



---

# *Journal of Statistical Software*

February 2009, Volume 29, Issue 6.

<http://www.jstatsoft.org/>

---

## PSAgraphics: An R Package to Support Propensity Score Analysis

James E. Helmreich  
Marist College

Robert M. Pruzek  
University at Albany

---

### Abstract

Propensity score analysis is a technique for adjusting for selection bias in observational data. Estimated propensity scores (probability of treatment given observed covariates) are used for stratification of observations. Within strata covariates should be more balanced between the two treatments than without the stratification. **PSAgraphics** is an R package that provides flexible graphical tools to assess within strata balance between treatment groups, as well as how covariate distributions differ across strata. Additional graphical tools facilitate estimation of treatment effects having adjusted for covariate differences. Several new and conventional numerical measures of balance are also provided.

*Keywords:* propensity score analysis, PSA, R, covariate balance, strata.

---

## 1. Introduction

When two treatments are compared using propensity score analysis (PSA), a key goal is to achieve similar covariate distributions between treatment groups across strata. Similarities are sought for both quantitative and categorical variables. The central aim of the **PSAgraphics** package is to provide a variety of flexible graphics functions to facilitate visualization of propensity score analysis results when two treatments are compared; also to help study details about covariate balance, principally through graphics. We aim to facilitate studies of covariate balance within and across strata for one covariate at a time. Numerical results are also provided by all functions, both to complement the graphical assessments of balance, and to provide detailed descriptive and inferential information about treatment effects.

Observational studies often suffer from a lack of covariate balance when comparing two treatment groups. For example, sicker patients are more likely to get an active treatment. This leads to outcome-defined effects being confounded with covariate differences. One strategy to overcome this problem is to stratify on the basis of what an analyst regards as especially

relevant or important covariates (cf. Cochran 1968) where he takes account of ages among smokers. But direct stratification on covariates is only practical when there are few covariates. An alternative due to Rosenbaum and Rubin (1983) is to estimate the probability that each observation is in a treatment group, i.e., the propensity score, using observed covariates, and then to stratify units into a few strata (five is standard) within which it may be reasonable to expect that at least the observed covariate distributions will be balanced across the treatment groups.

The package **PSAgraphics**, written for the statistical software platform R (R Development Core Team 2008), provides four functions to help assess the balance of covariates within strata. Two other functions help visualize the direction and strength of effects across strata and, using loess-based methods, help to see treatment effects across the range of the estimated propensity scores. Most functions are also highly versatile in they can accommodate numerous user-defined choices. Experience suggests that insights often follow from simultaneous examinations of covariate distributions within and across strata, as this method facilitates learning how strata work to achieve covariate distribution balance. The functions provided in this package are designed to improve data analyses by facilitating a detailed understanding of treatment effects, and by helping to learn where covariate confounding is and is not ignorable. In analyses involving few, though more than two treatments, good use of the package may be made using appropriately designed treatment contrasts. This package is made available in R with the intent that it can be used alongside a variety of other packages and functions; thus the user can expect to leverage any of several capabilities of the R platform in any one data analysis. **PSAgraphics** is available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=PSAgraphics>.

Conventional approaches to assessing covariate balance treat strata individually. For instance, paired boxplots or histograms are often used; but when strata are only compared individually, there are no visual benchmarks for comparing within-strata differences with across-strata differences. It can be particularly helpful to visualize what is distinctive about strata vis-à-vis a particular covariate in the context of particular propensity score estimates (and usually there will be more than one way to estimate propensity scores). Investigators generally need to be informed about how either quantitative or categorical covariate distributions differ between treatment groups, within strata. There may be concerns about location, variance, skewness or outliers. If the two treatment groups appear dissimilar (when derived propensity scores are not providing covariate balance) the investigator generally seeks knowledge about the details. This package aims to address these key concerns with two main functions, `box.psa` and `cat.psa`, which use boxplots and bargraphs to exhibit paired distributions within strata, as well as to facilitate across strata comparisons. The main goal is to provide good visual comparisons of these distributions, but additionally, four numeric balance statistics are calculated, two for each function.

For assessing effects, the package contains two other key functions. The first, `circ.psa`, is what we term a PSA assessment plot, a new graphic that displays effects for individual strata, as well as an overall summary of effects in the context of propensity-based adjustments. The graphic can be seen as an enhanced scatterplot that allows for strata of differing sizes, shows a weighted average of effects, and a graphical representation of the 95% confidence interval for the overall effect. Finally, `loess.psa`, which has a form first suggested by R. L. Obenchain (see Obenchain 2007), provides estimates of effects using non-parametric regressions, separately generated for the two groups being compared, across the range of estimated propensity scores,

without reference to strata.

