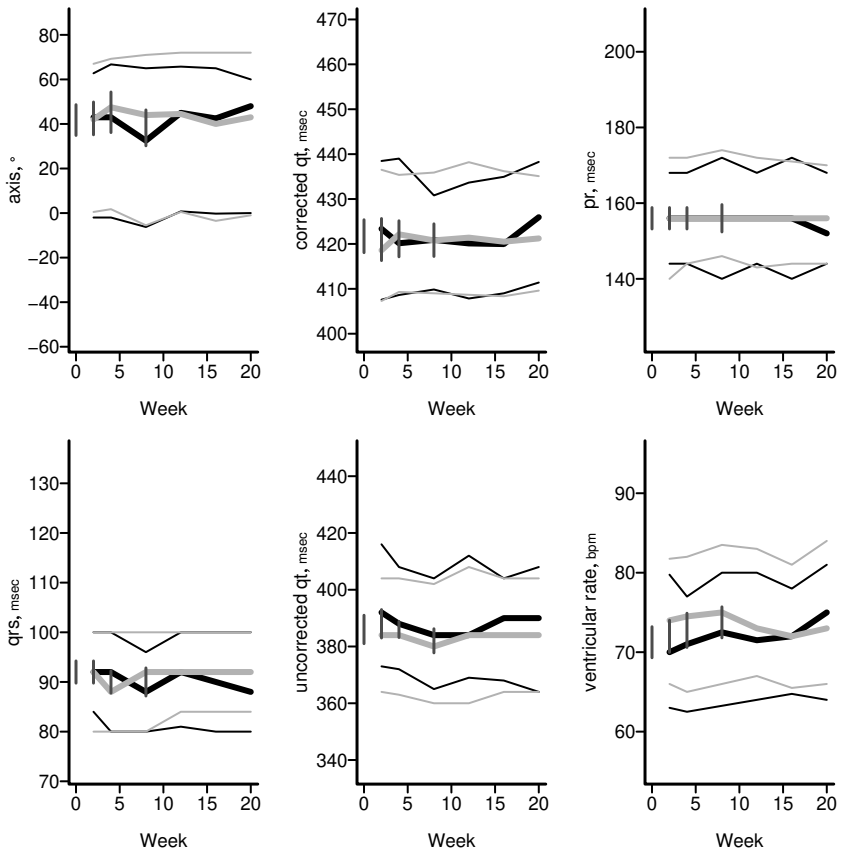

Figure 7: Quartiles of EKG variables over time. Outer lines are 25^{th} (lower line) and 75^{th} (upper line) percentiles. Thicker middle lines depict medians. y-axis is scaled to the pooled 5^{th} and 95^{th} quantiles. Vertical bars indicate half-widths of approximate 0.95 confidence intervals for differences in medians. When the distance between two medians exceeds the length of the bar, differences are significant at approximately the 0.05 level. A:——; B:——.

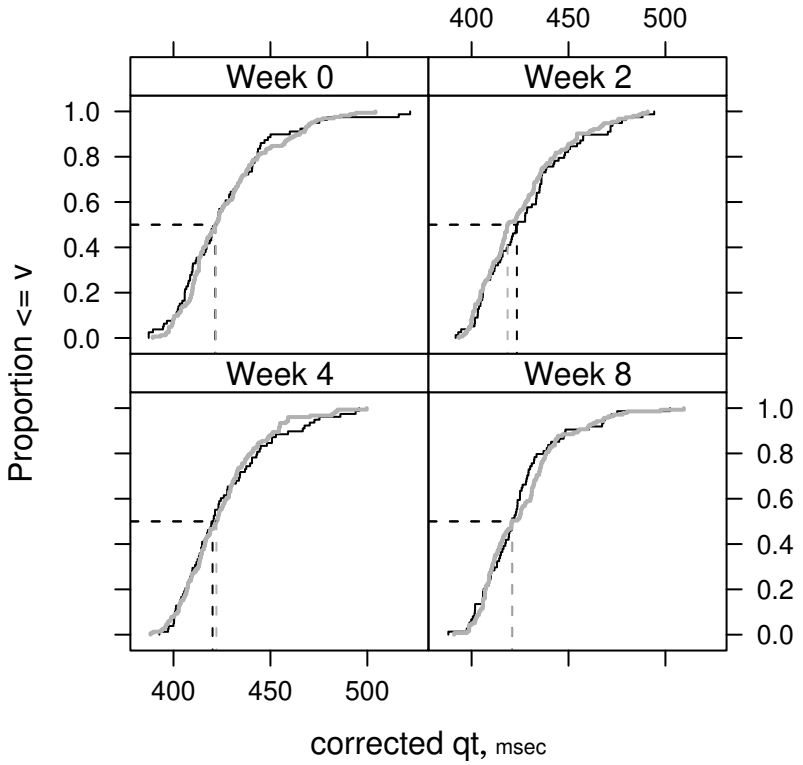

Figure 8: Empirical cumulative distribution function of corrected qt by treatment over time. A:——; B:——.

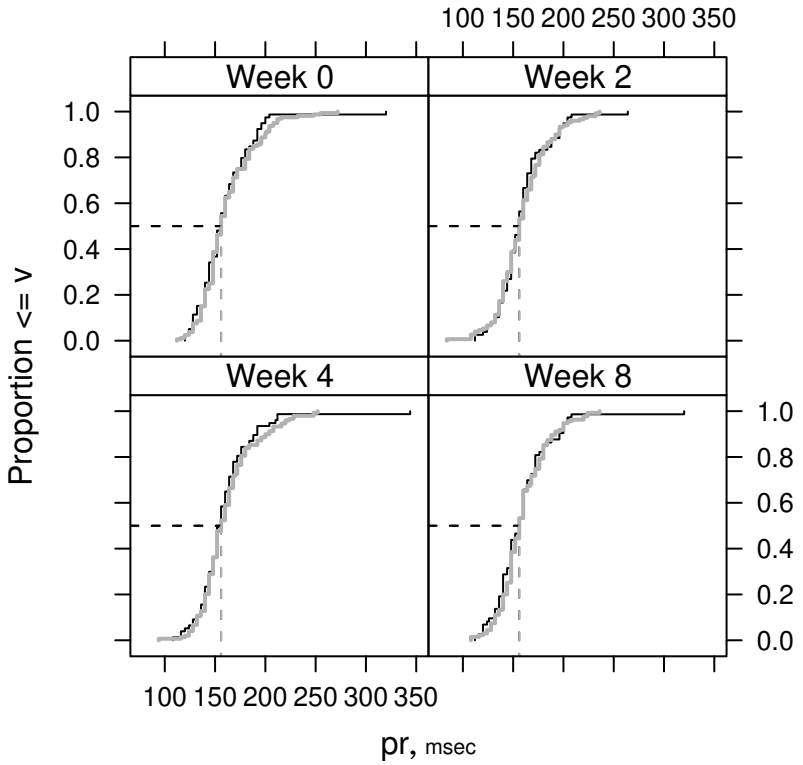

Figure 9: Empirical cumulative distribution function of pr by treatment over time. A:——; B:——.

