Example Closed Meeting Data Monitoring Committee Report

Frank Harrell

April 6, 2017

Contents

List of Figures

List of Tables

```
Load(ssafety)
ssafety ← upData(ssafety, rdate=as.Date(rdate),
                 smoking=factor(smoking, 0:1, c('No',
                      'Yes')),
                 labels=c(smoking='Smoking', bmi='BMI
                      ',
                   pack.yrs='Pack Years', age='Age',
                   height='Height', weight='Weight'),
                 units=c(age='years', height='cm',
                      weight='Kg'),
                 print=FALSE)
mtime ← function(f) format(file.info(f)$mtime)
datadate        ← mtime('ssafety.rda')
primarydatadate ← mtime('ssafety.rda')

## 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))))
week  ← ssafety$week
weeks ← sort(unique(week))
base  ← subset(ssafety, week==0)
denom ← c(c(enrolled=500, randomized=nrow(base)),
      table(base$trx))

setgreportOption(tx.var='trx', denom=denom, texwhere='
```

```
      ')
## Initialize app.tex
file ← sprintf('%s/app.tex', getgreportOption('texdir
      '))
cat('', file=file)
```

1 Philosophy

The reporting tools used here are based on a number of lessons learned from the intersection of the fields of statistical graphics, graphic design, and cognitive psychology, especially from the work of Bill Cleveland, Ralph McGill, John Tukey, Edward Tufte, and Jacques Bertin.

1. Whenever largely numerical information is displayed, graphs convey the information most often needed much better than tables.
   (a) Tables usually show more precision than is warranted by the sample information while hiding important features.
   (b) Graphics are much better than tables for seeing patterns and anomalies.
2. The best graphics are ones that make use of features that humans are most accurate in perceiving, namely position along a common scale.
3. Information across multiple data categories is usually easier to judge when the categories are sorted by the numeric quantity underlying the information[1].
4. The most robust and informative descriptive statistics for continuous variables are quantiles and whole distribution summaries[2].
5. For group comparisons, confidence intervals for individual means, medians, or proportions are not very useful, and whether or not two confidence intervals overlap is not the correct statistical approach for judging the significance of the difference between the two. The half-width of the confidence interval for the difference, when centered at the midpoint of the two estimates, provides a succinct precision display, and this half-interval touches the two estimates if and only if there is no significant difference between the two.
6. Each graphic needs a marker that provides the reader with a sense of exactly what fraction of the sample is being analyzed in that graphic.

[1] This also facilitates multivariate understanding of trends and differences. For example, if one sorted countries by the fraction of subjects who died and displayed also the fraction of subjects who suffered a stroke, the extent to which stroke incidence is also sorted by country is a measure of the correlation between mortality and stroke incidence across countries.

[2] In particular, the standard deviation is not very meaningful for asymmetric distributions, and is not robust to outliers.

7. Tables are best used as backups to graphics.

8. Tables should emphasize estimates that are not functions of the sample size. For categorical variables, proportions have interpretations independent of sample size so they are the featured estimates, and numerators and denominators are subordinate to the proportions. For continuous variables, minimum and maximum, while useful for data quality checking, are not population parameters, and they expand as $n \uparrow$, so they are not proper summary statistics.

2 Notation

Hyperlinks and Tables Some graphics and tables are hyperlinked to tables in the Appendix. For these, clicking anywhere in the graphic or table will move the pdf reader to the supporting table. Clicking on the appendix table will bring you back to the original figure. Other than for graphics, objects appearing in this color are hyperlinked.

Viewers You must use Adobe Acrobat Reader to view pdf files generated by `greport`, otherwise pop-ups will not work. Neither pop-ups nor hyperlinks will work if you view documents in a Web browser window. It is recommended that you click on `View ...Page Display ...Single Page` for optimum jumping between hyperlinks, i.e., do not use `Single Page Continuous` mode.

Figure Captions Needles represent the fraction of observations used in the current analysis. The first needle (red) shows the fraction of enrolled patients used. If randomization was taken into account, a second needle (green) represents the fraction of randomized subjects included in the analysis. When the analyses consider treatment assignment, two more needles may be added to the display, showing, respectively, the fraction of subjects randomized to treatment A used in the analysis and the fraction of subjects on treatment B who were analyzed. The colors of these last two needles are the colors used for the two treatments throughout the report. The following table shows some examples.

```
# dNeedle uses colors in setgreportOption(tx.col=,
      er.col=)
dNeedle(1,              'lttdemoa')
```

```
dNeedle(c(3,4)/4 ,    'lttdemob')
```

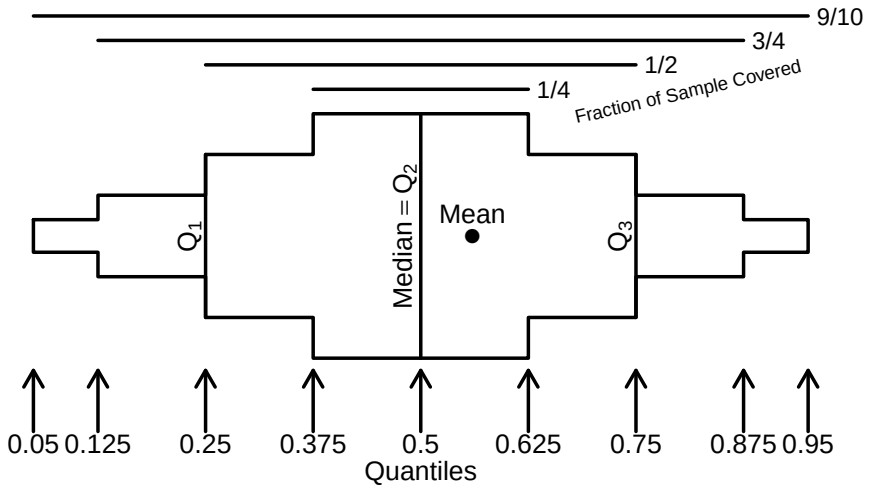
```
dNeedle(c(1,2)/4,     'lttdemoc')
```

```
dNeedle(c(1,2,3,1)/4,'lttdemod')
```

Needles	Interpretation
	All enrolled subjects analyzed, randomization not considered
	Analysis uses $\frac{3}{4}$ of enrolled subjects, and all randomized subjects
	Analysis uses $\frac{1}{4}$ of enrolled subjects, and $\frac{1}{2}$ of randomized subjects
	Same as previous example, and in addition the analysis utilized treatment assignment, analyzing $\frac{3}{4}$ of those randomized to A and $\frac{1}{4}$ of those randomized to B

Extended Box Plots For depicting distributions of continuous variables, many of the following displays use extended box plots, also called box–percentile plots. The plots are scaled to the marginal 0.01 and 0.99 quantiles of the variables. A prototype, with explanations, is below. When viewing the report, hovering the mouse over the word "box" will pop up this prototype as needed.

```
bpplt()
```



Typically violin plots are superimposed onto box plots in what follows. Violin plots show mirror images of the estimated probability density function for continuous variable, using a kernel density estimator. Violin plots are better able to show multimodality than quantile intervals.

Dot Charts Dot charts are used to present stratified proportions. In these charts the area of the symbols is proportional to the square root of the denom-

inator. The legend shows representative denominators and their corresponding symbol areas, using denominators that actually occurred in the data and extended from the minimum observed to the maximum observed sample size.

Longitudinal Analysis For continuous variables measured repeatedly, line plots show the median as a function of time. Next to each point is a series of thin vertical lines, one series on the left for treatment A and another series on the right for treatment B, when stratifying by treatment. Moving outward from the point showing the median, these lines depict the following quantile intervals—the same ones depicted in the extended box plots.

Horizontal Sequence	Quantiles Spanned			Fraction of Sample Covered
1	$\frac{1}{20}$	-	$\frac{19}{20}$	$\frac{9}{10}$
2	$\frac{1}{8}$	-	$\frac{7}{8}$	$\frac{3}{4}$
3	$\frac{1}{4}$	-	$\frac{3}{4}$	$\frac{1}{2}$
4	$\frac{3}{8}$	-	$\frac{5}{8}$	$\frac{1}{4}$

In addition there is a black vertical line centered at the midpoint of the two medians, with height equal to $\frac{1}{2}$ of the width of an approximate 0.95 confidence interval for the difference in the two medians. When the two medians touch this vertical bar, there is approximately no significant difference in the medians at the 0.05 level. The Harrell–Davis quantile estimator is used, along with its standard error estimate for the median.

Instead of quantile intervals, longitudinal plots may show vertical violin plots. Unlike the mirror–image violin plots shown on box plots, there there are two groups being compared the first group has a half-violin plot on the left of the point showing the median, and the second group has a half-violin plot on the right. Back–to–back comparisons of probability density functions are useful for comparing entire distributions of continuous variables. When for a group the sample size is less than 10 the violins are more faint, and when the sample size is less than 5 they are barely visible.

For binary variables measured repeatedly, the 0.95 Wilson confidence intervals are shown on either side of the proportions, and a vertical black line appears over the proportions. The height of this line is $\frac{1}{2}$ the length of the normal-approximation 0.95 confidence interval for the difference in the two proportions. When the proportions fail to touch this line, they are approximately significantly different at (at least) the 0.05 level.

For discrete variables that are not binary, such as adverse event presence and severity, means and bootstrap percentile 0.95 confidence intervals are shown, along with half-confidence intervals for the difference in means using as standard error the square root of the sum of squares of the two means' standard errors.

Survival Curves Graphs containing pairs of Kaplan-Meier survival curves show a shaded region centered at the midpoint of the two survival estimates and having a height equal to the half-width of the approximate 0.95 pointwise

confidence interval for the difference of the two survival probabilities. Time points at which the two survival estimates do not touch the shaded region denote approximately significantly different survival estimates, without any multiplicity correction.

3 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,[3] summary of previous closed report, interpretation, protocol changes, screening, eligibility, and waiting time until treatment commencement are not included in this example[4]. 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 2013-10-27 10:50:46. Source analysis files were last updated on 2013-10-27 10:50:46. See Section 12 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 darkblue, and external links to web sites are shown in red.

4 Accrual

```
accrualReport(randomize(rdate) ~ site(site), data=base
     ,
              dateRange=c('1990-01-01','1994-12-31'),
              targetDate='1994-12-31', targetN=300,
              closeDate=max(base$rdate))
```

Table 1: Study Numbers

Number	Category
20	Sites
250	Subjects randomized
12.5	Subjects per site
20	Sites randomizing
12.5	Subjects randomized per randomizing site
59.4	Months from first subject randomized (1990-01-03) to 1994-12-15
1101.7	Site-months for sites randomizing
55.1	Average months since a site first randomized
0.23	Subjects randomized per site per month

[3] Lan-DeMets monitoring bounds can be plotted using the open source R `gsDesign` package.

[4] 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.

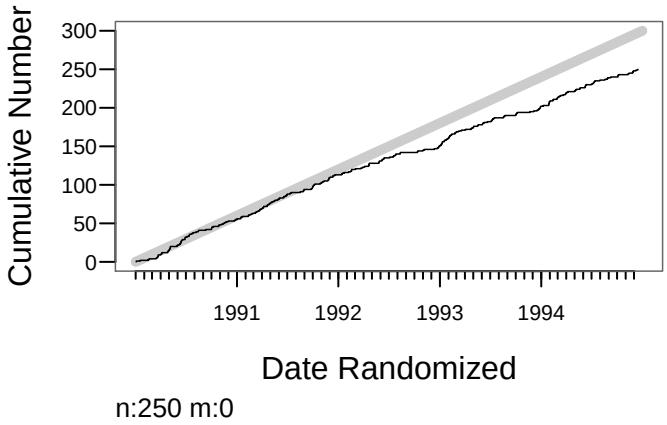


Figure 1: Subjects randomized over time. The solid back line depicts the cumulative frequency. The thick grayscale line represent targets.

