Example Closed Meeting Data Monitoring  
Committee Report  

Frank Harrell  
August 1, 2007  

Contents  
1 Introduction 3  
2 Baseline Variables 3  
3 Interrelationships Among Variables 6  
4 Compliance to Assigned Treatments 8  
5 Dropouts 9  
6 Adverse Events 10  
7 EKG Data 13  
8 Clinical Chemistry Data 18  
9 Programming 28  
9.1 Methods ................................................. 28  
9.2 Data Preparation ...................................... 28  
9.3 User L\TeX Code for This Document .................. 30  
9.4 User S Code ........................................ 34  
9.5 New Generic S Functions ............................. 36  

List of Figures  
1 Cumulative distribution plots of the continuous variables in the  
‘baseline’ table .............................................. 4  
2 Clustering of variables at Week 0 .......................... 6  
3 Clustering of variables at Week 8 ........................... 7  
4 Distribution of time until dropout from study ................ 9
List of Tables

1. Baseline ........................................... 5
2. Compliance by Treatment ..................... 8
3. Adverse Events at Week 4 ....................... 12
4. Adverse Events at Week 12 ..................... 12
5. Adverse Events at Any Time ................... 12
6. EKG Data at Week 0 ............................ 16
7. EKG Data at Week 2 ............................ 16
8. EKG Data at Week 4 ............................ 17
9. EKG Data at Week 8 ............................ 17
10. Clinical Chemistry Data at Week 0 ............. 18
11. Clinical Chemistry Data at Week 2 ............. 24
12. Clinical Chemistry Data at Week 4 ............. 25
13. Clinical Chemistry Data at Week 8 ............. 26
14. Clinical Chemistry Data at Week 12 .......... 27
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, summary of previous closed report, interpretation, protocol changes, screening, eligibility, and waiting time until treatment commencement are not included in this example. 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

---

1Lan-DeMets monitoring bounds can be plotted using the open source S \verb|lsbands| function in the Hmisc library for Linux and Unix systems.

2See 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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>250</td>
<td>60 65</td>
<td>70 61</td>
<td>0.104¹</td>
</tr>
<tr>
<td>Sex</td>
<td>250</td>
<td>78% (63)</td>
<td>73% (124)</td>
<td>0.453²</td>
</tr>
<tr>
<td></td>
<td>0% (0)</td>
<td>3% (5)</td>
<td></td>
<td>0.189²</td>
</tr>
<tr>
<td>Race</td>
<td>250</td>
<td>98% (79)</td>
<td>96% (163)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1% (1)</td>
<td>0% (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1% (1)</td>
<td>1% (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>250</td>
<td>164 169</td>
<td>175 168</td>
<td>173 0.185¹</td>
</tr>
<tr>
<td>weight</td>
<td>250</td>
<td>65 75</td>
<td>89 74</td>
<td>85 0.516¹</td>
</tr>
<tr>
<td>bmi</td>
<td>250</td>
<td>23.3 25.4</td>
<td>29.6 25.8</td>
<td>29.4 0.996¹</td>
</tr>
<tr>
<td>smoking</td>
<td>250</td>
<td>37% (30)</td>
<td>35% (59)</td>
<td>0.743²</td>
</tr>
<tr>
<td>pack_yrs</td>
<td>250</td>
<td>30 48</td>
<td>60 42</td>
<td>60 0.862¹</td>
</tr>
</tbody>
</table>

*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: ¹Wilcoxon test; ²Pearson 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 $\rho$ 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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Compliance</td>
<td>N</td>
</tr>
<tr>
<td>Week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>81</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>81</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>81</td>
<td>0.99</td>
</tr>
<tr>
<td>12</td>
<td>81</td>
<td>0.93</td>
</tr>
<tr>
<td>16</td>
<td>81</td>
<td>0.84</td>
</tr>
<tr>
<td>20</td>
<td>81</td>
<td>0.78</td>
</tr>
<tr>
<td>Overall</td>
<td>486</td>
<td>0.92</td>
</tr>
</tbody>
</table>
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 \(\chi^2\) tests.
Figure 6: Kaplan-Meier estimates of cumulative probabilities of adverse events by treatment over time (continued)
Table 3: Adverse Events at Week 4