Table 6: EKG Data at Week 0

		N	A $N = 81$	B $N = 169$	P-value
axis	degree	243	-4.00 **41.00** 62.50	1.75 **42.50** 69.00	0.347
corrected qt	msec	243	408 **422** 441	411 **422** 439	0.607
pr	msec	239	142 **156** 174	147 **156** 174	0.521
qrs	msec	243	82 **92** 100	80 **92** 100	0.871
uncorrected qt	msec	243	370 **384** 412	368 **388** 412	0.661
ventricular rate	bpm	243	63.0 **70.0** 78.0	64.0 **72.5** 82.0	0.23

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

Table 7: EKG Data at Week 2

		N	A $N = 81$	B $N = 169$	P-value
axis	degree	232	-2.0 **43.0** 62.8	0.5 **42.0** 67.0	0.886
corrected qt	msec	232	408 **423** 438	407 **419** 437	0.375
pr	msec	228	144 **156** 168	140 **156** 172	0.772
qrs	msec	232	84 **92** 100	80 **92** 100	0.564
uncorrected qt	msec	232	373 **392** 416	364 **384** 404	0.06
ventricular rate	bpm	232	63.0 **70.0** 79.8	66.0 **74.0** 81.8	0.098

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

Table 8: EKG Data at Week 4

		N	A $N = 81$	B $N = 169$	P-value
axis	degree	230	-2.00 **43.00** 66.75	1.75 **47.50** 69.25	0.771
corrected qt	msec	230	409 **420** 439	409 **422** 435	0.974
pr	msec	226	144 **156** 168	144 **156** 172	0.459
qrs	msec	230	80 **92** 100	80 **88** 100	0.778
uncorrected qt	msec	230	372 **388** 408	363 **384** 404	0.099
ventricular rate	bpm	230	62.5 **71.0** 77.0	65.0 **74.5** 82.0	0.061

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

Table 9: EKG Data at Week 8

		N	A $N = 81$	B $N = 169$	P-value
axis	degree	213	-6.25 **32.50** 65.00	-5.50 **44.00** 71.00	0.588
corrected qt	msec	213	410 **421** 431	409 **421** 436	0.587
pr	msec	208	140 **156** 172	146 **156** 174	0.515
qrs	msec	213	80 **88** 96	80 **92** 100	0.315
uncorrected qt	msec	213	365 **384** 404	360 **380** 402	0.401
ventricular rate	bpm	213	63.2 **72.5** 80.0	66.0 **75.0** 83.5	0.124

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

8 Clinical Chemistry Data

Table 10: Clinical Chemistry Data at Week 0

		N	A $N = 81$	B $N = 169$	P-value
neutrophils absolute	10^9/L	244	3.80 **4.42** 5.22	3.54 **4.62** 5.70	0.671
alanine aminotransferase	IU/L	246	11.0 **15.0** 21.5	12.0 **16.0** 23.0	0.558
albumin	G/L	247	41 **42** 43	41 **42** 44	0.268
alkaline phosphatase	IU/L	246	60.0 **76.0** 89.5	63.5 **77.0** 90.5	0.476
aspartate aminotransferase	IU/L	246	15.0 **18.0** 22.5	15.0 **19.0** 22.0	0.723
basophils	10^9/L	70	0.0157 **0.0295** 0.0513	0.0141 **0.0270** 0.0475	0.852
total bilirubin	UMOL/L	246	8 **10** 12	7 **10** 13	0.961
blood urea nitrogen	MMOL/L	247	4.55 **5.36** 6.43	4.57 **5.60** 6.70	0.552
chloride	MMOL/L	247	102 **104** 106	102 **104** 106	0.983
creatinine	UMOL/L	246	61.9 **74.0** 88.4	61.9 **73.0** 88.0	0.998
eosinophils	10^9/L	70	0.168 **0.211** 0.268	0.131 **0.211** 0.346	0.995
γ glutamyl transferase	IU/L	246	20.0 **35.0** 53.5	20.0 **26.0** 41.0	0.075
glucose - random	MMOL/L	234	5.10 **5.60** 6.40	4.90 **5.60** 6.55	0.732
hematocrit	%	245	42.2 **44.5** 47.5	41.1 **43.7** 46.3	0.07
hemoglobin	G/L	245	142 **150** 158	137 **146** 155	0.031
potassium	MMOL/L	246	4.2 **4.5** 4.7	4.3 **4.5** 4.7	0.681
lymphocytes	10^9/L	70	1.44 **1.97** 2.63	1.46 **1.80** 2.29	0.817
monocytes	10^9/L	70	0.324 **0.464** 0.642	0.323 **0.446** 0.587	0.545
sodium	MMOL/L	247	139 **140** 142	139 **141** 142	0.754
platelets	10^9/L	245	185 **214** 264	184 **225** 265	0.957
total protein	G/L	247	68.5 **71.0** 73.0	68.0 **71.0** 74.0	0.954
red blood cell count	10^{12}/L	245	4.4 **4.6** 5.0	4.4 **4.7** 5.0	0.92
uric acid	UMOL/L	247	264 **321** 383	260 **312** 375	0.792
white blood cell count	10^9/L	245	6.05 **7.15** 8.10	6.10 **7.40** 8.50	0.482