Category	*N*	Used in Analysis
Enrolled	500	250
Randomized	250	250

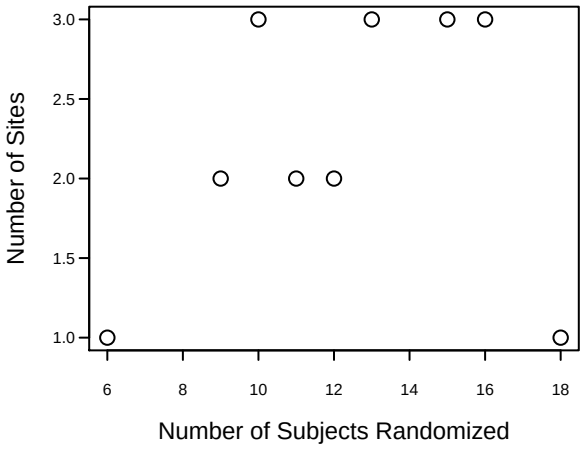


Figure 2: Number of sites having the given number of subjects randomized

Category	*N*	Used in Analysis
Enrolled	500	250
Randomized	250	250

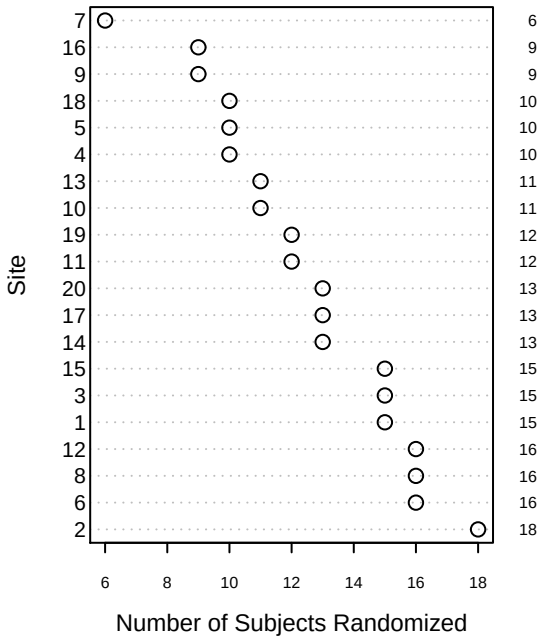


Figure 3: Counts of numbers of subjects and numbers of sites

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250

5 Baseline Variables

```
dReport(sex + race + smoking ~ trx, groups='trx', data
      =base,
         h=4, w=3.5)
```

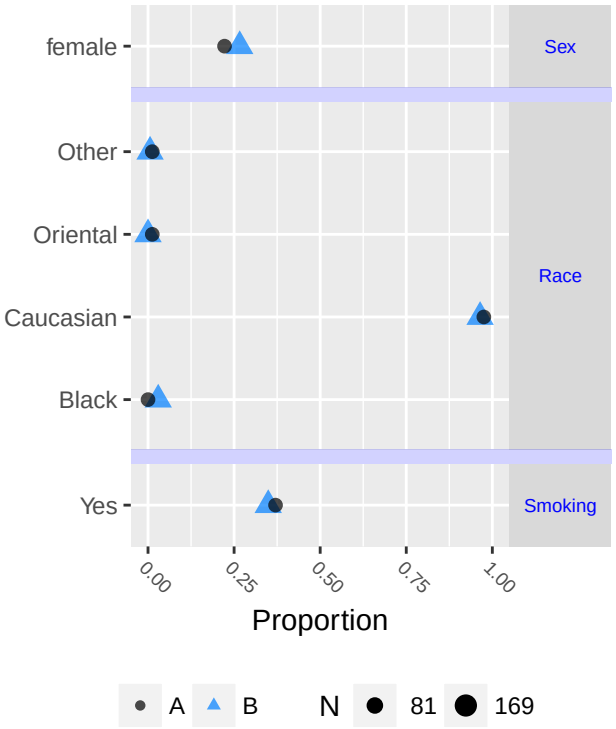


Figure 4: Proportions for sex, race, and smoking stratified by treatment. N=250 (Table 2)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
Sex	81	169
Race	81	169
Smoking	81	169

```
## Show spike histogram for raw data, 50 bins
dReport(age + height + weight + bmi + pack.yrs ~ trx,
      data=base,
         h=3.5,
         sopts=list(datadensity=TRUE,
           scat1d.opts=list(nhistSpike=1,
             col=adjustcolor('red', alpha.f=.5),
```

```
            nint=50)),
      append=TRUE)
```

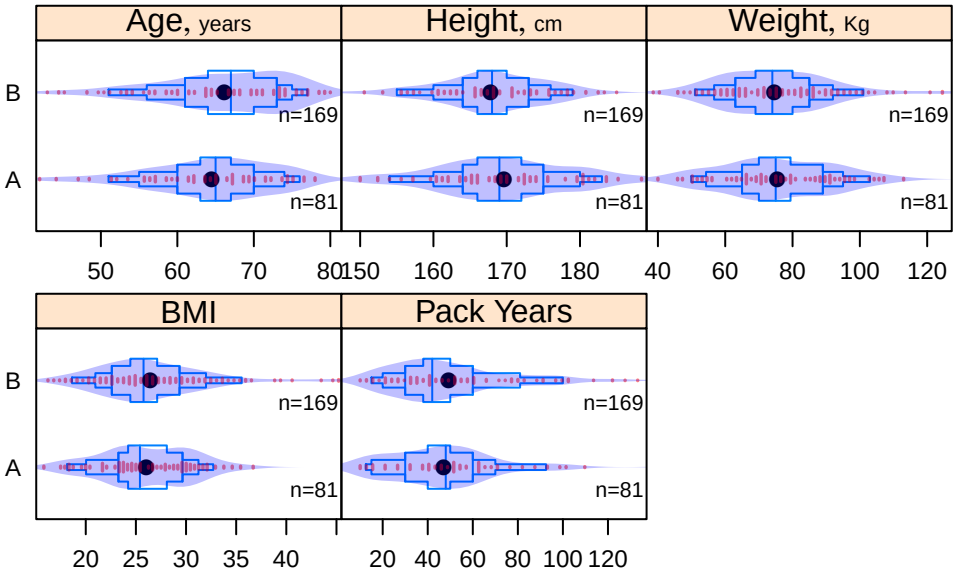


Figure 5: Extended box and violin plots for age, height, weight, BMI, and pack years stratified by treatment. N=250 (Table 3)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
Age	81	169
Height	81	169
Weight	81	169
BMI	81	169
Pack Years	81	169

6 Longitudinal Adverse Events

```
dReport(headache + ab.pain + nausea + dyspepsia +
     diarrhea +
        upper.resp.infect + coad ~ week + trx + id(id)
             ,
        groups='trx', data=ssafety, panel='longae',
             what='byx',
        popts=list(cex.strip=.57,
          key=list(x=.65, y=.2, lines=TRUE, points=
               FALSE)))
```

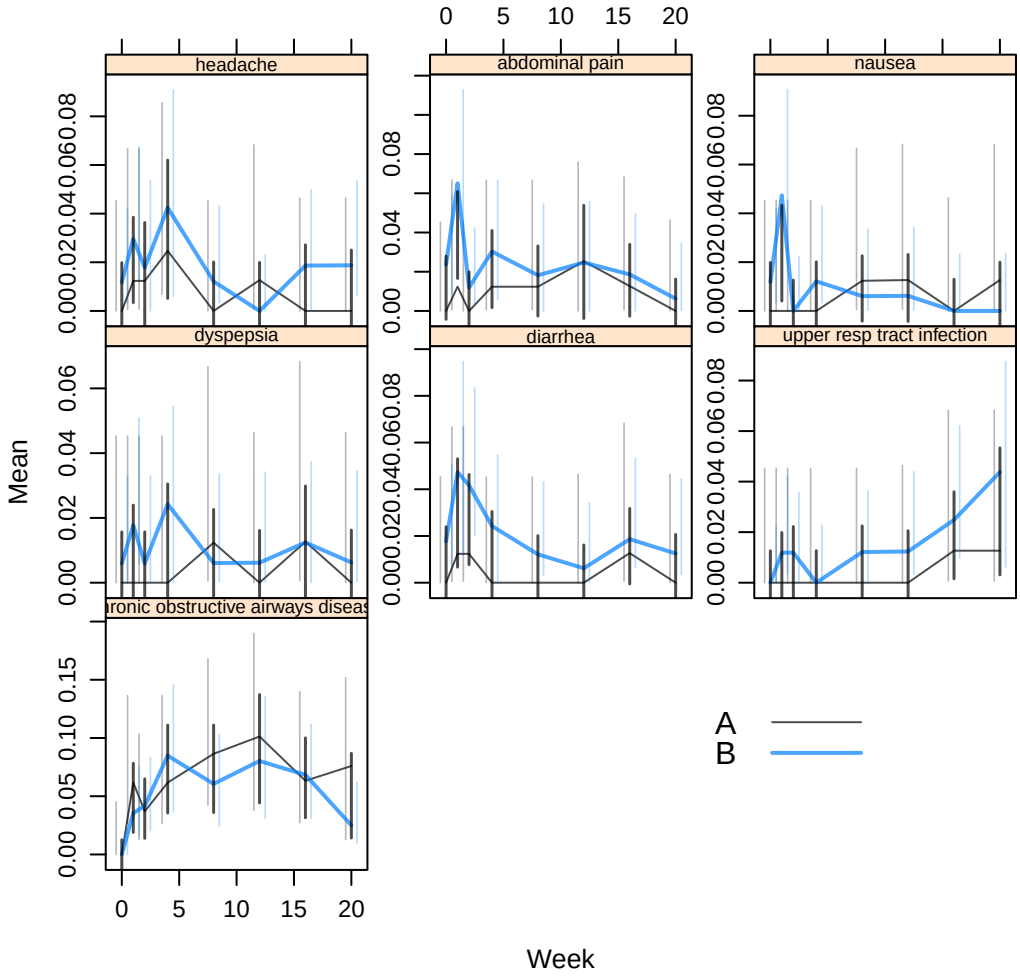


Figure 6: Means and 0.95 bootstrap percentile confidence limits for 7 variables vs. week stratified by treatment. N=250 (Table 4)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
headache	81	169
abdominal pain	81	169
nausea	81	169
dyspepsia	81	169
diarrhea	81	169
upper resp tract infection	81	169
chronic obstructive airways disease	81	169

7 Incidence of Adverse Events at Any Follow-up

```
## Reformat to one record per event per subject per
      time
aev ← vars$ae
ev  ← ssafety[ssafety$week > 0, c(aev, 'trx', 'id', '
      week')]
## Reshape to tall and thin format
evt ← reshape(ev, direction='long', idvar=c('id', '
      week'),
              varying=aev, v.names='sev', timevar='
                    event',
              times=aev)
## For each event, id and trx see if event occurred at
       any week
ne ← with(evt, summarize(sev, llist(id, trx, event),
                         function(y) any(y > 0, na.rm
                               =TRUE)))
## Remove non-occurrences of events
ne ← subset(ne, sev, select=c(id, trx, event))
## Replace event names with event labels
elab ← sapply(ssafety[aev], label)
ne$event ← elab[ne$event]
label(ne$trx) ← 'Treatment'

eReport(event ~ trx, data=ne, h=3.25)
```

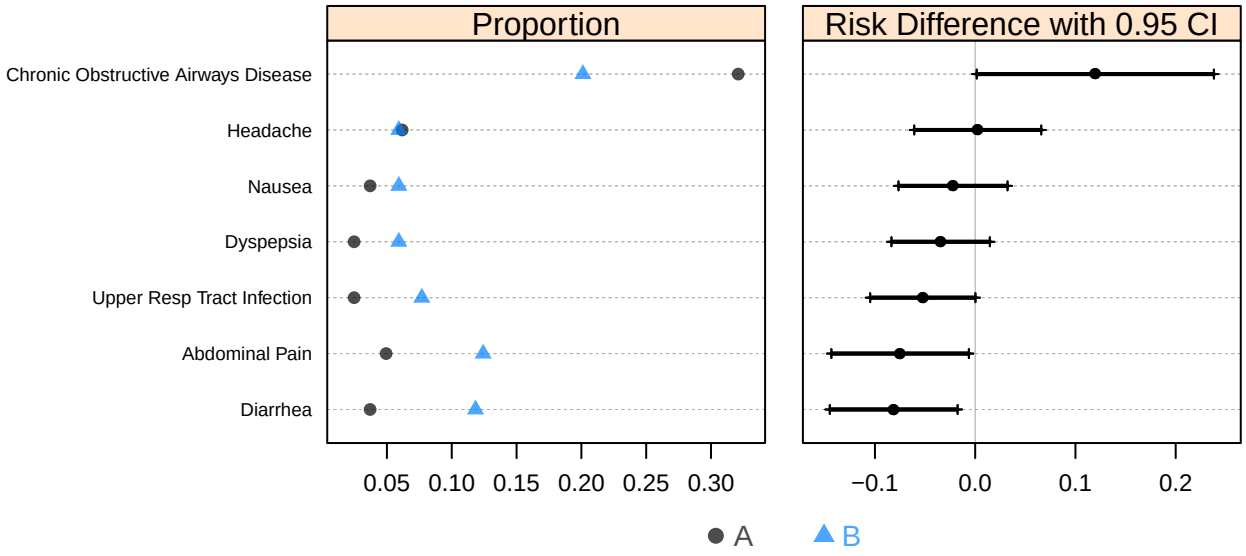


Figure 7: Proportion of adverse events and risk differences by Treatment sorted by risk difference (Table 5)

Category	*N*	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
event	45	118

8 Longitudinal EKG Data

```
dReport(axis + corr.qt + pr + qrs + uncorr.qt + hr ~
     week + trx +
        id(id),
        groups='trx', data=ssafety, panel='ekg', what=
             'byx', w=7)
```