<table>
<thead>
<tr>
<th></th>
<th>N = 81</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>headache</td>
<td>246</td>
<td>2% (2)</td>
<td>2% (4)</td>
<td>0.983</td>
</tr>
<tr>
<td>abdominal pain</td>
<td>246</td>
<td>1% (1)</td>
<td>2% (4)</td>
<td>0.534</td>
</tr>
<tr>
<td>nausea</td>
<td>246</td>
<td>0% (0)</td>
<td>1% (2)</td>
<td>0.32</td>
</tr>
<tr>
<td>dyspepsia</td>
<td>246</td>
<td>0% (0)</td>
<td>2% (3)</td>
<td>0.222</td>
</tr>
<tr>
<td>diarrhea</td>
<td>246</td>
<td>0% (0)</td>
<td>2% (3)</td>
<td>0.222</td>
</tr>
<tr>
<td>upper resp tract infection</td>
<td>246</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td></td>
</tr>
<tr>
<td>chronic obstructive airways disease</td>
<td>246</td>
<td>6% (5)</td>
<td>6% (10)</td>
<td>0.972</td>
</tr>
</tbody>
</table>

Table 4: Adverse Events at Week 12

<table>
<thead>
<tr>
<th></th>
<th>N = 81</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>headache</td>
<td>241</td>
<td>1% (1)</td>
<td>0% (0)</td>
<td>0.151</td>
</tr>
<tr>
<td>abdominal pain</td>
<td>241</td>
<td>1% (1)</td>
<td>2% (3)</td>
<td>0.738</td>
</tr>
<tr>
<td>nausea</td>
<td>241</td>
<td>1% (1)</td>
<td>1% (1)</td>
<td>0.602</td>
</tr>
<tr>
<td>dyspepsia</td>
<td>241</td>
<td>0% (0)</td>
<td>1% (1)</td>
<td>0.484</td>
</tr>
<tr>
<td>diarrhea</td>
<td>241</td>
<td>0% (0)</td>
<td>1% (1)</td>
<td>0.484</td>
</tr>
<tr>
<td>upper resp tract infection</td>
<td>241</td>
<td>0% (0)</td>
<td>1% (2)</td>
<td>0.321</td>
</tr>
<tr>
<td>chronic obstructive airways disease</td>
<td>241</td>
<td>9% (7)</td>
<td>6% (10)</td>
<td>0.444</td>
</tr>
</tbody>
</table>

Table 5: Adverse Events at Any Time

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>headache</td>
<td>6% (5)</td>
<td>7% (11)</td>
<td>0.919</td>
</tr>
<tr>
<td>abdominal pain</td>
<td>5% (4)</td>
<td>15% (25)</td>
<td>0.023</td>
</tr>
<tr>
<td>nausea</td>
<td>4% (3)</td>
<td>7% (12)</td>
<td>0.29</td>
</tr>
<tr>
<td>dyspepsia</td>
<td>2% (2)</td>
<td>7% (11)</td>
<td>0.178</td>
</tr>
<tr>
<td>diarrhea</td>
<td>4% (3)</td>
<td>14% (23)</td>
<td>0.016</td>
</tr>
<tr>
<td>upper resp tract infection</td>
<td>2% (2)</td>
<td>8% (13)</td>
<td>0.104</td>
</tr>
<tr>
<td>chronic obstructive airways disease</td>
<td>32% (26)</td>
<td>20% (34)</td>
<td>0.038</td>
</tr>
</tbody>
</table>
7 EKG Data