a **b** *c* represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

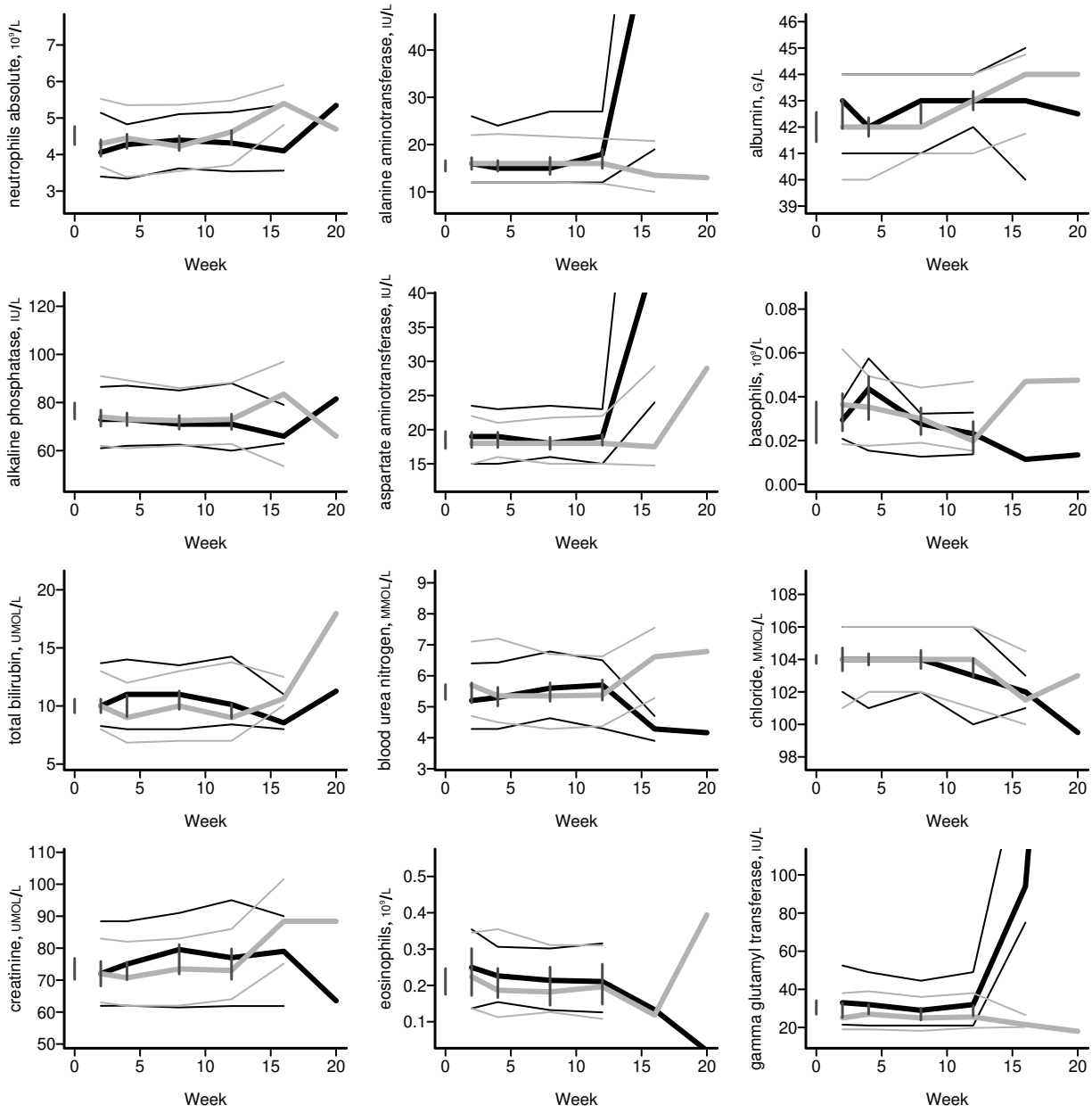

Figure 10: Quartiles of Clinical Chemistry variables over time. Outer lines are 25^{th} (lower line) and 75^{th} (upper line) percentiles. Thicker middle lines depict medians. y-axis is scaled to the pooled 5^{th} and 95^{th} quantiles. Vertical bars indicate half-widths of approximate 0.95 confidence intervals for differences in medians. When the distance between two medians exceeds the length of the bar, differences are significant at approximately the 0.05 level. A:——; B:——.

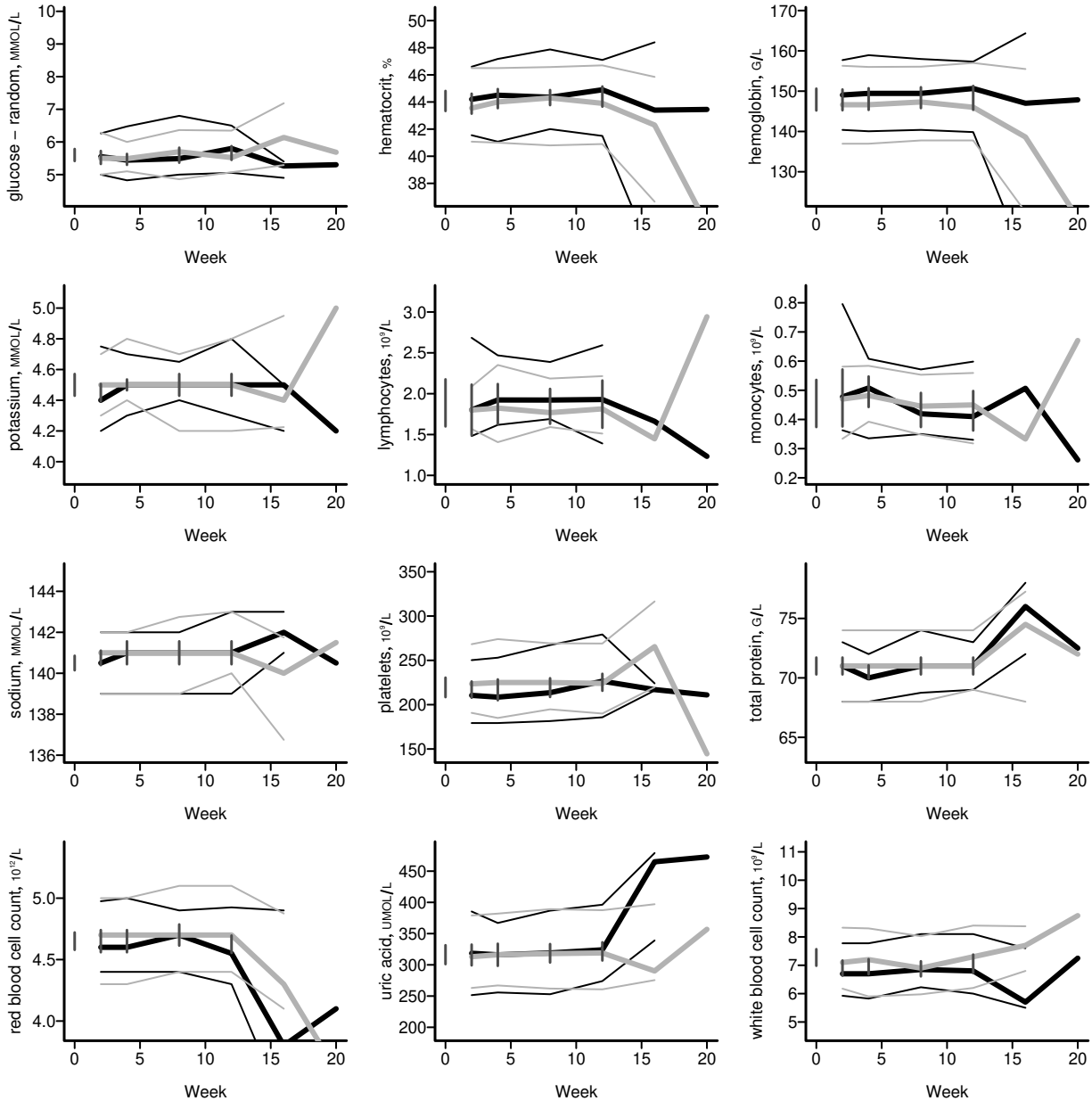

Figure 11: Quartiles of Clinical Chemistry variables over time (continued)