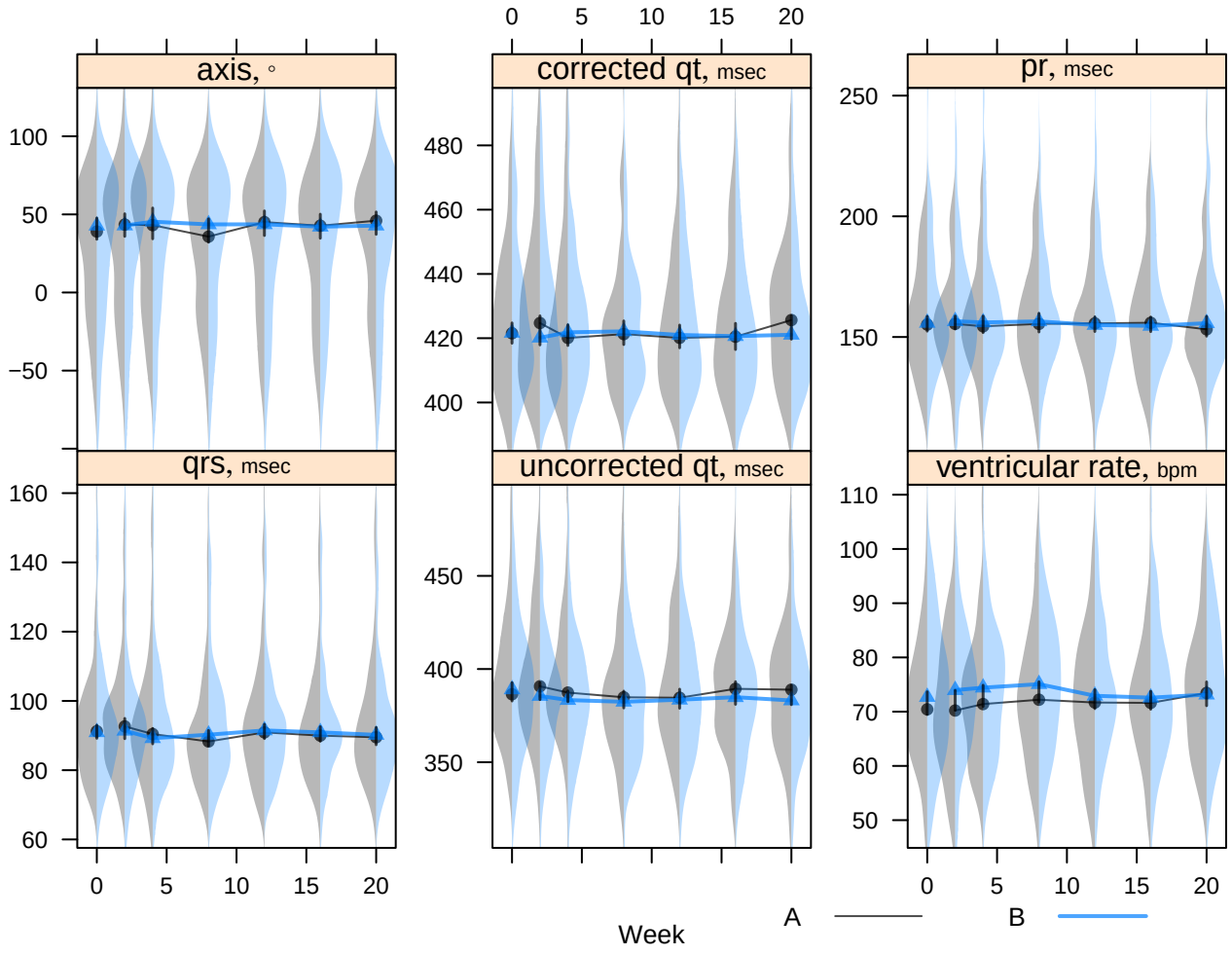


Figure 8: Medians with violin (density) plots for axis, corrected qt, pr, qrs, uncorrected qt, and ventricular rate vs. week stratified by treatment. N=248 to 250 (Table 6)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
axis	81	169
corrected qt	81	169
pr	81	167
qrs	81	169
uncorrected qt	81	169
ventricular rate	81	169

9 Longitudinal Clinical Chemistry Data

```
## Plot 6 variables per page
cvar ← split(vars$chem, rep(letters[1:4], each=6))
for(subpanel in names(cvar)) {
  form ← paste(cvar[[subpanel]], collapse=' + ')
  form ← as.formula(paste(form, 'week + trx + id(id)'
        , sep=' ~ '))
  dReport(form, groups='trx', data=ssafety, panel='
        cchem',
           subpanel=subpanel,
           what='byx', append=subpanel != 'a',
           popts=list(cex.strip=.7), w=7)
  }
```

```
## Repeat last figure using quantile intervals instead
       of violin densities
dReport(form, groups='trx', data=ssafety, panel='cchem
      ',
        subpanel='e', what='byx', byx.type='quantiles'
             , append=TRUE,
        popts=list(cex.strip=.7), w=7)
```

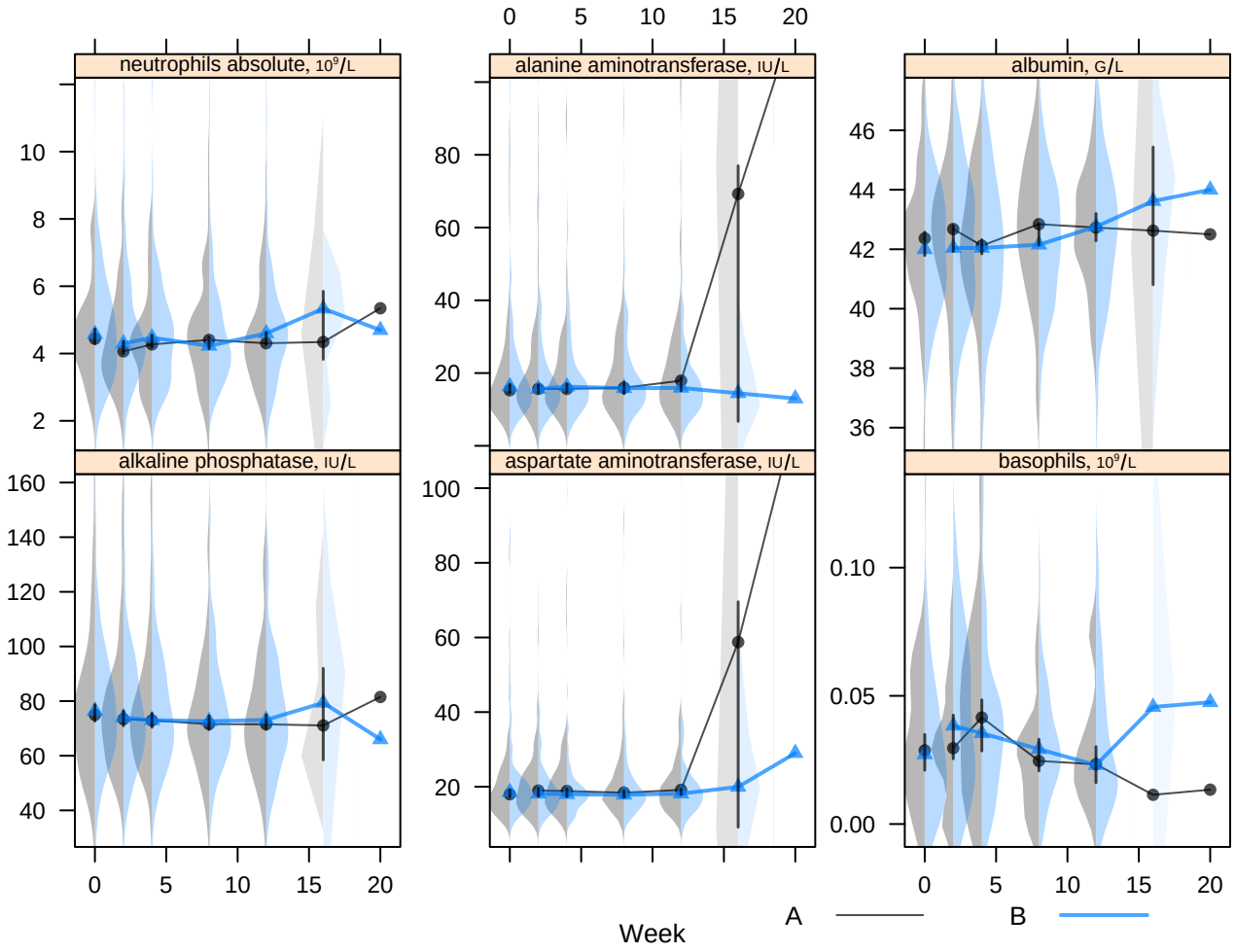


Figure 9: Medians with violin (density) plots for neutrophils absolute, alanine aminotransferase, albumin, alkaline phosphatase, aspartate aminotransferase, and basophils vs. week stratified by treatment. N=72 to 250 (Table 7)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
neutrophils absolute	81	169
alanine aminotransferase	81	169
albumin	81	169
alkaline phosphatase	81	169
aspartate aminotransferase	81	169
basophils	21	51

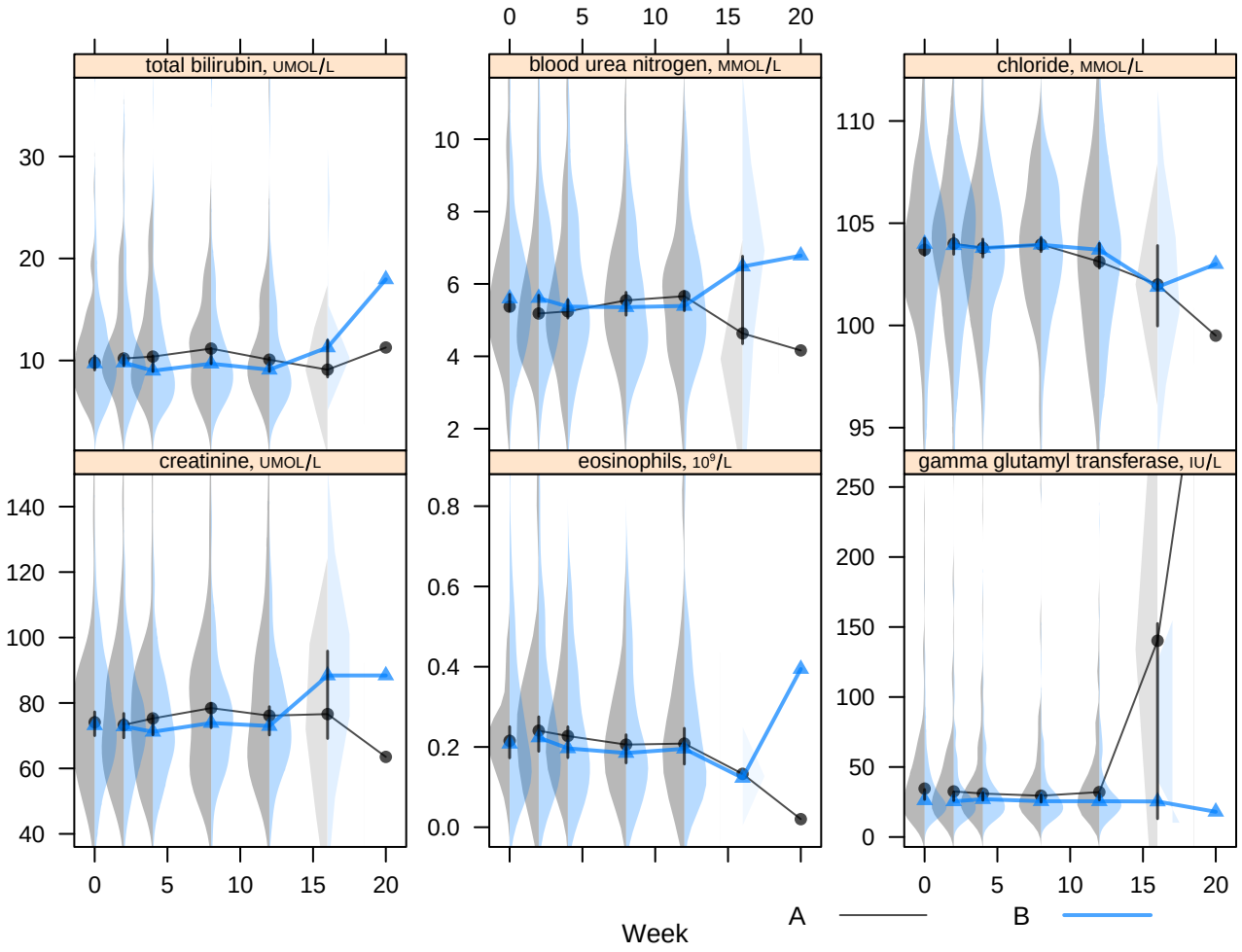


Figure 10: Medians with violin (density) plots for total bilirubin, blood urea nitrogen, chloride, creatinine, eosinophils, and gamma glutamyl transferase vs. week stratified by treatment. N=72 to 250 (Table 8)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
total bilirubin	81	169
blood urea nitrogen	81	169
chloride	81	169
creatinine	81	169
eosinophils	21	51
gamma glutamyl transferase	81	169