Figure 7: Quartiles of EKG variables over time. Outer lines are 25th (lower line) and 75th (upper line) percentiles. Thicker middle lines depict medians. $y$-axis is scaled to the pooled 5th and 95th 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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>81</td>
<td>169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axis</td>
<td>243</td>
<td>-4.00</td>
<td>41.00</td>
<td>62.50</td>
</tr>
<tr>
<td>corrected qt</td>
<td>243</td>
<td>408 42 441</td>
<td>411 422 439</td>
<td>0.607</td>
</tr>
<tr>
<td>pr</td>
<td>239</td>
<td>142 156 174</td>
<td>147 156 174</td>
<td>0.521</td>
</tr>
<tr>
<td>qrs</td>
<td>243</td>
<td>82 92 100</td>
<td>80 92 100</td>
<td>0.871</td>
</tr>
<tr>
<td>uncorrected qt</td>
<td>243</td>
<td>370 384 412</td>
<td>368 388 412</td>
<td>0.661</td>
</tr>
<tr>
<td>ventricular rate</td>
<td>243</td>
<td>63.0 70.0 78.0</td>
<td>64.0 72.5 82.0</td>
<td>0.23</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>81</td>
<td>169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axis</td>
<td>232</td>
<td>-2.0 43.0 62.8</td>
<td>0.5 42.0 67.0</td>
<td>0.886</td>
</tr>
<tr>
<td>corrected qt</td>
<td>232</td>
<td>408 423 438</td>
<td>407 419 437</td>
<td>0.375</td>
</tr>
<tr>
<td>pr</td>
<td>228</td>
<td>144 156 168</td>
<td>140 156 172</td>
<td>0.772</td>
</tr>
<tr>
<td>qrs</td>
<td>232</td>
<td>84 92 100</td>
<td>80 92 100</td>
<td>0.564</td>
</tr>
<tr>
<td>uncorrected qt</td>
<td>232</td>
<td>373 392 416</td>
<td>364 384 404</td>
<td>0.06</td>
</tr>
<tr>
<td>ventricular rate</td>
<td>232</td>
<td>63.0 70.0 79.8</td>
<td>66.0 74.0 81.8</td>
<td>0.098</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=81</td>
<td>N=169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axis degree</td>
<td>230</td>
<td>-2.00</td>
<td>43.00</td>
<td>66.75</td>
</tr>
<tr>
<td>corrected qt</td>
<td>230</td>
<td>409</td>
<td>420</td>
<td>439</td>
</tr>
<tr>
<td>pr msec</td>
<td>226</td>
<td>144</td>
<td>156</td>
<td>168</td>
</tr>
<tr>
<td>qrs msec</td>
<td>230</td>
<td>80</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>uncorrected qt</td>
<td>230</td>
<td>372</td>
<td>388</td>
<td>408</td>
</tr>
<tr>
<td>ventricular rate bpm</td>
<td>230</td>
<td>62.5</td>
<td>71.0</td>
<td>77.0</td>
</tr>
</tbody>
</table>

* 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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=81</td>
<td>N=169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axis degree</td>
<td>213</td>
<td>-6.25</td>
<td>32.50</td>
<td>65.00</td>
</tr>
<tr>
<td>corrected qt</td>
<td>213</td>
<td>410</td>
<td>421</td>
<td>431</td>
</tr>
<tr>
<td>pr msec</td>
<td>208</td>
<td>140</td>
<td>156</td>
<td>172</td>
</tr>
<tr>
<td>qrs msec</td>
<td>213</td>
<td>80</td>
<td>88</td>
<td>96</td>
</tr>
<tr>
<td>uncorrected qt</td>
<td>213</td>
<td>365</td>
<td>384</td>
<td>404</td>
</tr>
<tr>
<td>ventricular rate bpm</td>
<td>213</td>
<td>63.2</td>
<td>72.5</td>
<td>80.0</td>
</tr>
</tbody>
</table>