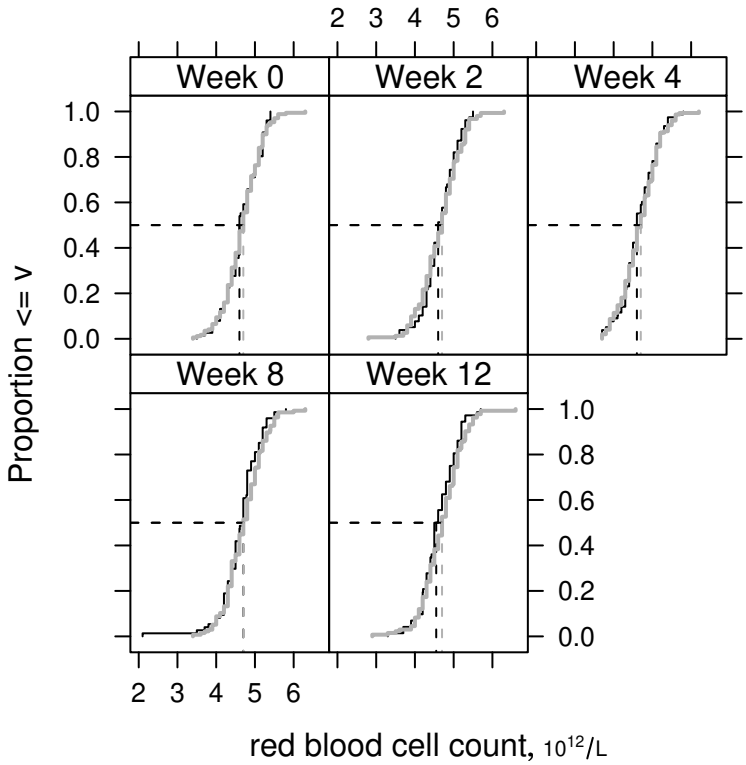

Figure 12: Empirical cumulative distribution function of red blood cell count by treatment over time. A:——; B:——.

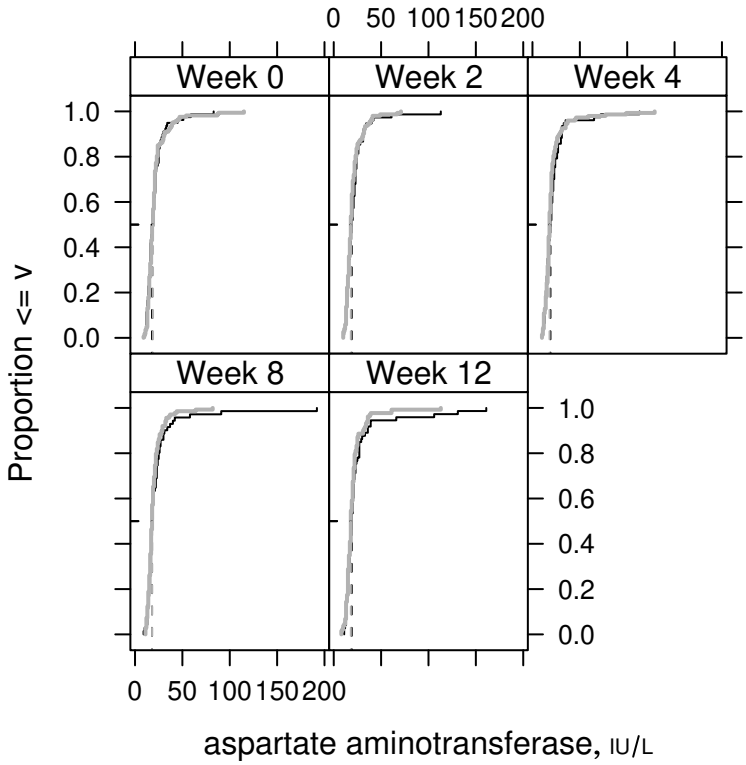

Figure 13: Empirical cumulative distribution function of aspartate aminotransferase by treatment over time. A:——; B:——.

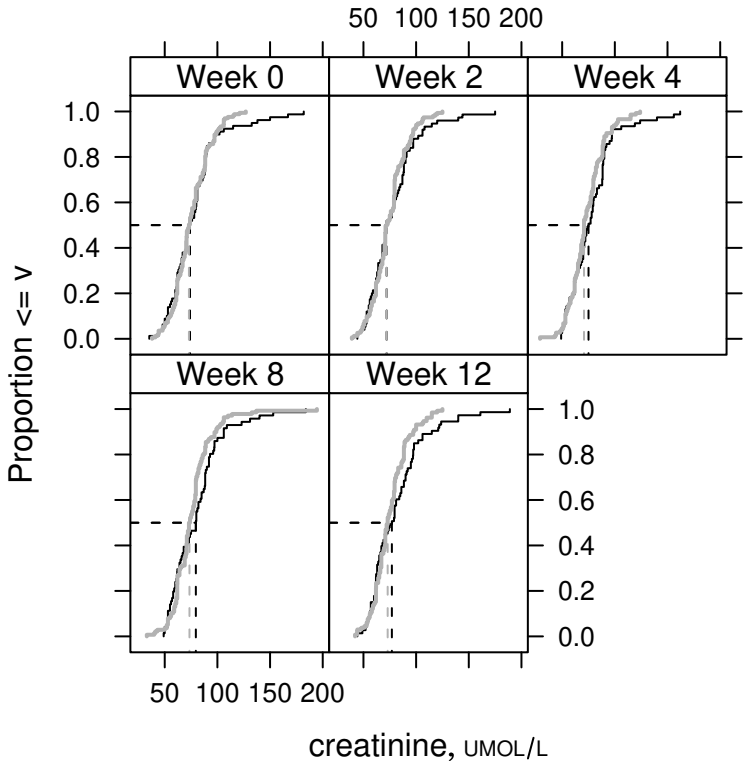

Figure 14: Empirical cumulative distribution function of creatinine by treatment over time. A:——; B:——.