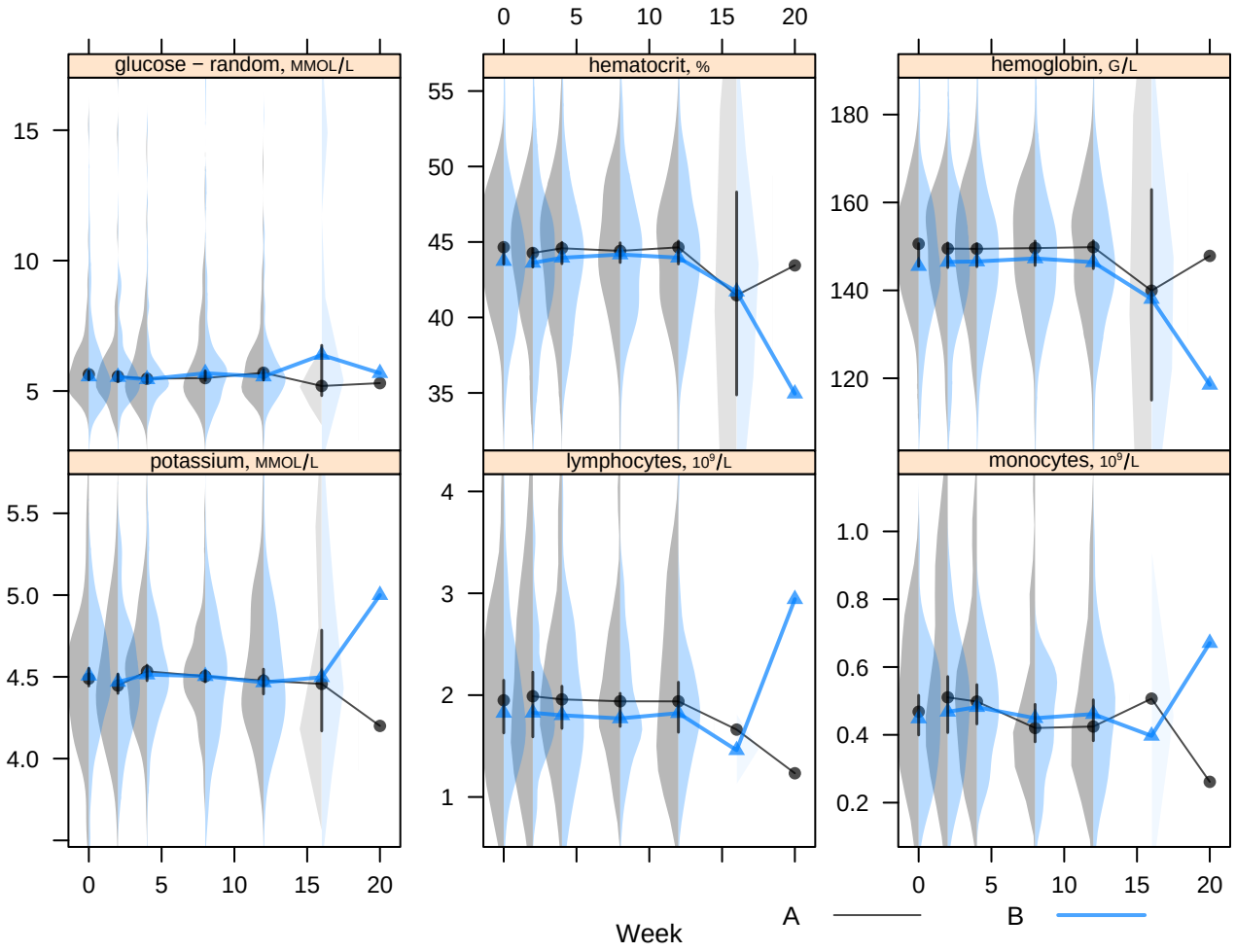


Figure 11: Medians with violin (density) plots for glucose - random, hematocrit, hemoglobin, potassium, lymphocytes, and monocytes vs. week stratified by treatment. N=72 to 250 (Table 9)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
glucose - random	81	163
hematocrit	81	169
hemoglobin	81	169
potassium	81	169
lymphocytes	21	51
monocytes	21	51

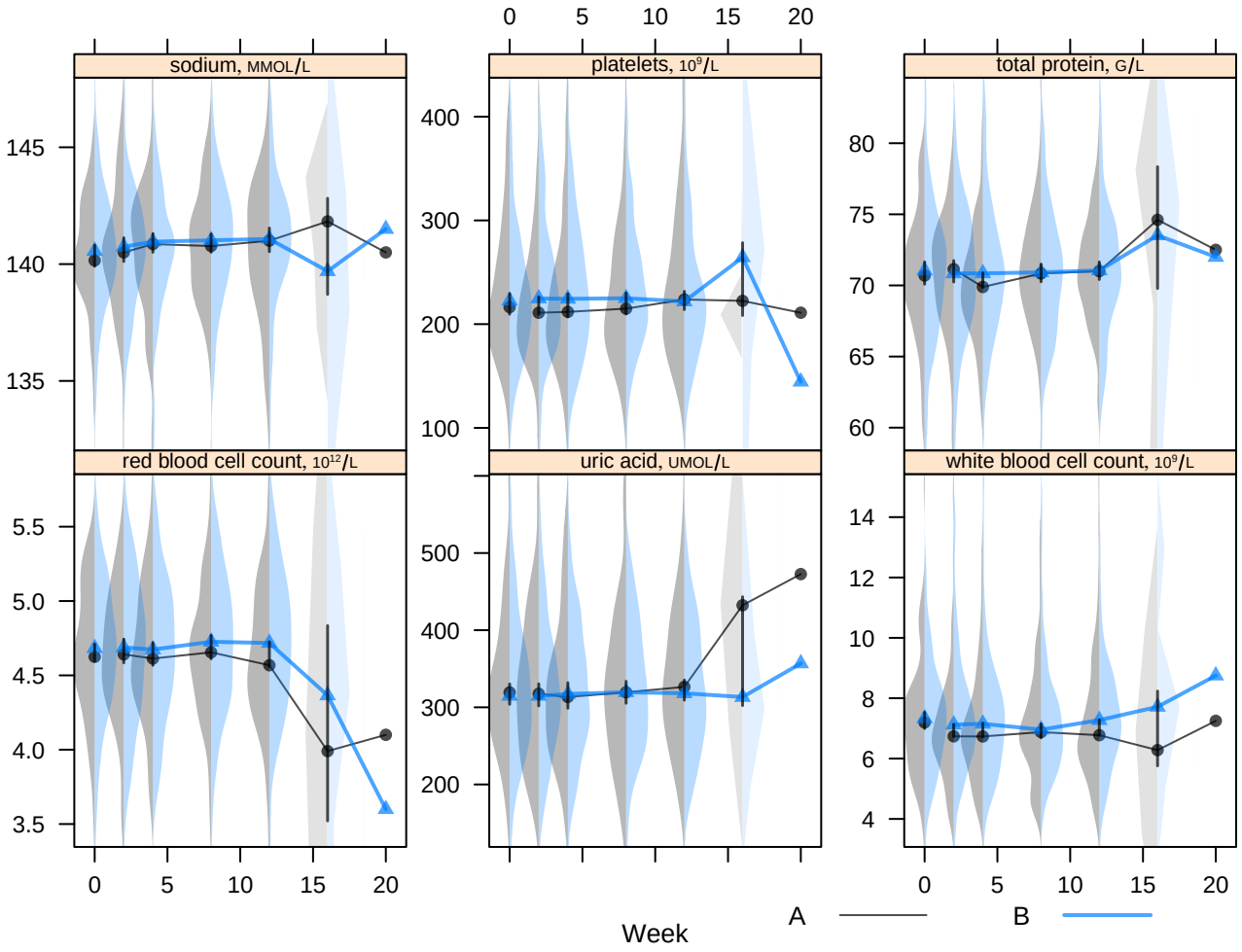


Figure 12: Medians with violin (density) plots for sodium, platelets, total protein, red blood cell count, uric acid, and white blood cell count vs. week stratified by treatment. N=250 (Table 10)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
sodium	81	169
platelets	81	169
total protein	81	169
red blood cell count	81	169
uric acid	81	169
white blood cell count	81	169

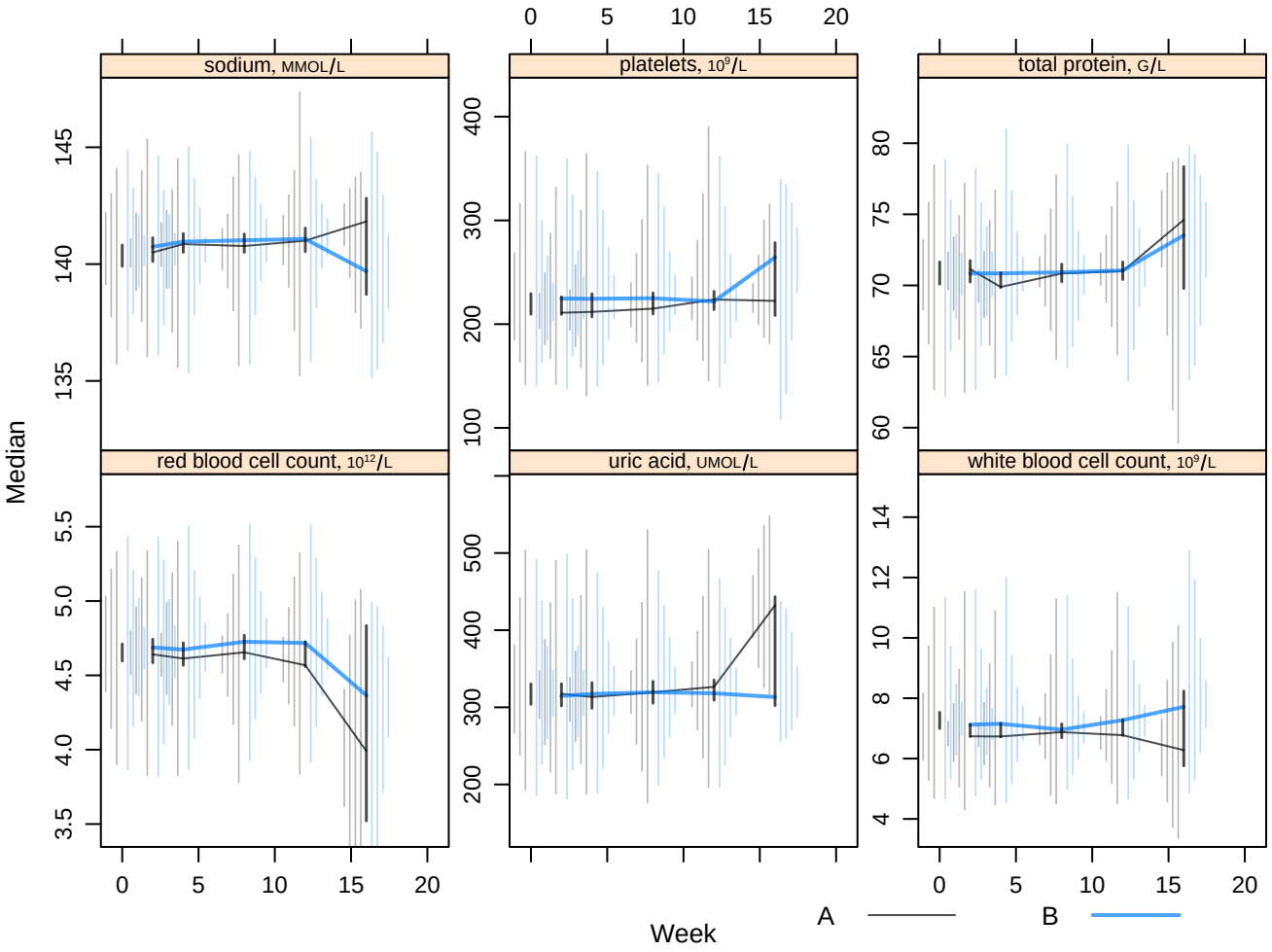


Figure 13: Medians with these quantile intervals for sodium, platelets, total protein, red blood cell count, uric acid, and white blood cell count vs. week stratified by treatment. N=250 (Table 11)

Category	N	Used in Analysis
Enrolled	500	250
Randomized	250	250
A	81	81
B	169	169

Variable	A	B
sodium	81	169
platelets	81	169
total protein	81	169
red blood cell count	81	169
uric acid	81	169
white blood cell count	81	169

10 Computing Environment

These analyses were done using the following versions of R[1], the operating system, and add-on packages `greport`[2], `Hmisc`[3], `rms`[4], and others:

- R version 3.3.3 (2017-03-06), `x86_64-pc-linux-gnu`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: data.table 1.10.4, Formula 1.2-1, ggplot2 2.2.1, greport 0.7-2, Hmisc 4.0-3, knitr 1.15.1, lattice 0.20-35, survival 2.41-3
- Loaded via a namespace (and not attached): acepack 1.4.1, backports 1.0.5, base64enc 0.1-3, checkmate 1.8.2, cluster 2.0.6, codetools 0.2-15, colorspace 1.3-2, digest 0.6.12, evaluate 0.10, foreign 0.8-67, grid 3.3.3, gridExtra 2.2.1, gtable 0.2.0, htmlTable 1.9, htmltools 0.3.5, htmlwidgets 0.8, labeling 0.3, latticeExtra 0.6-28, lazyeval 0.2.0, magrittr 1.5, MASS 7.3-45, Matrix 1.2-8, MatrixModels 0.4-1, multcomp 1.4-6, munsell 0.4.3, mvtnorm 1.0-6, nlme 3.1-131, nnet 7.3-12, plyr 1.8.4, polspline 1.1.12, quantreg 5.29, RColorBrewer 1.1-2, Rcpp 0.12.10, reshape2 1.4.2, rms 5.1-1, rpart 4.1-10, sandwich 2.3-4, scales 0.4.1, SparseM 1.76, splines 3.3.3, stringi 1.1.3, stringr 1.2.0, TH.data 1.0-8, tibble 1.3.0, tools 3.3.3, zoo 1.7-14

The reproducible research framework `knitr` [5] was used.