* 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

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

$a$ $b$ $c$ 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\textsuperscript{th} (lower line) and 75\textsuperscript{th} (upper line) percentiles. Thicker middle lines depict medians. \textit{y}-axis is scaled to the pooled 5\textsuperscript{th} and 95\textsuperscript{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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutrophils absolute</td>
<td>10^9/L</td>
<td>229</td>
<td>3.40 4.06 5.14</td>
<td>3.67 4.30 5.53</td>
</tr>
<tr>
<td>alanine aminotransferase</td>
<td>IU/L</td>
<td>228</td>
<td>12 16 26</td>
<td>12 16 22</td>
</tr>
<tr>
<td>albumin</td>
<td>g/L</td>
<td>228</td>
<td>41 43 44</td>
<td>40 42 44</td>
</tr>
<tr>
<td>alkaline phosphatase</td>
<td>IU/L</td>
<td>228</td>
<td>61.0 73.0 86.5</td>
<td>62.0 74.0 91.0</td>
</tr>
<tr>
<td>aspartate aminotransferase</td>
<td>IU/L</td>
<td>228</td>
<td>15.0 19.0 23.5</td>
<td>15.0 18.0 22.0</td>
</tr>
<tr>
<td>basophils</td>
<td>10^9/L</td>
<td>64</td>
<td>0.0208 0.0294 0.0384</td>
<td>0.0183 0.0364 0.0616</td>
</tr>
<tr>
<td>total bilirubin</td>
<td>UMOL/L</td>
<td>228</td>
<td>8.28 10.00 13.68</td>
<td>8.00 10.00 13.00</td>
</tr>
<tr>
<td>blood urea nitrogen</td>
<td>MMOL/L</td>
<td>229</td>
<td>4.28 5.20 6.40</td>
<td>4.70 5.70 7.10</td>
</tr>
<tr>
<td>chloride</td>
<td>MMOL/L</td>
<td>230</td>
<td>102 104 106</td>
<td>101 104 106</td>
</tr>
<tr>
<td>creatinine</td>
<td>UMOL/L</td>
<td>228</td>
<td>61.9 72.0 88.4</td>
<td>63.0 72.0 83.0</td>
</tr>
<tr>
<td>eosinophils</td>
<td>10^9/L</td>
<td>64</td>
<td>0.137 0.250 0.354</td>
<td>0.137 0.224 0.345</td>
</tr>
<tr>
<td>γ glutamyl transferase</td>
<td>IU/L</td>
<td>228</td>
<td>21.5 33.0 52.5</td>
<td>19.0 25.0 38.0</td>
</tr>
<tr>
<td>glucose - random</td>
<td>MMOL/L</td>
<td>215</td>
<td>5.00 5.55 6.26</td>
<td>5.00 5.50 6.29</td>
</tr>
<tr>
<td>hematocrit</td>
<td>%</td>
<td>230</td>
<td>41.6 44.2 46.6</td>
<td>41.1 43.5 46.5</td>
</tr>
<tr>
<td>hemoglobin</td>
<td>g/L</td>
<td>230</td>
<td>140 149 158</td>
<td>137 147 156</td>
</tr>
<tr>
<td>potassium</td>
<td>MMOL/L</td>
<td>229</td>
<td>4.20 4.40 4.75</td>
<td>4.30 4.50 4.70</td>
</tr>
<tr>
<td>lymphocytes</td>
<td>10^9/L</td>
<td>64</td>
<td>1.48 1.80 2.68</td>
<td>1.57 1.80 2.09</td>
</tr>
<tr>
<td>monocytes</td>
<td>10^9/L</td>
<td>64</td>
<td>0.363 0.477 0.796</td>
<td>0.334 0.470 0.581</td>
</tr>
<tr>
<td>sodium</td>
<td>MMOL/L</td>
<td>230</td>
<td>139 140 142</td>
<td>139 141 142</td>
</tr>
<tr>
<td>platelets</td>
<td>10^12/L</td>
<td>230</td>
<td>179 210 250</td>
<td>191 224 268</td>
</tr>
<tr>
<td>total protein</td>
<td>g/L</td>
<td>229</td>
<td>68 71 73</td>
<td>68 71 74</td>
</tr>
<tr>
<td>red blood cell count</td>
<td>10^12/L</td>
<td>230</td>
<td>4.46 4.60 4.98</td>
<td>4.30 4.70 5.00</td>
</tr>
<tr>
<td>uric acid</td>
<td>UMOL/L</td>
<td>229</td>
<td>252 318 385</td>
<td>263 313 379</td>
</tr>
<tr>
<td>white blood cell count</td>
<td>10^9/L</td>
<td>230</td>
<td>5.93 6.70 7.78</td>
<td>6.17 7.10 8.33</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 81</td>
<td>N = 169</td>
<td></td>
</tr>
<tr>
<td>neutrophils absolute</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>226</td>
<td>3.34</td>
<td>4.28</td>
</tr>
<tr>
<td>alanine aminotransferase</td>
<td>IU/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>12.0</td>
<td>15.0</td>
</tr>
<tr>
<td>albumin</td>
<td>g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>224</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>alkaline phosphatase</td>
<td>IU/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>62.0</td>
<td>73.0</td>
</tr>
<tr>
<td>aspartate aminotransferase</td>
<td>IU/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>basophils</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>0.0154</td>
<td>0.0436</td>
</tr>
<tr>
<td>total bilirubin</td>
<td>UMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>8.00</td>
<td>11.00</td>
</tr>
<tr>
<td>blood urea nitrogen</td>
<td>MMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>4.28</td>
<td>5.30</td>
</tr>
<tr>
<td>chloride</td>
<td>MMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>226</td>
<td>101</td>
<td>104</td>
</tr>
<tr>
<td>creatinine</td>
<td>UMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>224</td>
<td>62.0</td>
<td>75.0</td>
</tr>
<tr>
<td>eosinophils</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>0.154</td>
<td>0.226</td>
</tr>
<tr>
<td>γ glutamyl transferase</td>
<td>IU/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>224</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>glucose - random</td>
<td>MMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213</td>
<td>4.83</td>
<td>5.44</td>
</tr>
<tr>
<td>hematocrit</td>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>227</td>
<td>41.1</td>
<td>44.5</td>
</tr>
<tr>
<td>hemoglobin</td>
<td>g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>227</td>
<td>140</td>
<td>149</td>
</tr>
<tr>
<td>potassium</td>
<td>MMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>226</td>
<td>4.3</td>
<td>4.5</td>
</tr>
<tr>
<td>lymphocytes</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>1.62</td>
<td>1.92</td>
</tr>
<tr>
<td>monocytes</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>0.335</td>
<td>0.508</td>
</tr>
<tr>
<td>sodium</td>
<td>MMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>226</td>
<td>139</td>
<td>141</td>
</tr>
<tr>
<td>platelets</td>
<td>10^9/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>227</td>
<td>179</td>
<td>208</td>
</tr>
<tr>
<td>total protein</td>
<td>g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>68</td>
<td>70</td>
</tr>
<tr>
<td>red blood cell count</td>
<td>10^12/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>227</td>
<td>4.4</td>
<td>4.6</td>
</tr>
<tr>
<td>uric acid</td>
<td>UMOL/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>256</td>
<td>316</td>
</tr>
<tr>
<td>white blood cell count</td>
<td>10^12/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>227</td>
<td>5.83</td>
<td>6.70</td>
</tr>
</tbody>
</table>