Table 11: Clinical Chemistry Data at Week 2

		N	A	B	P-value
			$N = 81$	$N = 169$	
neutrophils absolute	10^9/L	229	3.40 **4.06** 5.14	3.67 **4.30** 5.53	0.195
alanine aminotransferase	IU/L	228	12 **16** 26	12 **16** 22	0.674
albumin	G/L	228	41 **43** 44	40 **42** 44	0.208
alkaline phosphatase	IU/L	228	61.0 **73.0** 86.5	62.0 **74.0** 91.0	0.716
aspartate aminotransferase	IU/L	228	15.0 **19.0** 23.5	15.0 **18.0** 22.0	0.42
basophils	10^9/L	64	0.0208 **0.0294** 0.0384	0.0183 **0.0364** 0.0616	0.224
total bilirubin	UMOL/L	228	8.28 **10.00** 13.68	8.00 **10.00** 13.00	0.334
blood urea nitrogen	MMOL/L	229	4.28 **5.20** 6.40	4.70 **5.70** 7.10	0.093
chloride	MMOL/L	230	102 **104** 106	101 **104** 106	0.76
creatinine	UMOL/L	228	61.9 **72.0** 88.4	63.0 **72.0** 83.0	0.718
eosinophils	10^9/L	64	0.137 **0.250** 0.354	0.137 **0.224** 0.345	0.961
γ glutamyl transferase	IU/L	228	21.5 **33.0** 52.5	19.0 **25.0** 38.0	0.048
glucose - random	MMOL/L	215	5.00 **5.55** 6.26	5.00 **5.50** 6.29	0.839
hematocrit	%	230	41.6 **44.2** 46.6	41.1 **43.5** 46.5	0.367
hemoglobin	G/L	230	140 **149** 158	137 **147** 156	0.148
potassium	MMOL/L	229	4.20 **4.40** 4.75	4.30 **4.50** 4.70	0.909
lymphocytes	10^9/L	64	1.48 **1.80** 2.68	1.57 **1.80** 2.09	0.586
monocytes	10^9/L	64	0.363 **0.477** 0.796	0.334 **0.470** 0.581	0.214
sodium	MMOL/L	230	139 **140** 142	139 **141** 142	0.904
platelets	10^9/L	230	179 **210** 250	191 **224** 268	0.101
total protein	G/L	229	68 **71** 73	68 **71** 74	0.985
red blood cell count	10^{12}/L	230	4.40 **4.60** 4.98	4.30 **4.70** 5.00	0.902
uric acid	UMOL/L	229	252 **318** 385	263 **313** 379	0.916
white blood cell count	10^9/L	230	5.93 **6.70** 7.78	6.17 **7.10** 8.33	0.13

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

Table 12: Clinical Chemistry Data at Week 4

		N	A	B	P-value
			$N = 81$	$N = 169$	
neutrophils absolute	10^9/L	226	3.34 **4.28** 4.83	3.39 **4.45** 5.35	0.3
alanine aminotransferase	IU/L	225	12.0 **15.0** 24.0	12.0 **16.0** 22.2	0.918
albumin	G/L	224	41 **42** 44	40 **42** 44	0.716
alkaline phosphatase	IU/L	225	62.0 **73.0** 87.0	61.0 **73.0** 89.2	0.861
aspartate aminotransferase	IU/L	225	15 **19** 23	16 **18** 21	0.403
basophils	10^9/L	62	0.0154 **0.0436** 0.0575	0.0176 **0.0352** 0.0494	0.517
total bilirubin	UMOL/L	225	8.00 **11.00** 14.00	6.84 **9.00** 12.00	0.081
blood urea nitrogen	MMOL/L	225	4.28 **5.30** 6.43	4.50 **5.36** 7.20	0.41
chloride	MMOL/L	226	101 **104** 106	102 **104** 106	0.513
creatinine	UMOL/L	224	62.0 **75.0** 88.4	61.9 **70.7** 82.0	0.191
eosinophils	10^9/L	62	0.154 **0.226** 0.306	0.112 **0.187** 0.355	0.582
γ glutamyl transferase	IU/L	224	21 **32** 49	19 **27** 39	0.162
glucose - random	MMOL/L	213	4.83 **5.44** 6.48	5.10 **5.49** 6.00	0.994
hematocrit	%	227	41.1 **44.5** 47.2	41.0 **44.0** 46.5	0.326
hemoglobin	G/L	227	140 **149** 159	137 **147** 156	0.149
potassium	MMOL/L	226	4.3 **4.5** 4.7	4.4 **4.5** 4.8	0.786
lymphocytes	10^9/L	62	1.62 **1.92** 2.47	1.41 **1.82** 2.35	0.363
monocytes	10^9/L	62	0.335 **0.508** 0.608	0.392 **0.482** 0.584	0.766
sodium	MMOL/L	226	139 **141** 142	139 **141** 142	0.822
platelets	10^9/L	227	179 **208** 253	185 **225** 274	0.264
total protein	G/L	225	68 **70** 72	68 **71** 74	0.14
red blood cell count	10^{12}/L	227	4.4 **4.6** 5.0	4.3 **4.7** 5.0	0.692
uric acid	UMOL/L	225	256 **316** 367	267 **316** 382	0.665
white blood cell count	10^9/L	227	5.83 **6.70** 7.78	5.90 **7.20** 8.30	0.17

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

Table 13: Clinical Chemistry Data at Week 8

		N	A	B	P-value
			$N = 81$	$N = 169$	
neutrophils absolute	10^9/L	210	3.62 **4.39** 5.11	3.55 **4.22** 5.36	0.996
alanine aminotransferase	IU/L	209	12.0 **15.0** 27.0	12.0 **16.0** 21.8	0.38
albumin	G/L	210	41 **43** 44	41 **42** 44	0.102
alkaline phosphatase	IU/L	209	62.5 **71.0** 85.0	62.0 **72.5** 86.0	0.556
aspartate aminotransferase	IU/L	209	16.0 **18.0** 23.5	15.0 **18.0** 21.8	0.24
basophils	10^9/L	59	0.0126 **0.0276** 0.0323	0.0190 **0.0300** 0.0442	0.163
total bilirubin	UMOL/L	209	8.0 **11.0** 13.5	7.0 **10.0** 13.0	0.186
blood urea nitrogen	MMOL/L	210	4.63 **5.60** 6.78	4.28 **5.35** 6.70	0.543
chloride	MMOL/L	210	102 **104** 106	102 **104** 106	0.881
creatinine	UMOL/L	209	61.4 **79.6** 91.0	62.0 **73.5** 83.0	0.321
eosinophils	10^9/L	59	0.132 **0.214** 0.302	0.127 **0.182** 0.311	0.856
γ glutamyl transferase	IU/L	209	21.0 **29.0** 44.5	18.2 **25.0** 36.0	0.089
glucose - random	MMOL/L	200	5.00 **5.49** 6.80	4.86 **5.70** 6.36	0.98
hematocrit	%	210	42.0 **44.4** 47.9	40.8 **44.3** 46.6	0.431
hemoglobin	G/L	210	140 **149** 158	138 **147** 156	0.251
potassium	MMOL/L	209	4.40 **4.50** 4.65	4.20 **4.50** 4.70	0.691
lymphocytes	10^9/L	59	1.69 **1.92** 2.39	1.59 **1.77** 2.19	0.392
monocytes	10^9/L	59	0.350 **0.419** 0.572	0.348 **0.445** 0.553	0.887
sodium	MMOL/L	210	139 **141** 142	139 **141** 143	0.48
platelets	10^9/L	210	182 **214** 267	195 **225** 269	0.305
total protein	G/L	210	68.8 **71.0** 74.0	68.0 **71.0** 74.0	0.754
red blood cell count	10^{12}/L	210	4.4 **4.7** 4.9	4.4 **4.7** 5.1	0.252
uric acid	UMOL/L	210	253 **319** 387	262 **318** 389	0.724
white blood cell count	10^9/L	210	6.22 **6.85** 8.10	5.98 **6.90** 8.03	0.647

a *b* *c* represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. N is the number of non–missing values. Test used: Wilcoxon test