References

[1] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. Available from http://www.R-project.org. 27

[2] Frank E. Harrell. `greport`: R functions for graphical reporting of clinical trials. Available from biostat.mc.vanderbilt.edu/Greport, 2014. 27

[3] Frank E. Harrell. `Hmisc`: A package of miscellaneous R functions. Available from biostat.mc.vanderbilt.edu/Hmisc, 2014. 27

[4] Frank E. Harrell. `rms`: S functions for biostatistical/epidemiologic modeling, testing, estimation, validation, graphics, prediction, and typesetting by storing enhanced model design attributes in the fit. Available from biostat.mc.vanderbilt.edu/rms, 2013. Implements methods in *Regression Modeling Strategies*, New York:Springer, 2001. 27

[5] Yihui Xie. *Dynamic Documents with R and* ***knitr***. Chapman and Hall, 2013. ISBN 978-1482203530. 27

11 Appendix: Supporting Tables

Table 2: Proportions for sex, race, and smoking stratified by treatment (Figure 4).

		A	B
Sex			
	female	0.222 $\frac{18}{81}$	0.266 $\frac{45}{169}$
Race			
	Black	0.000 $\frac{0}{81}$	0.030 $\frac{5}{169}$
	Caucasian	0.975 $\frac{79}{81}$	0.964 $\frac{163}{169}$
	Oriental	0.012 $\frac{1}{81}$	0.000 $\frac{0}{169}$
	Other	0.012 $\frac{1}{81}$	0.006 $\frac{1}{169}$
Smoking			
	Yes	0.370 $\frac{30}{81}$	0.349 $\frac{59}{169}$

Table 3: Statistics for age, height, weight, BMI, and pack years stratified by treatment. $_a$ b $_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. (Figure 5).

	A $N = 81$	B $N = 169$
Age years	60.0 **65.0** 70.0 (64.5 ± 8.1)	61.0 **67.0** 73.0 (66.1 ± 8.4)
Height cm	164.0 **169.0** 175.0 (169.6 ± 9.1)	164.0 **168.0** 173.0 (167.8 ± 7.4)
Weight Kg	65.0 **75.0** 89.0 (75.4 ±16.7)	63.0 **74.0** 85.0 (74.6 ±16.5)
BMI	23.3 **25.4** 29.6 (26.0 ± 4.6)	22.6 **25.8** 29.4 (26.4 ± 5.4)
Pack Years	30.0 **48.0** 60.0 (47.0 ±25.3)	30.0 **42.0** 60.0 (49.2 ±27.1)

Table 4: Means and 0.95 bootstrap CLs for 7 variables vs. week stratified by treatment (Figure 6).

		A				B			
	Week	n	Mean	Lower	Upper	n	Mean	Lower	Upper
Headache									
	0	81	0.00	0.00	0.05	169	0.01	0.00	0.04
	1	81	0.01	0.00	0.07	169	0.03	0.01	0.07
	2	81	0.01	0.00	0.07	168	0.02	0.00	0.05
	4	81	0.02	0.01	0.09	165	0.04	0.01	0.09
	8	81	0.00	0.00	0.05	165	0.01	0.00	0.04
	12	79	0.01	0.00	0.07	162	0.00	0.00	0.02
	16	79	0.00	0.00	0.05	161	0.02	0.00	0.05
	20	79	0.00	0.00	0.05	160	0.02	0.01	0.05
Abdominal Pain									
	0	81	0.00	0.00	0.05	169	0.02	0.01	0.06
	1	81	0.01	0.00	0.07	169	0.07	0.04	0.11
	2	81	0.00	0.00	0.05	168	0.01	0.00	0.04
	4	81	0.01	0.00	0.07	165	0.03	0.01	0.07
	8	81	0.01	0.00	0.07	165	0.02	0.00	0.05
	12	79	0.03	0.00	0.08	162	0.02	0.00	0.06
	16	79	0.01	0.00	0.07	161	0.02	0.00	0.05
	20	79	0.00	0.00	0.05	160	0.01	0.00	0.03
Nausea									
	0	81	0.00	0.00	0.05	169	0.01	0.00	0.04
	1	81	0.00	0.00	0.05	169	0.05	0.02	0.09
	2	81	0.00	0.00	0.05	168	0.00	0.00	0.02
	4	81	0.00	0.00	0.05	165	0.01	0.00	0.04
	8	81	0.01	0.00	0.07	165	0.01	0.00	0.03
	12	79	0.01	0.00	0.07	162	0.01	0.00	0.03
	16	79	0.00	0.00	0.05	161	0.00	0.00	0.02
	20	79	0.01	0.00	0.07	160	0.00	0.00	0.02
Dyspepsia									
	0	81	0.00	0.00	0.05	169	0.01	0.00	0.03
	1	81	0.00	0.00	0.05	169	0.02	0.01	0.05
	2	81	0.00	0.00	0.05	168	0.01	0.00	0.03
	4	81	0.00	0.00	0.05	165	0.02	0.00	0.05
	8	81	0.01	0.00	0.07	165	0.01	0.00	0.03
	12	79	0.00	0.00	0.05	162	0.01	0.00	0.03
	16	79	0.01	0.00	0.07	161	0.01	0.00	0.04
	20	79	0.00	0.00	0.05	160	0.01	0.00	0.03
Diarrhea									
	0	81	0.00	0.00	0.05	169	0.02	0.01	0.05
	1	81	0.01	0.00	0.07	169	0.05	0.01	0.09
	2	81	0.01	0.00	0.07	168	0.04	0.02	0.08
	4	81	0.00	0.00	0.05	165	0.02	0.00	0.05
	8	81	0.00	0.00	0.05	165	0.01	0.00	0.04
	12	79	0.00	0.00	0.05	162	0.01	0.00	0.03
	16	79	0.01	0.00	0.07	161	0.02	0.01	0.05
	20	79	0.00	0.00	0.05	160	0.01	0.00	0.04
Upper Resp Tract Infection									
	0	81	0.00	0.00	0.05	169	0.00	0.00	0.02
	1	81	0.00	0.00	0.05	169	0.01	0.00	0.04
	2	81	0.00	0.00	0.05	168	0.01	0.00	0.04
	4	81	0.00	0.00	0.05	165	0.00	0.00	0.02
	8	81	0.00	0.00	0.05	165	0.01	0.00	0.04
	12	79	0.00	0.00	0.05	162	0.01	0.00	0.04
	16	79	0.01	0.00	0.07	161	0.02	0.01	0.06
	20	79	0.01	0.00	0.07	160	0.04	0.01	0.09
Chronic Obstructive Airways Disease									
	0	81	0.00	0.00	0.05	169	0.00	0.00	0.02
	1	81	0.06	0.03	0.14	169	0.04	0.02	0.08
	2	81	0.04	0.01	0.10	168	0.04	0.02	0.08
	4	81	0.06	0.03	0.14	165	0.08	0.04	0.15
	8	81	0.09	0.04	0.17	165	0.06	0.02	0.10
	12	79	0.10	0.04	0.19	162	0.08	0.03	0.14
	16	79	0.06	0.03	0.14	161	0.07	0.03	0.11
	20	79	0.08	0.01	0.15	160	0.02	0.01	0.06

Table 5: Proportion of adverse events and risk differences by Treatment (Figure 7).

Event	A	B	Difference	Lower	Upper
Abdominal Pain	0.049	0.124	-0.075	-0.143	-0.006
Chronic Obstructive Airways Disease	0.321	0.201	0.120	0.002	0.238
Diarrhea	0.037	0.118	-0.081	-0.145	-0.018
Dyspepsia	0.025	0.059	-0.034	-0.084	0.015
Headache	0.062	0.059	0.003	-0.061	0.066
Nausea	0.037	0.059	-0.022	-0.077	0.032
Upper Resp Tract Infection	0.025	0.077	-0.052	-0.105	0.000

Table 6: Medians for axis, corrected qt, pr, qrs, uncorrected qt, and ventricular rate vs. week stratified by treatment (Figure 8).

			A				B		
		Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Axis	**degree**									
		0	79	39.0	−4.8	63.0	164	42.8	0.2	69.6
		1	0				0			
		2	78	43.5	−2.1	64.2	154	42.9	0.4	67.8
		4	78	43.1	−1.5	66.3	152	45.3	0.8	69.2
		8	74	35.8	−4.6	65.6	139	43.6	−4.1	70.3
		12	74	45.1	1.3	65.8	134	43.6	−1.0	71.5
		16	72	42.7	2.5	64.6	131	42.1	−2.7	72.0
		20	69	46.0	−0.1	62.2	129	42.8	2.1	71.0
Corrected Qt	**msec**									
		0	79	421.5	408.0	439.4	164	421.7	410.8	438.9
		1	0				0			
		2	78	424.7	408.4	440.8	154	420.2	407.5	437.6
		4	78	420.1	408.6	438.9	152	421.8	409.3	435.6
		8	74	421.3	409.6	432.0	139	422.1	409.0	436.2
		12	74	420.1	407.5	436.1	134	421.0	408.4	438.6
		16	72	420.5	408.5	436.9	131	420.6	408.1	436.5
		20	69	425.7	410.9	438.1	129	421.1	409.9	436.5
Pr	**msec**									
		0	79	155.1	141.9	173.2	160	155.8	145.0	175.9
		1	0				0			
		2	78	155.4	144.2	167.8	150	156.5	141.8	172.7
		4	77	154.4	143.1	168.8	149	156.0	144.2	172.7
		8	73	155.4	140.8	170.9	135	156.5	145.5	173.7
		12	73	155.6	143.5	169.0	132	155.0	141.8	172.6
		16	71	155.9	140.6	171.6	130	154.6	144.8	172.4
		20	68	153.2	142.8	169.5	127	155.7	144.4	171.2
Qrs	**msec**									
		0	79	91.2	81.8	99.5	164	90.9	81.7	100.7
		1	0				0			
		2	78	92.6	83.1	100.3	154	91.4	81.5	99.6
		4	78	90.4	81.5	100.1	152	89.2	80.5	101.1
		8	74	88.3	80.3	97.9	139	90.2	81.6	101.5
		12	74	90.9	81.7	100.8	134	91.5	83.0	100.1
		16	72	90.0	81.2	99.9	131	90.9	84.0	99.5
		20	69	89.5	81.6	100.6	129	90.2	82.0	100.9
Uncorrected Qt	**msec**									
		0	79	386.5	370.5	412.6	164	389.1	367.3	410.4
		1	0				0			
		2	78	390.7	372.2	418.4	154	385.6	363.4	404.1
		4	78	387.4	372.0	410.0	152	383.4	362.5	404.7
		8	74	384.8	365.1	405.5	139	382.4	359.2	402.4
		12	74	384.6	367.6	413.9	134	383.5	361.4	407.9
		16	72	389.4	367.2	408.3	131	384.9	364.8	403.5
		20	69	388.9	364.4	407.8	129	383.1	362.9	403.4
Ventricular Rate	**bpm**									
		0	79	70.4	63.5	78.0	164	72.7	63.9	81.7
		1	0				0			
		2	78	70.2	62.8	80.1	154	74.0	65.8	81.6
		4	78	71.4	62.9	77.8	152	74.4	64.9	82.1
		8	74	72.2	64.1	80.3	139	75.1	65.9	83.8
		12	74	71.7	63.2	80.3	134	72.9	66.1	82.5
		16	72	71.6	64.8	78.2	131	72.6	65.3	81.1
		20	69	73.4	64.4	80.6	129	73.1	65.9	83.6

Table 7: Medians for neutrophils absolute, alanine aminotransferase, albumin, alkaline phosphatase, aspartate aminotransferase, and basophils vs. week stratified by treatment (Figure 9).