abc 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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N = 81</td>
<td>N = 169</td>
<td></td>
</tr>
<tr>
<td>neutrophils absolute</td>
<td>210</td>
<td>3.62 <strong>4.39</strong> 5.11</td>
<td>3.55 <strong>4.22</strong> 5.36</td>
<td>0.996</td>
</tr>
<tr>
<td>alanine aminotransferase</td>
<td>209</td>
<td>12.0 <strong>15.0</strong> 27.0</td>
<td>12.0 <strong>16.0</strong> 21.8</td>
<td>0.38</td>
</tr>
<tr>
<td>albumin</td>
<td>210</td>
<td>41 <strong>43</strong> 44</td>
<td>41 <strong>42</strong> 44</td>
<td>0.102</td>
</tr>
<tr>
<td>alkaline phosphatase</td>
<td>209</td>
<td>62.5 <strong>71.0</strong> 85.0</td>
<td>62.0 <strong>72.5</strong> 86.0</td>
<td>0.556</td>
</tr>
<tr>
<td>aspartate aminotransferase</td>
<td>209</td>
<td>16.0 <strong>18.0</strong> 23.5</td>
<td>15.0 <strong>18.0</strong> 21.8</td>
<td>0.24</td>
</tr>
<tr>
<td>basophils</td>
<td>59</td>
<td>0.0126 <strong>0.0276</strong> 0.0323</td>
<td>0.0190 <strong>0.0300</strong> 0.0442</td>
<td>0.163</td>
</tr>
<tr>
<td>total bilirubin</td>
<td>209</td>
<td>8.0 <strong>11.0</strong> 13.5</td>
<td>7.0 <strong>10.0</strong> 13.0</td>
<td>0.186</td>
</tr>
<tr>
<td>blood urea nitrogen</td>
<td>210</td>
<td>4.63 <strong>5.60</strong> 6.78</td>
<td>4.28 <strong>5.35</strong> 6.70</td>
<td>0.543</td>
</tr>
<tr>
<td>chloride</td>
<td>210</td>
<td>102 <strong>104</strong> 106</td>
<td>102 <strong>104</strong> 106</td>
<td>0.881</td>
</tr>
<tr>
<td>creatinine</td>
<td>209</td>
<td>61.4 <strong>79.6</strong> 91.0</td>
<td>62.0 <strong>73.5</strong> 83.0</td>
<td>0.321</td>
</tr>
<tr>
<td>eosinophils</td>
<td>59</td>
<td>0.132 <strong>0.214</strong> 0.302</td>
<td>0.127 <strong>0.182</strong> 0.311</td>
<td>0.856</td>
</tr>
<tr>
<td>γ glutamyl transferase</td>
<td>209</td>
<td>21.0 <strong>29.0</strong> 44.5</td>
<td>18.2 <strong>25.0</strong> 36.0</td>
<td>0.089</td>
</tr>
<tr>
<td>glucose - random</td>
<td>200</td>
<td>5.00 <strong>5.49</strong> 6.80</td>
<td>4.86 <strong>5.70</strong> 6.36</td>
<td>0.98</td>
</tr>
<tr>
<td>hematocrit</td>
<td>210</td>
<td>42.6 <strong>44.4</strong> 47.9</td>
<td>40.8 <strong>44.3</strong> 46.6</td>
<td>0.431</td>
</tr>
<tr>
<td>hemoglobin</td>
<td>210</td>
<td>140 <strong>149</strong> 158</td>
<td>138 <strong>147</strong> 156</td>
<td>0.251</td>
</tr>
<tr>
<td>potassium</td>
<td>209</td>
<td>4.40 <strong>4.50</strong> 4.65</td>
<td>4.20 <strong>4.50</strong> 4.70</td>
<td>0.691</td>
</tr>
<tr>
<td>lymphocytes</td>
<td>59</td>
<td>1.69 <strong>1.92</strong> 2.39</td>
<td>1.59 <strong>1.77</strong> 2.19</td>
<td>0.392</td>
</tr>
<tr>
<td>monocytes</td>
<td>59</td>
<td>0.350 <strong>0.419</strong> 0.572</td>