Table 14: Clinical Chemistry Data at Week 12

		N	A	B	P-value
			$N = 81$	$N = 169$	
neutrophils absolute	10^9/L	205	3.54 **4.32** 5.16	3.71 **4.62** 5.48	0.347
alanine aminotransferase	IU/L	205	12.0 **18.0** 27.0	11.8 **16.0** 21.2	0.148
albumin	G/L	205	42 **43** 44	41 **43** 44	0.662
alkaline phosphatase	IU/L	205	60.0 **71.0** 88.0	62.8 **73.0** 88.2	0.759
aspartate aminotransferase	IU/L	205	15 **19** 23	15 **18** 22	0.294
basophils	10^9/L	56	0.0137 **0.0231** 0.0327	0.0152 **0.0198** 0.0469	0.787
total bilirubin	UMOL/L	204	8.41 **10.13** 14.25	7.00 **9.00** 13.76	0.208
blood urea nitrogen	MMOL/L	205	4.30 **5.70** 6.50	4.38 **5.38** 6.62	0.772
chloride	MMOL/L	206	100 **103** 106	101 **104** 106	0.325
creatinine	UMOL/L	205	61.9 **77.0** 95.0	64.0 **73.0** 86.0	0.38
eosinophils	10^9/L	56	0.126 **0.211** 0.316	0.108 **0.196** 0.310	0.866
γ glutamyl transferase	IU/L	205	21.0 **32.0** 49.0	19.8 **25.5** 38.0	0.043
glucose - random	MMOL/L	194	5.05 **5.80** 6.50	5.07 **5.52** 6.35	0.623
hematocrit	%	205	41.5 **44.9** 47.1	40.9 **43.9** 46.7	0.529
hemoglobin	G/L	205	140 **151** 157	138 **146** 157	0.265
potassium	MMOL/L	205	4.3 **4.5** 4.8	4.2 **4.5** 4.8	0.539
lymphocytes	10^9/L	56	1.39 **1.93** 2.59	1.51 **1.81** 2.21	0.467
monocytes	10^9/L	56	0.330 **0.410** 0.598	0.318 **0.450** 0.559	0.946
sodium	MMOL/L	206	139 **141** 143	140 **141** 143	0.766
platelets	10^9/L	205	186 **226** 279	190 **224** 269	0.664
total protein	G/L	205	69 **71** 73	69 **71** 74	0.975
red blood cell count	10^{12}/L	205	4.30 **4.55** 4.93	4.40 **4.70** 5.10	0.186
uric acid	UMOL/L	206	274 **324** 396	261 **319** 387	0.452
white blood cell count	10^9/L	205	6.0 **6.8** 8.1	6.2 **7.3** 8.4	0.24

a *b* *c* represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. *N* is the number of non–missing values. Test used: Wilcoxon test

9 Programming

9.1 Methods

This report was produced using high-quality open source, freely available R and LaTeX packages. R uses virtually the same S language used by S-Plus but with some enhancements[3]. High-level S graphics and LaTeX table making functions in FE Harrell's Hmisc library were used. New S language functions `completenessReport`, `accrualReport`, `baselineReport`, `repVarclus`, `complianceReport`, `dropoutReport`, `aeReport`, and `labReport` were written to package these functions, using the philosophy of program-controlled generation of LaTeX text, figures, and tables. When figures were plotted in R, LaTeX figure legends and graphics insertion macro calls were automatically generated. Some of the functions produce both open (with pooling of treatment groups) and closed (stratifying on treatment) meeting reports. Automatically created graphics and `.tex` files for the open report have names beginning with `0`.

The `.pdf` file containing the report was generated using `pdflatex` so as to automatically generate hyperlinks (shown in blue) to all the figures and tables for easy navigation when viewing on the screen.

Tables were created using the Hmisc library's `summary.formula` function which also performs statistical tests across columns of tables using S functions.

Output for open meeting components is shown in a separate report.

9.2 Data Preparation

Variable labels are used in much of the graphical and tabular output, so it is advisable to attach `label` attributes to almost all variables. Variable names are used when `labels` are not defined. Units of measurement also appear in the output, so most continuous variables should have a `units` attribute. The `units` may contain mathematical expressions such as `cm^2` which will be properly typeset in tables and plots, using superscripts, subscripts, etc. Variables that are not binary (0/1, `Y/N`, etc.) but are categorical should have `levels` (value labels) defined (e.g., using the `factor` function) that will be attractive in the report. The Hmisc library `upData` function is useful for annotating variables with labels, units of measurement, and value labels. See Alzola and Harrell, 2006 and FE Harrell's lecture notes for details about setting up analysis files.

S code that created the analysis file for this report is shown below. For this particular application, `units` and some of the `labels` were actually obtained from separate data tables as shown in the code.

```
load('all.sav')
load('labcodes.sav')
load('ecgcodes.sav')
load('vitcodes.sav')
row.names(labcodes) ← tolower(labcodes$lablabel)
```