			A				B			
		Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Neutrophils Absolute	10^9/L									
		0	76	4.4397	3.7258	5.2161	168	4.5789	3.5287	5.7013
		1	0				0			
		2	78	4.0648	3.3825	5.1414	151	4.2988	3.6155	5.4803
		4	78	4.2710	3.3565	4.9506	148	4.4709	3.3806	5.4140
		8	74	4.4056	3.5020	5.2038	136	4.2296	3.5533	5.3471
		12	72	4.3054	3.5380	5.3420	133	4.5964	3.7107	5.4461
		16	5	4.3430	2.4731	6.4518	6	5.3344	4.1894	5.9051
		20	2	5.3450			2	4.6950		
Alanine Aminotransferase	IU/L									
		0	79	15.2663	11.1730	22.5401	167	16.3126	11.7193	22.8366
		1	0				0			
		2	75	15.6605	12.1158	24.9389	153	15.6409	11.6878	21.8960
		4	77	15.6518	12.0340	23.5054	148	16.2617	11.9288	22.7533
		8	71	15.9708	12.3113	26.6913	138	15.8491	11.7525	21.5796
		12	73	17.9051	12.2189	26.9154	132	15.9316	11.5314	21.8103
		16	5	69.2640	28.5319	116.4191	6	14.4518	9.2086	21.4861
		20	2	118.0000			2	13.0000		
Albumin	G/L									
		0	79	42.3666	41.0279	43.4679	168	42.0000	40.6297	43.5820
		1	0				0			
		2	76	42.6779	41.2642	43.9077	152	42.0449	40.1725	43.9503
		4	76	42.1190	40.9380	43.7581	148	42.0427	40.3419	43.8788
		8	72	42.8440	41.3587	44.2927	138	42.1524	40.6837	43.9263
		12	73	42.7276	41.6121	43.9923	132	42.7576	40.9668	44.0354
		16	5	42.6246	38.8248	45.8731	6	43.6193	41.4765	45.0456
		20	2	42.5000			2	44.0000		
Alkaline Phosphatase	IU/L									
		0	79	74.9537	59.9445	89.2797	167	76.4320	63.1867	91.2775
		1	0				0			
		2	75	73.4181	60.7393	86.4430	153	73.9787	62.1072	90.8339
		4	77	72.9318	61.6680	85.9529	148	73.0075	61.4313	89.0862
		8	71	71.5796	61.6872	84.6627	138	72.6242	62.0784	87.5008
		12	73	71.5091	60.1041	88.0216	132	73.1198	61.9448	88.5549
		16	5	71.0858	60.9128	94.5131	6	79.3870	53.9745	97.8150
		20	2	81.5000			2	66.0000		
Aspartate Aminotransferase	IU/L									
		0	79	17.9750	14.9040	23.3245	167	18.5310	15.2728	21.9740
		1	0				0			
		2	75	18.9557	15.1797	23.6707	153	18.1252	15.0217	22.0339
		4	77	18.8499	15.1357	23.7954	148	18.0004	15.5743	21.0850
		8	71	18.3871	16.0637	23.9242	138	17.8465	15.3363	21.5931
		12	73	19.1500	15.2237	24.1486	132	18.1470	15.1467	21.8040
		16	5	58.7536	27.4556	124.3370	6	19.9588	14.0015	32.3066
		20	2	119.0000			2	29.0000		
Basophils	10^9/L									
		0	20	0.0287	0.0143	0.0520	50	0.0272	0.0121	0.0482
		1	0				0			
		2	21	0.0296	0.0145	0.0447	43	0.0384	0.0180	0.0645
		4	20	0.0415	0.0162	0.0627	42	0.0354	0.0182	0.0512
		8	20	0.0246	0.0090	0.0333	39	0.0292	0.0189	0.0445
		12	20	0.0233	0.0139	0.0393	36	0.0230	0.0146	0.0498
		16	2	0.0114			3	0.0457	0.0155	0.0747
		20	1	0.0134			2	0.0475		

Table 8: Medians for total bilirubin, blood urea nitrogen, chloride, creatinine, eosinophils, and gamma glutamyl transferase vs. week stratified by treatment (Figure 10).

			A				B			
		Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Total Bilirubin	UMOL/L									
		0	79	9.792	7.676	12.550	167	9.716	7.344	13.042
		1	0				0			
		2	75	10.199	8.252	13.992	153	9.802	7.606	13.063
		4	77	10.384	7.871	14.165	148	8.997	6.902	12.483
		8	71	11.172	7.781	13.836	138	9.688	7.304	12.682
		12	72	10.077	8.044	14.201	132	9.100	6.994	13.514
		16	5	9.103	6.896	11.425	6	11.283	9.645	17.402
		20	2	11.275			2	17.955		
Blood Urea Nitrogen	MMOL/L									
		0	79	5.373	4.463	6.557	168	5.598	4.554	6.689
		1	0				0			
		2	76	5.183	4.240	6.370	153	5.611	4.585	7.028
		4	77	5.244	4.259	6.538	148	5.374	4.451	7.070
		8	72	5.546	4.552	6.720	138	5.359	4.287	6.735
		12	73	5.661	4.312	6.637	132	5.392	4.393	6.643
		16	5	4.635	3.045	8.212	6	6.481	4.640	7.752
		20	2	4.164			2	6.783		
Chloride	MMOL/L									
		0	79	103.703	101.662	106.055	168	103.988	101.554	105.898
		1	0				0			
		2	76	104.000	101.601	105.863	154	103.924	101.362	105.999
		4	77	103.786	101.234	105.628	149	103.783	102.002	105.829
		8	72	103.959	101.787	106.025	138	103.946	101.878	106.122
		12	74	103.118	100.012	105.862	132	103.700	101.161	105.842
		16	5	102.000	100.321	103.679	6	101.881	99.894	104.583
		20	2	99.500			2	103.000		
Creatinine	UMOL/L									
		0	79	74.096	61.687	87.565	167	73.195	62.716	87.166
		1	0				0			
		2	75	73.286	61.654	88.635	153	72.859	63.290	83.767
		4	77	75.269	62.579	88.522	147	71.196	62.415	82.803
		8	71	78.433	61.095	91.309	138	73.887	62.898	83.488
		12	73	76.119	62.224	93.825	132	72.965	63.973	85.037
		16	5	76.603	61.436	87.817	6	88.392	73.014	103.825
		20	2	63.520			2	88.400		
Eosinophils	10^9/L									
		0	20	0.215	0.160	0.273	50	0.208	0.124	0.355
		1	0				0			
		2	21	0.241	0.134	0.338	43	0.224	0.133	0.369
		4	20	0.227	0.150	0.330	42	0.196	0.109	0.345
		8	20	0.206	0.128	0.300	39	0.185	0.111	0.333
		12	20	0.208	0.121	0.327	36	0.195	0.107	0.329
		16	2	0.133			3	0.123	0.093	0.156
		20	1	0.020			2	0.394		
Gamma Glutamyl Transferase	IU/L									
		0	79	34.536	20.152	54.819	167	26.235	19.790	41.753
		1	0				0			
		2	75	32.551	20.586	52.182	153	25.490	18.532	39.833
		4	77	31.033	20.476	50.683	147	27.024	18.978	40.208
		8	71	29.489	20.902	47.519	138	25.587	18.597	38.188
		12	73	32.086	20.960	50.753	132	25.558	19.561	38.657
		16	5	140.185	67.120	388.722	6	25.369	20.096	67.396
		20	2	424.500			2	18.000		

Table 9: Medians for glucose - random, hematocrit, hemoglobin, potassium, lymphocytes, and monocytes vs. week stratified by treatment (Figure 11).

		A				B			
	Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Glucose - Random MMOL/L									
	0	77	5.637	5.062	6.406	157	5.556	4.865	6.632
	1	0				0			
	2	72	5.559	5.003	6.327	143	5.537	5.013	6.389
	4	74	5.472	4.856	6.543	139	5.449	5.065	6.084
	8	69	5.494	4.959	6.789	131	5.682	4.868	6.397
	12	70	5.700	5.034	6.464	124	5.555	5.037	6.305
	16	5	5.193	4.758	5.587	6	6.382	5.287	9.892
	20	2	5.302			2	5.684		
Hematocrit %									
	0	76	44.653	41.895	47.444	169	43.740	40.892	46.225
	1	0				0			
	2	78	44.270	41.427	46.669	152	43.616	40.875	46.447
	4	78	44.583	41.157	47.234	149	43.954	40.911	46.486
	8	74	44.411	41.906	47.674	136	44.170	40.883	46.773
	12	72	44.644	41.407	47.100	133	43.960	40.988	46.787
	16	5	41.469	33.937	47.725	6	41.712	36.044	45.719
	20	2	43.450			2	34.950		
Hemoglobin G/L									
	0	76	150.586	141.716	157.914	169	145.524	137.151	155.593
	1	0				0			
	2	78	149.503	140.323	157.872	152	146.475	137.239	155.986
	4	78	149.449	139.587	158.858	149	146.569	137.104	155.576
	8	74	149.616	140.160	158.679	136	147.265	137.738	155.998
	12	72	149.834	139.816	157.784	133	146.391	137.923	156.531
	16	5	139.921	113.529	160.527	6	138.046	119.689	154.016
	20	2	147.842			2	118.500		
Potassium MMOL/L									
	0	79	4.489	4.222	4.722	167	4.507	4.295	4.732
	1	0				0			
	2	75	4.447	4.231	4.765	154	4.469	4.252	4.709
	4	77	4.534	4.291	4.727	149	4.514	4.365	4.755
	8	71	4.503	4.365	4.663	138	4.504	4.220	4.728
	12	73	4.477	4.281	4.831	132	4.466	4.232	4.762
	16	5	4.457	4.167	4.985	6	4.498	4.248	4.956
	20	2	4.200			2	5.000		
Lymphocytes 10^9/L									
	0	20	1.949	1.394	2.588	50	1.826	1.486	2.313
	1	0				0			
	2	21	1.989	1.452	2.758	43	1.827	1.549	2.177
	4	20	1.959	1.570	2.495	42	1.802	1.368	2.356
	8	20	1.940	1.610	2.428	39	1.771	1.542	2.276
	12	20	1.939	1.403	2.660	36	1.823	1.503	2.197
	16	2	1.662			3	1.459	1.377	1.551
	20	1	1.233			2	2.942		
Monocytes 10^9/L									
	0	20	0.468	0.320	0.648	50	0.448	0.318	0.583
	1	0				0			
	2	21	0.511	0.368	0.808	43	0.468	0.323	0.593
	4	20	0.498	0.341	0.633	42	0.482	0.371	0.591
	8	20	0.420	0.332	0.655	39	0.449	0.338	0.560
	12	20	0.424	0.323	0.629	36	0.461	0.320	0.567
	16	2	0.507			3	0.397	0.302	0.547
	20	1	0.261			2	0.671		

Table 10: Medians for sodium, platelets, total protein, red blood cell count, uric acid, and white blood cell count vs. week stratified by treatment (Figure 12).

			A				B		
	Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Sodium MMOL/L									
	0	79	140.16	139.13	142.21	168	140.56	139.02	142.10
	1	0				0			
	2	76	140.50	138.88	142.19	154	140.74	138.97	142.11
	4	77	140.86	138.98	142.28	149	140.96	139.18	142.40
	8	72	140.77	138.99	142.14	138	141.02	139.29	142.55
	12	74	141.00	139.00	142.96	132	141.08	139.84	142.59
	16	5	141.83	139.42	143.22	6	139.69	136.63	142.94
	20	2	140.50			2	141.50		
Platelets 10^9/L									
	0	76	216.38	185.08	268.75	169	222.62	184.97	265.39
	1	0				0			
	2	78	210.98	180.22	249.59	152	224.74	189.11	270.61
	4	78	211.83	178.49	256.85	149	224.43	184.85	273.90
	8	74	214.92	182.32	267.60	136	224.96	193.15	269.28
	12	72	223.77	184.65	280.55	133	221.90	187.20	266.65
	16	5	222.44	199.91	266.78	6	264.39	184.93	317.07
	20	2	211.00			2	144.50		
Total Protein G/L									
	0	79	70.70	68.29	73.22	168	71.05	67.67	73.60
	1	0				0			
	2	76	71.14	68.26	73.40	153	70.85	67.89	74.11
	4	77	69.89	67.75	72.37	148	70.85	67.90	73.78
	8	72	70.84	68.55	73.59	138	70.92	67.54	73.81
	12	73	71.01	68.82	73.51	132	71.05	68.51	73.96
	16	5	74.61	66.50	77.92	6	73.52	67.23	77.72
	20	2	72.50			2	72.00		
Red Blood Cell Count 10^{12}/L									
	0	76	4.62	4.39	5.03	169	4.68	4.37	5.02
	1	0				0			
	2	78	4.64	4.38	4.96	152	4.69	4.30	5.01
	4	78	4.61	4.37	4.99	149	4.67	4.34	5.03
	8	74	4.65	4.36	4.91	136	4.73	4.38	5.06
	12	72	4.57	4.31	4.96	133	4.72	4.36	5.06
	16	5	3.99	3.34	4.77	6	4.37	3.72	4.83
	20	2	4.10			2	3.60		
Uric Acid UMOL/L									
	0	79	319.30	266.46	381.00	168	314.90	259.07	378.68
	1	0				0			
	2	76	317.51	251.41	388.04	153	314.83	261.81	379.32
	4	77	313.37	255.82	372.81	148	317.28	265.84	385.43
	8	72	319.46	252.32	388.06	138	319.54	261.51	390.48
	12	74	326.50	270.96	398.38	132	318.21	261.73	388.66
	16	5	432.30	351.61	505.48	6	313.34	271.47	396.52
	20	2	472.66			2	356.88		
White Blood Cell Count 10^9/L									
	0	76	7.20	5.96	8.16	169	7.33	6.13	8.47
	1	0				0			
	2	78	6.74	5.91	7.82	152	7.12	6.13	8.34
	4	78	6.73	5.80	7.85	149	7.15	5.90	8.34
	8	74	6.88	5.99	8.15	136	6.96	6.00	8.07
	12	72	6.77	5.93	8.28	133	7.27	6.23	8.47
	16	5	6.28	4.57	8.59	6	7.72	6.21	9.98
	20	2	7.25			2	8.75		

Table 11: Medians for sodium, platelets, total protein, red blood cell count, uric acid, and white blood cell count vs. week stratified by treatment (Figure 13).