<td>0.348 <strong>0.445</strong> 0.553</td>
<td>0.887</td>
</tr>
<tr>
<td>sodium</td>
<td>210</td>
<td>139 <strong>141</strong> 142</td>
<td>139 <strong>141</strong> 143</td>
<td>0.48</td>
</tr>
<tr>
<td>platelets</td>
<td>210</td>
<td>182 <strong>214</strong> 267</td>
<td>195 <strong>225</strong> 269</td>
<td>0.305</td>
</tr>
<tr>
<td>total protein</td>
<td>210</td>
<td>68.8 <strong>71.0</strong> 74.0</td>
<td>68.0 <strong>71.0</strong> 74.0</td>
<td>0.754</td>
</tr>
<tr>
<td>red blood cell count</td>
<td>210</td>
<td>4.4 <strong>4.7</strong> 4.9</td>
<td>4.4 <strong>4.7</strong> 5.1</td>
<td>0.252</td>
</tr>
<tr>
<td>uric acid</td>
<td>210</td>
<td>253 <strong>319</strong> 387</td>
<td>262 <strong>318</strong> 389</td>
<td>0.724</td>
</tr>
<tr>
<td>white blood cell count</td>
<td>210</td>
<td>6.22 <strong>6.85</strong> 8.10</td>
<td>5.98 <strong>6.90</strong> 8.03</td>
<td>0.647</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophils absolute</td>
<td>205</td>
<td>205</td>
<td>81</td>
</tr>
<tr>
<td>Alanine aminotransferase</td>
<td>205</td>
<td>205</td>
<td>12.0</td>
</tr>
<tr>
<td>Albumin</td>
<td>205</td>
<td>205</td>
<td>42</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>205</td>
<td>205</td>
<td>60.0</td>
</tr>
<tr>
<td>Aspartate aminotransferase</td>
<td>205</td>
<td>205</td>
<td>15</td>
</tr>
<tr>
<td>Basophils</td>
<td>56</td>
<td>56</td>
<td>0.0137</td>
</tr>
<tr>
<td>Total bilirubin</td>
<td>204</td>
<td>205</td>
<td>8.41</td>
</tr>
<tr>
<td>Blood urea nitrogen</td>
<td>205</td>
<td>206</td>
<td>4.30</td>
</tr>
<tr>
<td>Chloride</td>
<td>206</td>
<td>205</td>
<td>100</td>
</tr>
<tr>
<td>Creatinine</td>
<td>205</td>
<td>206</td>
<td>61.9</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>56</td>
<td>56</td>
<td>0.126</td>
</tr>
<tr>
<td>Glucose - random</td>
<td>194</td>
<td>205</td>
<td>5.05</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>205</td>
<td>205</td>
<td>41.5</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>205</td>
<td>205</td>
<td>140</td>
</tr>
<tr>
<td>Potassium</td>
<td>205</td>
<td>205</td>
<td>4.3</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>56</td>
<td>56</td>
<td>1.39</td>
</tr>
<tr>
<td>Monocytes</td>
<td>56</td>
<td>56</td>
<td>0.330</td>
</tr>
<tr>
<td>Sodium</td>
<td>206</td>
<td>206</td>
<td>139</td>
</tr>
<tr>
<td>Platelets</td>
<td>205</td>
<td>205</td>
<td>186</td>
</tr>
<tr>
<td>Total protein</td>
<td>205</td>
<td>205</td>
<td>69</td>
</tr>
<tr>
<td>Red blood cell count</td>
<td>205</td>
<td>206</td>
<td>4.38</td>
</tr>
<tr>
<td>Uric acid</td>
<td>206</td>
<td>206</td>
<td>274</td>
</tr>
<tr>
<td>White blood cell count</td>
<td>205</td>
<td>205</td>
<td>6.0</td>
</tr>
</tbody>
</table>