[3]The primary enhancements used here are mathematical, greek, and varying sized symbols in plots, and finer control over the names used for automatically-generated multi-page plots.

```
row.names(ecgcodes) ← tolower(ecgcodes$ecglabel)
row.names(vitcodes) ← tolower(vitcodes$vitlabel)

set.seed(13)
ns ← 250
id ← sample(unique(as.character(all$pid)), ns, replace=FALSE)
ssafety ← subset(all, pid %in% id)

for(i in names(ssafety)) {
  label(ssafety[[i]]) ← attr(all[[i]],'label')
  if(i %in% row.names(labcodes))
    units(ssafety[[i]]) ← as.character(labcodes[i,'lab.unit'])
  if(i %in% row.names(ecgcodes))
    units(ssafety[[i]]) ← tolower(as.character(ecgcodes[i,'ecg.unit']))
}
# labelled class was not used in all

v ← c(lab034='wbc', ecg008='hr', ecg007='uncorr.qt',
       lab016='creatinine', lab030='platelets', lab031='protein',
       lab003='lymphocytes.abs', lab014='bun', ecg002='axis',
       ae045='headache', ae104='ab.pain', ae109='dyspepsia',
       ae224='upper.resp.infect', ae392='coad',
       lab007='albumin', lab008='alk.phos', lab013='bilirubin',
       lab020='glucose', lab021='hematocrit', lab023='potassium',
       lab032='rbc', ecg001='atrial.rate', ecg003='corr.qt',
       ecg004='pr', ecg005='qrs', ecg006='rr',
       lab001='basophils.abs', lab002='eosinophils.abs',
       lab004='monocytes.abs',
       lab005='neutrophils', lab006='alat', lab009='amylase',
       lab010='asat', lab011='aty.lymph', lab012='basophils',
       lab015='chloride',lab017='eosinophils',lab018='ggt',
       lab019='glucose.fasting',lab022='hemoglobin',lab024='lymphocytes',
       lab025='monocytes',lab026='sodium',lab027='neutrophils.total',
       lab028='neutrophil.bands',lab029='neutrophils.seg',
       lab033='uric.acid',smk.stat="smoking")

# Generate uniformly distributed dates from 1990-1995
dates ← as.POSIXct(round(structure(365.25*24*60*60*(20+runif(ns,0,5)),
                                   class=c('POSIXt','POSIXct')),'days'))
# names(dates) ← id

# Get a compliance model.  Solve for a and b such that
# plogis(a+2b)=.99 and plogis(a+20b)=.75
b ← (qlogis(.75)-qlogis(.99))/18
a ← qlogis(.99)-2*b
# Note: compliance variable was not generated to be consistent with
# actual data availability.  Same for date of dropout.
d.dropout ← round(rexp(ns, 1/(7*40)))
dropout ← 1*(d.dropout <= 7*20)
d.dropout ← pmin(d.dropout, 7*20)
```

```
ssafety ←   upData(ssafety,
          id=as.integer(pid),
          trx=factor(ifelse(is.na(trx),ifelse(trx.seq==’P/P’,’A’,’B’),
            ifelse(trx==’PLACEBO’,’A’,’B’))),
          nausea  =ae126,
          diarrhea=ae077,
          smoking=ifelse(is.na(smoking),NA,1*(smoking==’Y’)),
          rdate=dates[id],
          site=factor(sample(1:20,ns,TRUE)[id]),
          comply=ifelse(week<=1, NA,
            ifelse(runif(length(week)) <= plogis(a+b*week), 1, 0)),
          d.dropout=d.dropout[id], dropout=dropout[id],
          rename=v,
          labels=c(trx=’Treatment’,rdate=’Randomization Date’,
            week=’Week’,nausea=’nausea’, diarrhea=’diarrhea’,
            sex=’Sex’, race=’Race’,
            comply=’Compliance’,
            d.dropout=’Days from Randomization to Dropout’,
            dropout=’Dropout/censoring Indicator’),
          units=c(d.dropout=’Day’),
          levels=list(sex=c(’female’,’male’)),
          drop=Cs(pid,trx.seq,ae126,ae077,prot,neutrophils.total))

save(ssafety, file=’ssafety.sav’, compress=T)
latex(describe(ssafety, format=’%d%b%y’))
summary(comply  ∼ week, data=ssafety)
```

9.3 User LaTeX Code for This Document

```
% To compile: pdflatex --shell-escape report (=pdflatexs report)
% Produces: report.pdf
\documentclass{article}
\usepackage{graphics}
\usepackage{relsize}    % for \smaller etc.
\usepackage{ctable}
\usepackage{moreverb}   % for \verbatimtabinput
\usepackage{sinput}     % local package for pretty printing of S code
\usepackage{fancyhdr}   % for fancy headers
\usepackage{lscape}     % for landscape tables (not used here)
\usepackage{color}      % for makeTreatKey function (draws grayscale
                        % line in LaTeX)
\usepackage{epic,calc}  % for micro dotcharts from summary.formula

\def\linkcol{blue}      % usually blue; can use black for hard copy
\newcommand{\titl}{Example Closed Meeting Data Monitoring Committee Report}
\usepackage[pdftex,bookmarks,pagebackref,pdfpagemode=UseOutlines,
    colorlinks,linkcolor=\linkcol,
    pdfauthor={Frank E Harrell Jr},
    pdftitle={\titl}]{hyperref}
```

```
% Remove colorlinks and linkcolor options to hyperref to box the
% hyperlinked items (for screen only)

\graphicspath{{pdf/}}

\newcommand{\scom}[1]{{\rm\scriptsize \# #1}}  % used by sinput
\newcommand{\code}[1]{\texttt{\smaller #1}}  % format software names
% smaller implemented by relsize: use 1 size smaller than current font

\author{Frank Harrell}
\title{\titl}
\date{\today}
\pagestyle{fancy}              % used for running headers, footers (rhead)
\renewcommand{\subsectionmark}[1]{} % suppress subsection titles in headers

\input{gentex/params.tex}             % created by sample.s; defines constants
                               % and LaTeX \treatkey macro for captions
\def\inclcode{1}               % 0=exclude code from report
\begin{document}
\maketitle
\tableofcontents
\listoffigures
\listoftables
\clearpage
\rhead{\scriptsize The {\em EXAMPLE} Study \\
    Protocol xyz--001 \\
    \today}

\section{Introduction}
This is a sample of the part of a closed meeting Data Monitoring
Committee report that contains software generated results.  Components
related to efficacy, study design, data monitoring
plan\footnote{Lan-DeMets monitoring bounds can be plotted using the
  open source S
  \href{http://biostat.mc.vanderbilt.edu/s/Hmisc/html/ldBands.html}
{\code{ldBands} function} in the Hmisc library for Linux and Unix systems.},
summary of previous closed report, interpretation, protocol changes,
screening, eligibility, and waiting time until treatment commencement
are not included in this example\footnote{See Ellenberg, Fleming, and
  DeMets, \emph{Data Monitoring Committees in Clinical Trials} (Wiley,
  2002), pp.\ 73-74 for recommended components in open and closed data
  monitoring committee reports.}.  This report used a random sample of
safety data from a randomized clinical trial.  Randomization date,
dropouts, and compliance variables were simulated, the latter two not
being made consistent with the presence or absence of actual data in
the random sample.  The date and time that the analysis file used here
was last updated was \datadate.  Source analysis files were last
updated on \primarydatadate.
\ifnum\inclcode=1{
 See Section~\ref{program} for information about software used.
```

```
\LaTeX's \code{hyperref} style was used to produce a \code{pdf}
file with hyperlinks for easy navigation to sections, tables, and
graphs using Adobe Acrobat Reader.  Internal hyperlinks are shown in
\linkcol, and external links to web sites are shown in red.
}\fi

See the example open meeting report for subject accrual, data
availability and completeness, and analyses not stratified by
treatment.

\section{Baseline Variables}
\input{gentex/baseline}
\clearpage

\section{Interrelationships Among Variables}
\input{gentex/varclus}
\clearpage

\section{Compliance to Assigned Treatments}
\input{gentex/compliance}
\clearpage

\section{Dropouts}
\input{gentex/dropout}
\clearpage

\section{Adverse Events}
\input{gentex/ae}
\clearpage

\section{EKG Data}
\input{gentex/ekg}
\clearpage

\section{Clinical Chemistry Data}
\input{gentex/chem}
\clearpage

\ifnum\inclcode=1{
\section{Programming}\label{program}
\subsection{Methods}
This report was produced using high-quality open source, freely
available R and \LaTeX\ packages.  R uses virtually the same S
language used by \textsc{S-Plus} but with some
enhancements\footnote{The primary enhancements used here are
  mathematical, greek, and varying sized symbols in plots, and finer
  control over the names used for automatically-generated multi-page
  plots.}.  High-level S graphics and \LaTeX\ table making functions
in FE Harrell's Hmisc library were used.  New S language functions
```

```
\code{completeness\-Report}, \code{accrual\-Report},
\code{baseline\-Report}, \code{rep\-Varclus},
\code{compliance\-Report}, \code{dropout\-Report}, \code{ae\-Report},
and \code{lab\-Report} were written to package these functions, using
the philosophy of program-controlled generation of \LaTeX\ text,
figures, and tables.  When figures were plotted in R, \LaTeX\ figure
legends and graphics insertion macro calls were automatically
generated.  Some of the functions produce both open (with pooling of
treatment groups) and closed (stratifying on treatment) meeting
reports.  Automatically created graphics and \code{.tex} files for
the open report have names beginning with \code{O}.

The \code{.pdf} file containing the report was generated using
\code{pdflatex} so as to automatically generate hyperlinks (shown in
blue) to all the figures and tables for easy navigation when viewing
on the screen.

Tables were created using the
\href{http://biostat.mc.vanderbilt.edu/s/Hmisc}
{Hmisc library}'s
\href{http://biostat.mc.vanderbilt.edu/twiki/pub/Main/StatReport/summary.pdf}
{\code{summary.formula}} function which also performs statistical
tests across columns of tables using S functions.

Output for open meeting components is shown in a separate report.

\subsection{Data Preparation}
Variable labels are used in much of the graphical and tabular output,
so it is advisable to attach \code{label} attributes to almost all
variables.  Variable names are used when \code{label}s are not
defined.  Units of measurement also appear in the output, so most
continuous variables should have a \code{units} attribute.  The
\code{units} may contain mathematical expressions such as
\verb|cm^2| which will be properly typeset in tables and plots, using
superscripts, subscripts, etc.  Variables that are not binary (0/1,
\code{Y/N}, etc.) but are categorical should have \code{levels}
(value labels) defined (e.g., using the \code{factor} function) that
will be attractive in the report.  The Hmisc library \code{upData}
function is useful for annotating variables with labels, units of
measurement, and value labels.  See
\href{http://biostat.mc.vanderbilt.edu/twiki/pub/Main/RS/sintro.pdf}
{Alzola and Harrell, 2006} and
\href{http://biostat.mc.vanderbilt.edu/twiki/pub/Main/StatCompCourse/sCompGraph.pdf}
{FE Harrell's lecture notes} for details about setting up analysis
files.

S code that created the analysis file for this report is shown
below.  For this particular application, \code{units} and some of
the \code{labels} were actually obtained from separate data tables
as shown in the code.
```

```
{\small\sinput{../createTest.s}}

\subsection{User \LaTeX\ Code for This Document}
{\small\verbatimtabinput{report.tex}}

\subsection{User S Code}
In the S code below, the \code{vars} object (a list) defines all the
vectors of variable names to analyze separately.  The \code{setdiff}
function is useful here, to define a set of variables except for a
specified vector of names (the second argument to \code{setdiff}).
In the future, some of the components of this list may be lists
themselves, for the adverse event data.  This will be used to organize
the variables by body system.
{\small\sinput{sample.s}}

\subsection{New Generic S Functions}
See the
\href{http://biostat.mc.vanderbilt.edu/Rreport}{\code{subversion} repository}.

}\fi

\end{document}
```