			A				B		
	Week	n	Median	Q1	Q3	n	Median	Q1	Q3
Sodium MMOL/L									
	0	79	140.16	139.13	142.21	168	140.56	139.02	142.10
	1	0				0			
	2	76	140.50	138.88	142.19	154	140.74	138.97	142.11
	4	77	140.86	138.98	142.28	149	140.96	139.18	142.40
	8	72	140.77	138.99	142.14	138	141.02	139.29	142.55
	12	74	141.00	139.00	142.96	132	141.08	139.84	142.59
	16	5	141.83	139.42	143.22	6	139.69	136.63	142.94
	20	2	140.50			2	141.50		
Platelets 10^9/L									
	0	76	216.38	185.08	268.75	169	222.62	184.97	265.39
	1	0				0			
	2	78	210.98	180.22	249.59	152	224.74	189.11	270.61
	4	78	211.83	178.49	256.85	149	224.43	184.85	273.90
	8	74	214.92	182.32	267.60	136	224.96	193.15	269.28
	12	72	223.77	184.65	280.55	133	221.90	187.20	266.65
	16	5	222.44	199.91	266.78	6	264.39	184.93	317.07
	20	2	211.00			2	144.50		
Total Protein G/L									
	0	79	70.70	68.29	73.22	168	71.05	67.67	73.60
	1	0				0			
	2	76	71.14	68.26	73.40	153	70.85	67.89	74.11
	4	77	69.89	67.75	72.37	148	70.85	67.90	73.78
	8	72	70.84	68.55	73.59	138	70.92	67.54	73.81
	12	73	71.01	68.82	73.51	132	71.05	68.51	73.96
	16	5	74.61	66.50	77.92	6	73.52	67.23	77.72
	20	2	72.50			2	72.00		
Red Blood Cell Count 10^{12}/L									
	0	76	4.62	4.39	5.03	169	4.68	4.37	5.02
	1	0				0			
	2	78	4.64	4.38	4.96	152	4.69	4.30	5.01
	4	78	4.61	4.37	4.99	149	4.67	4.34	5.03
	8	74	4.65	4.36	4.91	136	4.73	4.38	5.06
	12	72	4.57	4.31	4.96	133	4.72	4.36	5.06
	16	5	3.99	3.34	4.77	6	4.37	3.72	4.83
	20	2	4.10			2	3.60		
Uric Acid UMOL/L									
	0	79	319.30	266.46	381.00	168	314.90	259.07	378.68
	1	0				0			
	2	76	317.51	251.41	388.04	153	314.83	261.81	379.32
	4	77	313.37	255.82	372.81	148	317.28	265.84	385.43
	8	72	319.46	252.32	388.06	138	319.54	261.51	390.48
	12	74	326.50	270.96	398.38	132	318.21	261.73	388.66
	16	5	432.30	351.61	505.48	6	313.34	271.47	396.52
	20	2	472.66			2	356.88		
White Blood Cell Count 10^9/L									
	0	76	7.20	5.96	8.16	169	7.33	6.13	8.47
	1	0				0			
	2	78	6.74	5.91	7.82	152	7.12	6.13	8.34
	4	78	6.73	5.80	7.85	149	7.15	5.90	8.34
	8	74	6.88	5.99	8.15	136	6.96	6.00	8.07
	12	72	6.77	5.93	8.28	133	7.27	6.23	8.47
	16	5	6.28	4.57	8.59	6	7.72	6.21	9.98
	20	2	7.25			2	8.75		

12 Programming

12.1 Methods

This report was produced using high-quality open source, freely available R and LaTeX packages. High-level R graphics and LaTeX making functions in FE Harrell's `Hmisc` package were used in the context of the R `knitr` package. A new R package `greport` contains functions `accrualReport`, `dReport`, `exReport`, `eReport`, and `survReport` 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.

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. If using pop-up method `poptype=1`, the user must install the following LaTeX packages: `acrotex`, `ocgtools`, and `asymptote`. If using `poptype=2`, the user must install the `tooltip` style. See http://biostat.mc.vanderbilt.edu/Greport for more information. `poptype=3` just puts the denominator information as tiny tables in figure captions and don't require use of any special LaTeX packages. This approach solves a problem with Macs and Ipads not handling pop-ups correctly.

Before running the R code to produce the report components, create the following directories underneath your project directory: `pdf` and `gentex`, to hold `pdf` graphics and generated LaTeX code, respectively. You can change the name of these directories using the `setgreportOption` function.

The entire process is best managed by creating a single `.Rnw` file that is executed using the `knitr` package in R. **Note**: When using `knitr` with `cache=TRUE` it is assumed that no cached chunks produce appendix tables. For debugging, it is recommended that slow chunks be cached, then to make sure the entire appendix is generated turn off all caching and re-run the program.

The user musc define a function `spar` using the one in http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RConfiguration/Rprofile as a template. `spar` sets good default graphical parameters (e.g., space between axis labels and axes) for non-`lattice` R graphics. The `Rprofile` example there also defines the `knitrSet` function that is used in this example, to pretty-print R code.

12.2 Changing Lattice Graphics Parameters

The most common change needed in Lattice graphics is the font size in strip labels, especially to allow longer labels to fit. Here is a summary of how to change this in a few `greport` contexts.

violin and box plots with `dReport` : `sopts=list(cex.strip=.6)`

Proportion charts with `dReport` : `popts=list(par.strip=list(cex=.6))`

12.3 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 `label`s 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 for details about setting up analysis files.

R 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.

12.3.1 Data Assumptions

1. Non-randomized subjects are marked by missing data of randomization
2. The treatment variable is always the same for every dataset and is defined in `tx.var` on `setgreportOption`.
3. For some graphics there must be either no treatment variable or exactly two treatment levels.
4. If there are treatments the design is a parallel-group design.
5. Whenever a dataset is specified to one of the `greport` functions and subject have repeated measurements (> 1 record), an `id` variable must be given.