*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\textsuperscript{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 $O$.

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\textsuperscript{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.

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

\textsuperscript{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='gtt',
    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)
\begin{verbatim}
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)
l latex(describe(ssafety, format = '%d%b%y'))
summary(comply ~ week, data = ssafety)
\end{verbatim}

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 \verbatimtableinput
\usepackage{sinput} % 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,
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]{{\tt\smaller #1}} % format software names
% smaller implemented by relsize: use 1 size smaller than current font
\author{Frank Harrell}
\title{The \textit{EXAMPLE} Study \ Protocol xyz--001 \ August 1, 2007}
\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 \textit{EXAMPLE} Study \ Protocol xyz--001 \ August 1, 2007}
\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, interpretation, protocol changes, screening, eligibility, and waiting time until treatment commencement are not included in this example. See Ellenberg, Fleming, and DeMets, \textit{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 \texttt{hyperref} style was used to produce a \texttt{pdf} file with hyperlinks for easy navigation to sections, tables, and graphs using Adobe Acrobat Reader. Internal hyperlinks are shown in \texttt{linkcol}, 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.

\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

32
The \textit{EXAMPLE} Study
Protocol xyz-001
August 1, 2007

The \texttt{completeness-Report}, \texttt{accrual-Report}, \texttt{baseline-Report}, \texttt{rep-Varclus}, \\texttt{compliance-Report}, \texttt{dropout-Report}, \texttt{ae-Report}, and \texttt{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 \texttt{.tex} files for the open report have names beginning with \texttt{O}.

The \texttt{.pdf} file containing the report was generated using \texttt{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 \texttt{Hmisc} library's \texttt{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.

\section*{Data Preparation}

Variable labels are used in much of the graphical and tabular output, so it is advisable to attach \texttt{label} attributes to almost all variables. Variable names are used when \texttt{label}s are not defined. Units of measurement also appear in the output, so most continuous variables should have a \texttt{units} attribute. The \texttt{units} may contain mathematical expressions such as \texttt{cm^2} which will be properly typeset in tables and plots, using superscripts, subscripts, etc. Variables that are not binary (0/1, \texttt{Y/N}, etc.) but are categorical should have \texttt{levels} (value labels) defined (e.g., using the \texttt{factor} function) that will be attractive in the report. The \texttt{Hmisc} library \texttt{upData} function is useful for annotating variables with labels, units of measurement, and value labels. See \texttt{Alzola and Harrell, 2006} and \texttt{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, \texttt{units} and some of the \texttt{labels} were actually obtained from separate data tables as shown in the code.

\begin{verbatim}
end{verbatim}
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.

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

```s
# 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')
```

```s
# 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.