9.4 User S Code

In the S code below, the `vars` object (a list) defines all the vectors of variable names to analyze separately. The `setdiff` function is useful here, to define a set of variables except for a specified vector of names (the second argument to `setdiff`). In the future, some of the components of this list may be lists themselves, for the adverse event data. This will be used to organize the variables by body system.

```
library(Hmisc)
source('~/tmp/hmisc.s')
source('~/tmp/rreport.s')
load('../ssafety.sav')

# Save last modification date/time for source data files in
# LaTeX variables datadate and primarydatadate in file params.tex
cat('\\def\\datadate{',   format(file.info('../ssafety.sav')$mtime),'}\n',
    '\\def\\primarydatadate{',format(file.info('../all.sav')$mtime),'}\n',
    sep='', file='gentex/params.tex')

# List of lab variables that are missing too much to be used
omit ← Cs(amylase,aty.lymph,glucose.fasting,neutrophil.bands)
```

```
# Make a list that separates variables into major categories
vars ← list(baseline=Cs(age,sex,race,height,weight,bmi,smoking,pack.yrs),
            ae  =Cs(headache, ab.pain, nausea, dyspepsia, diarrhea,
                    upper.resp.infect, coad),
            ekg =setdiff(names(ssafety)[c(49:53,55:56)],'atrial.rate'),
            chem=setdiff(names(ssafety)[16:48],
              c(omit, Cs(lymphocytes.abs,atrial.rate,monocytes.abs,
                         neutrophils.seg,eosinophils.abs,basophils.abs))))


gtype ← c('ps','pdf','interactive')[2]

library(lattice)

week ← ssafety$week
weeks ← sort(unique(week))

base ← subset(ssafety, week==0)

# Make key for different line styles for inclusion in figure captions
makeTreatKey(levels(base$trx), append=TRUE)  # adds to params.tex

accrualReport(Minor=base$site, MinorLabel='site',
              EntryDate1=as.chron(base$rdate),
              EntryDate1cap='randomized subjects',
              dateRange=c('1990-01-01','1994-12-31'),
              targetDate='1994-12-31', targetN=300, hdotchart=4)

completenessReport(base, vars$baseline,  'baseline',
                   append=FALSE)
completenessReport(ssafety, vars$ae,     'ae',  week,
                   longPanel='adverse events')
completenessReport(ssafety, vars$ekg,    'ekg',
                   week, weeks[weeks!=1], longPanel='EKG')
completenessReport(ssafety, vars$chem,   'chem',
                   week, weeks[weeks %nin% c(1,16,20)],
                   longPanel='clinical chemistry')

complianceReport(ssafety$comply, ssafety$trx, ssafety$week,
                 weeks[weeks > 1])

baselineReport(base, vars$baseline, treat='trx', cdf=TRUE, long=FALSE)

repVarclus(ssafety[unlist(vars)], week, c(0,8))

dropoutReport(base$d.dropout, base$dropout, base$trx, time.inc=14)

aeReport(ssafety,  vars$ae, 'trx', 'week', weeks, 'id',
         times.tables=c(4,12), ylim=c(0,.15), forceBinary=TRUE)
```

```
labReport(ssafety, vars$ekg, 'ekg', 'trx', 'id', 'week', c(0,2,4,8),
          longPanel='EKG', cdf=c('corr.qt','pr'), clearPlots=TRUE)


labReport(ssafety, vars$chem, 'chem', 'trx', 'id', 'week', c(0,2,4,8,12),
          longPanel='Clinical Chemistry',
          cdf=c('rbc','asat','creatinine'))


if(gtype=='ps') dirps2pdf()  # if want to convert all new ps to pdf files
```

9.5 New Generic S Functions

See the `subversion` repository.