12.4 User `knitr` Source File for This Document

```
% Usage: Copy report.Rnw and ssafety.rda to a temporary directory and
% make directories gentex and pdf under it, then run knitr on report.Rnw
% To compile: knitr report; pdflatex report
% Produces: report.pdf
\documentclass{article}
\usepackage{knitrl}
\usepackage{moreverb}   % for \verbatimtabinput
\usepackage{fancyhdr}   % for fancy headers
\usepackage{url}        % for \url{}
\usepackage{changepage} % for exReport \begin{adjustwidth}

\def\titl{Example Closed Meeting Data Monitoring Committee Report}
\definecolor{darkblue}{RGB}{0,0,139}
\def\linkcol{darkblue}
```

```
\usepackage[pdftex,hidelinks,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)

\def\poptype{3}   % 0=no pop-up tooltips  1=ocgtools  2=movable pop-ups
                  % 3=no pop-ups, put as tiny tables in figure captions
\usepackage{greport}

\graphicspath{{pdf/}}

\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

\begin{document}
\maketitle
\tableofcontents
\listoffigures
\listoftables
\clearpage
\rhead{\scriptsize The {\em EXAMPLE} Study \\
    Protocol xyz--001 \\
    \today}

<<setup,echo=FALSE,results='asis'>>=
echo <- TRUE     # include code in report
# echo <- FALSE    # exclude code in report
cat('%------------------\n')
cat(sprintf('\\def\\inclcode{%s}\n', 1 * echo))
require(greport)
knitrSet(echo=echo)
@

<<>>=
Load(ssafety)
ssafety <- upData(ssafety, rdate=as.Date(rdate),
                  smoking=factor(smoking, 0:1, c('No','Yes')),
                  labels=c(smoking='Smoking', bmi='BMI',
                    pack.yrs='Pack Years', age='Age',
                    height='Height', weight='Weight'),
                  units=c(age='years', height='cm', weight='Kg'),
                  print=FALSE)
```

```
mtime <- function(f) format(file.info(f)$mtime)
datadate        <- mtime('ssafety.rda')
primarydatadate <- mtime('ssafety.rda')

## 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))))
week  <- ssafety$week
weeks <- sort(unique(week))
base  <- subset(ssafety, week==0)
denom <- c(c(enrolled=500, randomized=nrow(base)), table(base$trx))

setgreportOption(tx.var='trx', denom=denom, texwhere='')
## Initialize app.tex
file <- sprintf('%s/app.tex', getgreportOption('texdir'))
cat('', file=file)
@

\section{Philosophy}
The reporting tools used here are based on a number of lessons learned
from the intersection of the fields of statistical graphics, graphic
design, and cognitive psychology, especially from the work of Bill Cleveland,
Ralph McGill, John Tukey, Edward Tufte, and Jacques Bertin.
\begin{enumerate}
\item Whenever largely numerical information is displayed, graphs
  convey the information most often needed much better than tables.
  \begin{enumerate}
    \item Tables usually show more precision than is warranted by the
      sample information while hiding important features.
    \item Graphics are much better than tables for seeing patterns and
      anomalies.
    \end{enumerate}
\item The best graphics are ones that make use of features that humans are most
  accurate in perceiving, namely position along a common scale.
\item Information across multiple data categories is usually easier to
  judge when the categories are sorted by the numeric quantity
  underlying the information\footnote{This also facilitates
    multivariate understanding of trends and differences.  For
    example, if one sorted countries by the fraction of subjects who
```

```
    died and displayed also the fraction of subjects who suffered a
    stroke, the extent to which stroke incidence is also sorted by
    country is a measure of the correlation between mortality and
    stroke incidence across countries.}.
\item The most robust and informative descriptive statistics for
  continuous variables are quantiles and whole distribution
  summaries\footnote{In particular, the standard deviation is not very
    meaningful for asymmetric distributions, and is not robust to outliers.}.
\item For group comparisons, confidence intervals for individual
  means, medians, or proportions are not very useful, and whether or
  not two confidence intervals overlap is not the correct statistical
  approach for judging the significance of the difference between the
  two.  The half-width of the confidence interval for the difference,
  when centered at the midpoint of the two estimates, provides a
  succinct precision display, and this half-interval touches the
  two estimates if and only if there is no significant difference
  between the two.
\item Each graphic needs a marker that provides the reader with a
  sense of exactly what fraction of the sample is being analyzed in
  that graphic.
\item Tables are best used as backups to graphics.
\item Tables should emphasize estimates that are not functions of the
  sample size.  For categorical variables, proportions have
  interpretations independent of sample size so they are the featured
  estimates, and numerators and denominators are subordinate to the
  proportions.  For
  continuous variables, minimum and maximum, while useful for data
  quality checking, are not population parameters, and they expand as
  $n\uparrow$, so they are not proper summary statistics.
\end{enumerate}

\section{Notation}
\ifnum\poptype > 0
\ifnum\poptype < 3
\paragraph{Pop-up Tooltips}
Certain elements of the report, signaled by
\textcolor[gray]{0.5}{$\mapsto$}, have pop-up tooltips behind them.
More information will pop up when viewing the report under Acrobat
Reader when the mouse hovers over \textcolor[gray]{0.5}{$\mapsto$}.
\ifnum\poptype=1
Clicking on the information in the pop-up will make it ``stick'', and
clicking on the \textcolor{red}{X} will make it disappear.  For
graphics that have pop-up tables you can also click anywhere inside
the graph.  When the pop-up is a wide table, it will use full-page
mode.  If the table is tall you may need to scroll vertically.  To do
that, click on the table when it pops up to make it stick, then
scroll, then click again to make it disappear.
\fi
\ifnum\poptype=2
Clicking on the pop-up and releasing will allow you to move the pop-up
```

```
with a mouse gesture (do not hold the mouse button down).  Click on
the pop-up to make it stick in a certain location.  Hover over
\textcolor[gray]{0.5}{$\mapsto$} to make the pop-up disappear, or
click on the pop-up again to unstick it.
\fi
\fi
\fi

\paragraph{Hyperlinks and Tables}
Some graphics and tables are hyperlinked to tables
in the Appendix.  For these, clicking anywhere in the graphic or table
will move the pdf reader to the supporting table.  Clicking on the
appendix table will bring you back to the original figure.
\ifnum\pdfstrcmp{\linkcol}{black}=0
%
\else
Other than for graphics, objects appearing in
\textcolor{\linkcol}{this color} are hyperlinked.
\fi

\paragraph{Viewers}
You must use Adobe Acrobat Reader to view pdf files generated by
\code{greport}, otherwise pop-ups will not work. Neither pop-ups nor
hyperlinks will work if you view documents in a Web browser window.
It is recommended that you click on \texttt{View \dots Page Display
  \dots Single Page} for optimum jumping between hyperlinks, i.e., do
  not use \texttt{Single Page Continuous} mode.

\paragraph{Figure Captions}
Needles represent the fraction of observations used in the current
analysis.  The first needle (red) shows the fraction of enrolled
patients used.  If randomization was taken into account, a second
needle (green) represents the fraction of randomized subjects included
in the analysis.  When the analyses consider treatment assignment, two
more needles may be added to the display, showing, respectively, the
fraction of subjects randomized to treatment A used in the analysis
and the fraction of subjects on treatment B who were analyzed.  The
colors of these last two needles are the colors used for the two
treatments throughout the report.  The following table shows some
examples.
<<results='asis'>>=
# dNeedle uses colors in setgreportOption(tx.col=, er.col=)
dNeedle(1,           'lttdemoa')
dNeedle(c(3,4)/4 ,   'lttdemob')
dNeedle(c(1,2)/4,    'lttdemoc')
dNeedle(c(1,2,3,1)/4,'lttdemod')
@

\begin{center}
  \begin{tabular}{ll}
```

```
      \textbf{Needles} & \textbf{Interpretation} \\ \hline
      \lttdemoa & All enrolled subjects analyzed, randomization not considered\\
      \lttdemob & Analysis uses $\frac{3}{4}$ of enrolled subjects,
                  and all randomized subjects\\
      \lttdemoc & Analysis uses $\frac{1}{4}$ of enrolled subjects,
                  and $\frac{1}{2}$ of randomized subjects\\
      \lttdemod & Same as previous example, and in addition the analysis\\
                & utilized treatment assignment, analyzing $\frac{3}{4}$ of
                  those\\
                & randomized to A and $\frac{1}{4}$ of those randomized to B\\
      \hline
  \end{tabular}
\end{center}
\ifnum\poptype > 0
\ifnum\poptype < 3
There are pop-up tooltips embedded in the needles.  When hovering the
mouse over \textcolor[gray]{0.5}{$\mapsto$} a table of subject counts
will pop up.
\fi
\fi

\paragraph{Extended Box Plots}
% For poptype 1 and 2:
%\newcommand{\eboxpopup}[1]{\tooltipm{#1}{\includegraphics{bpplt-proto-1}}}
% For poptype 3:
\newcommand{\eboxpopup}[1]{\hyperlink{bpplt}{#1}}
% To not generate pop-up use: \newcommand{\eboxpopup}[1]{}
For depicting distributions of continuous variables, many of the
following displays use extended box plots, also called
box--percentile plots.  The plots are scaled to the marginal 0.01 and
0.99 quantiles of the variables.  A prototype, with explanations, is
below.  When viewing the report, hovering the
mouse over the word ``box'' will pop up this prototype as needed.
<<bpplt-proto,w=5,h=3.5>>=
bpplt()
@
\hypertarget{bpplt}{}

Typically violin plots are superimposed onto box plots in what
follows.  Violin plots show mirror images of the estimated probability
density function for continuous variable, using a kernel density
estimator.  Violin plots are better able to show multimodality than
quantile intervals.

\paragraph{Dot Charts}
Dot charts are used to present stratified proportions.  In these
charts the area of the symbols is proportional to the square root of
the denominator.  The legend shows representative denominators and
their corresponding symbol areas, using denominators that actually
occurred in the data and extended from the minimum observed to the
```

```
maximum observed sample size.


\paragraph{Longitudinal Analysis}
For continuous variables measured repeatedly, line plots show the
median as a function of time.  Next to each point is a series of thin
vertical lines, one series on the left for treatment A and another
series on the right for treatment B, when stratifying by treatment.
Moving outward from the point showing the median, these lines depict
the following quantile intervals---the same ones depicted in the
extended box plots.
\def\quantint{
\begin{tabular}{crclc} \hline
  Horizontal & \multicolumn{3}{c}{Quantiles} & Fraction of\\
  Sequence   & \multicolumn{3}{c}{Spanned}   & Sample Covered \\ \hline
  ~~~~~~$1$  & $\frac{1}{20}$ &-& $\frac{19}{20}$  & $\frac{9}{10}$ \\
  ~~~~$2$    & $\frac{1}{8}$  &-& $\frac{7}{8}$    & ~~$\frac{3}{4}$  \\
  ~~$3$      & $\frac{1}{4}$  &-& $\frac{3}{4}~$   & ~~~~$\frac{1}{2}$  \\
  $4$        & $\frac{3}{8}$  &-& $\frac{5}{8}$    & ~~~~~~$\frac{1}{4}$  \\ \hline
\end{tabular}
}
\begin{center}
\quantint\hypertarget{quantint}{}
\end{center}
%\newcommand{\qintpopup}[1]{\tooltipn{#1}{\quantint}}
\newcommand{\qintpopup}{\hyperlink{quantint}{these~}}
% To not generate pop-up use \newcommand{\qintpopup}[1]{}
\ifnum\poptype > 0
\ifnum\poptype < 3
These definitions will pop-up when hovering the mouse at the end of
the phrase ``quantile intervals'' in captions.
\fi
\fi

In addition there is a black vertical line centered at the midpoint of
the two medians, with height equal to $\frac{1}{2}$ of the width of an
approximate 0.95 confidence interval for the difference in the two
medians.  When the two medians touch this vertical bar, there is
approximately no significant difference in the medians at the 0.05
level.  The Harrell--Davis quantile estimator is used, along with its
standard error estimate for the median.

Instead of quantile intervals, longitudinal plots may show vertical
violin plots.  Unlike the mirror--image violin plots shown on box
plots, there there are two groups being compared the first group has a
half-violin plot on the left of the point showing the median, and the
second group has a half-violin plot on the right. Back--to--back
comparisons of probability density functions are useful for comparing
entire distributions of continuous variables.  When for a group the
sample size is less than 10 the violins are more faint, and when the
```

```
sample size is less than 5 they are barely visible.

For binary variables measured repeatedly, the 0.95 Wilson confidence
intervals are shown on either side of the proportions, and a vertical
black line appears over the proportions.  The height of this line is
$\frac{1}{2}$ the length of the normal-approximation 0.95 confidence
interval for the difference in the two proportions.  When the
proportions fail to touch this line, they are approximately
significantly different at (at least) the 0.05 level.

For discrete variables that are not binary, such as adverse event
presence and severity, means and bootstrap percentile 0.95 confidence
intervals are shown, along with half-confidence intervals for the
difference in means using as standard error the square root of the sum
of squares of the two means' standard errors.

\paragraph{Survival Curves}
Graphs containing pairs of Kaplan-Meier survival curves show a shaded
region centered at the midpoint of the two survival estimates and
having a height equal to the half-width of the approximate 0.95 pointwise
confidence interval for the difference of the two survival
probabilities.  Time points at which the two survival estimates do not
touch the shaded region denote approximately significantly different
survival estimates, without any multiplicity correction.
\clearpage

\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 \R\ \code{gsDesign} package.}
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 \Sexpr{datadate}.  Source analysis files were last
updated on \Sexpr{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{Accrual}
<<accrual,results='asis'>>=
accrualReport(randomize(rdate) ~ site(site), data=base,
              dateRange=c('1990-01-01','1994-12-31'),
              targetDate='1994-12-31', targetN=300,
              closeDate=max(base$rdate))
@
\clearpage
\section{Baseline Variables}
<<baseline,results='asis'>>=
dReport(sex + race + smoking ~ trx, groups='trx', data=base,
        h=4, w=3.5)

## Show spike histogram for raw data, 50 bins
dReport(age + height + weight + bmi + pack.yrs ~ trx, data=base,
        h=3.5,
        sopts=list(datadensity=TRUE,
          scat1d.opts=list(nhistSpike=1,
            col=adjustcolor('red', alpha.f=.5),
            nint=50)),
        append=TRUE)
@
\clearpage

%\section{Compliance to Assigned Treatments}
%complianceReport(ssafety$comply, ssafety$trx, ssafety$week,
%                 weeks[weeks > 1])

%\section{Dropouts}
%dropoutReport(base$d.dropout, base$dropout, base$trx, time.inc=14)

\section{Longitudinal Adverse Events}
<<longae,cache=FALSE,results='asis'>>=
dReport(headache + ab.pain + nausea + dyspepsia + diarrhea +
        upper.resp.infect + coad ~ week + trx + id(id),
        groups='trx', data=ssafety, panel='longae', what='byx',
        popts=list(cex.strip=.57,
          key=list(x=.65, y=.2, lines=TRUE, points=FALSE)))
@
\clearpage

\section{Incidence of Adverse Events at Any Follow-up}
<<anyae,results='asis'>>=
```

```
## Reformat to one record per event per subject per time
aev <- vars$ae
ev  <- ssafety[ssafety$week > 0, c(aev, 'trx', 'id', 'week')]
## Reshape to tall and thin format
evt <- reshape(ev, direction='long', idvar=c('id', 'week'),
               varying=aev, v.names='sev', timevar='event',
               times=aev)
## For each event, id and trx see if event occurred at any week
ne <- with(evt, summarize(sev, llist(id, trx, event),
                          function(y) any(y > 0, na.rm=TRUE)))
## Remove non-occurrences of events
ne <- subset(ne, sev, select=c(id, trx, event))
## Replace event names with event labels
elab <- sapply(ssafety[aev], label)
ne$event <- elab[ne$event]
label(ne$trx) <- 'Treatment'

eReport(event ~ trx, data=ne, h=3.25)
@
\clearpage

\section{Longitudinal EKG Data}
<<ekg,results='asis'>>=
dReport(axis + corr.qt + pr + qrs + uncorr.qt + hr ~ week + trx +
        id(id),
        groups='trx', data=ssafety, panel='ekg', what='byx', w=7)
@
\clearpage

\section{Longitudinal Clinical Chemistry Data}
<<cchem,cache=FALSE,results='asis'>>=
## Plot 6 variables per page
cvar <- split(vars$chem, rep(letters[1:4], each=6))
for(subpanel in names(cvar)) {
  form <- paste(cvar[[subpanel]], collapse=' + ')
  form <- as.formula(paste(form, 'week + trx + id(id)', sep=' ~ '))
  dReport(form, groups='trx', data=ssafety, panel='cchem',
          subpanel=subpanel,
          what='byx', append=subpanel != 'a',
          popts=list(cex.strip=.7), w=7)
  }
## Repeat last figure using quantile intervals instead of violin densities
dReport(form, groups='trx', data=ssafety, panel='cchem',
        subpanel='e', what='byx', byx.type='quantiles', append=TRUE,
        popts=list(cex.strip=.7), w=7)
@
\clearpage  % needed to get last tooltips to work

\section{Computing Environment}
These analyses were done using the following versions of R\cite{Rsystem}, the
```

```
operating system, and add-on packages \code{greport}\cite{greport},
\code{Hmisc}\cite{Hmisc}, \code{rms}\cite{rrms}, and others:
<<echo=FALSE,results='asis'>>=
toLatex(sessionInfo(), locale=FALSE)
@
The reproducible research framework \code{knitr}~\cite{knitrbook} was used.

\bibliography{feh.bib}
\bibliographystyle{unsrt}
\clearpage

\section{Appendix: Supporting Tables}
\input{gentex/app}
\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.  High-level \R\ graphics and \LaTeX\
making functions in FE Harrell's \code{Hmisc} package were used in the
context of the \R\ \code{knitr} package.
A new \R\ package \code{greport} contains functions
%\code{completeness\-Report},
\code{accrual\-Report},
\code{dReport}, %\code{rep\-Varclus},
%\code{compliance\-Report}, \code{dropout\-Report},
\code{ex\-Report}, \code{e\-Report}, and \code{surv\-Report}
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.  If using pop-up method \code{poptype=1},
the user must install the following \LaTeX\ packages:
\code{acrotex}, \code{ocgtools}, and \code{asymptote}.  If using
\code{poptype=2}, the user must install the \code{tooltip} style.
See \url{http://biostat.mc.vanderbilt.edu/Greport} for more information.
\code{poptype=3} just puts the denominator information as tiny tables
in figure captions and don't require use of any special \LaTeX\
packages.  This approach solves a problem with Macs and Ipads not
handling pop-ups correctly.
```

```
Before running the \R\ code to produce the report components, create the
following directories underneath your project directory: \code{pdf}
and \code{gentex}, to hold \code{pdf} graphics and generated \LaTeX\
code, respectively.  You can change the name of these directories
using the \code{setgreportOption} function.

The entire process is best managed by creating a single \code{.Rnw}
file that is executed using the \code{knitr} package in \R.
\textbf{Note}: When using \code{knitr} with
\code{cache=TRUE} it is assumed that no cached chunks produce appendix
tables.  For debugging, it is recommended that slow chunks be cached,
then to make sure the entire appendix is generated turn off all
caching and re-run the program.

The user musc define a function \code{spar} using the one in
\url{http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RConfiguration/Rprofile}
as a template.  \code{spar} sets good default graphical parameters
(e.g., space between axis labels and axes) for non-\code{lattice} \R\
graphics.  The \code{Rprofile} example there also defines the
\code{knitrSet} function that is used in this example, to pretty-print
\R\ code.

\subsection{Changing Lattice Graphics Parameters}
The most common change needed in Lattice graphics is the font size in strip labels, especially to allo

\begin{description}
\item[violin and box plots with \code{dReport}]: \code{sopts=list(cex.strip=.6)}
\item[Proportion charts with \code{dReport}]: \code{popts=list(par.strip=list(cex=.6))}
\end{description}

\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/wiki/pub/Main/RS/sintro.pdf}
{Alzola and Harrell, 2006},
\href{http://biostat.mc.vanderbilt.edu/StatGraphCourse}, and
\href{http://biostat.mc.vanderbilt.edu/StatCompCourse} for details
about setting up analysis files.
```

```
\R\ 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.

\subsubsection{Data Assumptions}
\begin{enumerate}
  \item Non-randomized subjects are marked by missing data of randomization
  \item The treatment variable is always the same for every dataset
    and is defined in \code{tx.var} on \code{setgreportOption}.
  \item For some graphics there must be either no treatment variable
    or exactly two treatment levels.
  \item If there are treatments the design is a parallel-group design.
  \item Whenever a dataset is specified to one of the \code{greport}
    functions and subject have repeated measurements ($>1$ record), an
    \code{id} variable must be given.
\end{enumerate}

\subsection{User \code{knitr} Source File for This Document}
{\small\verbatimtabinput{report.Rnw}}

}\fi

\end{document}
```