## 2. Lindner dataset

We make use of the observational data frame `lindner`, first provided with the R package **USPS**, by [Obenchain \(2007\)](#). The `lindner` data contain data on 996 patients treated at the Lindner Center, Christ Hospital, Cincinnati in 1997. Patients received a Percutaneous Coronary Intervention (PCI). The data consists of 10 variables. Two are outcome: `lifepres` ranges over two values, 11.4 or 0 depending on whether patients survived to six months post treatment or did not survive to six months (respectively), where 11.4 is the mean expected preserved life years for those patients who survive to six months. Secondly, `cardbill` contains the costs in 1998 dollars for the first six months (or less if patient did not survive) after treatment with the drug abciximab. For simplicity we confine our analysis to `cardbill`, focusing on the logarithm of this variate. The treatment variable is `abcix`, where 0 indicates standard PCI treatment and 1 indicates standard PCI treatment and additional treatment in some form with abciximab. Covariates include `acutemi`, 1 indicating a recent acute myocardial infarction and 0 not; `ejecfrac` for the left ventricle ejection fraction, a percentage from 0 to 90; `ves1proc` giving the number of vessels (0 to 5) involved in the initial PCI; `stent` with 1 indicating coronary stent inserted, 0 not; `diabetic` where 1 indicates that the patient has been diagnosed with diabetes, 0 not; `height` in centimeters and `female` coding the sex of the patient, 1 for female, 0 male.

## 3. Estimation of propensity scores, production of strata

We estimate propensity scores in two ways: via logistic regression and a recursively partitioned tree.

```
R> data("lindner")
R> attach(lindner)
R> lindner.log <- glm(abcix ~ stent + height + female + diabetic +
+   acutemi + ejecfrac + ves1proc, data = lindner, family = binomial)
R> ps <- lindner.log$fitted
R> lindner.s5 <- cut(ps, quantile(ps, seq(0, 1, 1/5)),
+   include.lowest = TRUE, labels = FALSE)
R> lindner.s10 <- cut(ps, quantile(ps, seq(0, 1, 1/10)),
+   include.lowest = TRUE, labels = FALSE)
```

In `lindner.s5` we produced five strata of roughly equal size, ten strata in `lindner.s10`. Alternatively, propensity scores can be estimated and strata defined using a recursively partitioned tree. This generally provides strata that generally differ in their sizes. We use the R function `rpart` from the package of the same name by [Therneau and Atkinson \(2002\)](#) to generate a tree with six bins or strata; the tree is not reproduced here.

```
R> library("rpart")
R> lindner.rpart <- rpart(abcix ~ stent + height + female + diabetic +
```

```
+   acutemi + ejecfrac + ves1proc, data = lindner, method = "class")
R> lindner.tree <- factor(lindner.rpart$where, labels = 1:6)
```

Each illustration of these methods represents only one way of defining strata. One of the main goals of several of the functions in the package **PSAgraphics** is to aid in the choice and refinement of appropriate bins. Note that the user can compare graphics for different numbers of strata or bins, for different ways of estimating propensity scores, etc. In addition, the functions aim to facilitate comparisons of balance and treatment effects with respect to various choices in the context of particular data analyses.

#### 4. Balance graphics for continuous covariates

Having defined strata based upon estimated propensity scores, we aim to assess balance within each stratum for individual covariates between the two treatment groups, as well as to examine these individual covariate distributions across strata. We generally consider one group to be a control (C) and the other treatment (T). For continuous covariates, the function `box.psa` generates a graphic that shows pairs of boxplots of the treatment and control groups across all strata. In general, balance is assessed one covariate at a time (see Figure 1). It is notable that [Sarkar \(2008\)](#) has developed a comprehensive R package **lattice** to facilitate many trellis style graphics. So for example, were an analyst to examine balance for two covariates at a time across several strata, Chapter 5 of [Sarkar \(2008\)](#) provides relevant code. A lattice of scatterplots, one for each stratum, with different characters for treatment and control entities could be helpful.

```
R> box.psa(ejecfrac, abcix, lindner.s5, xlab = "Strata",
+   ylab = "Covariate: ejecfrac", legend.xy = c(2, 110), balance = FALSE)
```

Boxplots of the covariate left ventricle ejection fraction, `ejecfrac`, are presented by pairs across all strata. Means of control and treatment are compared using heavy black lines. Visually, the slope of these lines gives an indication of mean differences between the treatment groups. However the boxplot comparisons also provide details as to how the distributions compare both within each stratum and across strata. In this case, we see that each stratum is reasonably balanced as the distributions of `ejecfrac` for control and treatment groups are visually similar for each stratum. Furthermore, the distribution does not change dramatically across strata. Note that the sizes of treatment groups are printed below the boxplots. The user may specify an optional `trim` parameter, and the appropriate trimmed mean will be calculated and graphed. The maximum `trim = 0.5` will of course produce and connect the medians already shown in the boxplot. In addition, optional arguments (e.g., `notch`) may be passed to the (internally called) `boxplot` function.

The `rpart` tree-derived strata yield a slightly different picture (see Figure 2).

```
R> box.psa(ejecfrac, abcix, lindner.tree, xlab = "Strata",
+   ylab = "Covariate: ejecfrac", legend.xy = c(2, 110), balance = TRUE)
```

In contrast to the logit-based strata, tree-derived strata yield one small stratum that identifies most (of the few) patients with a high left ventricle ejection fraction. The balance graphic

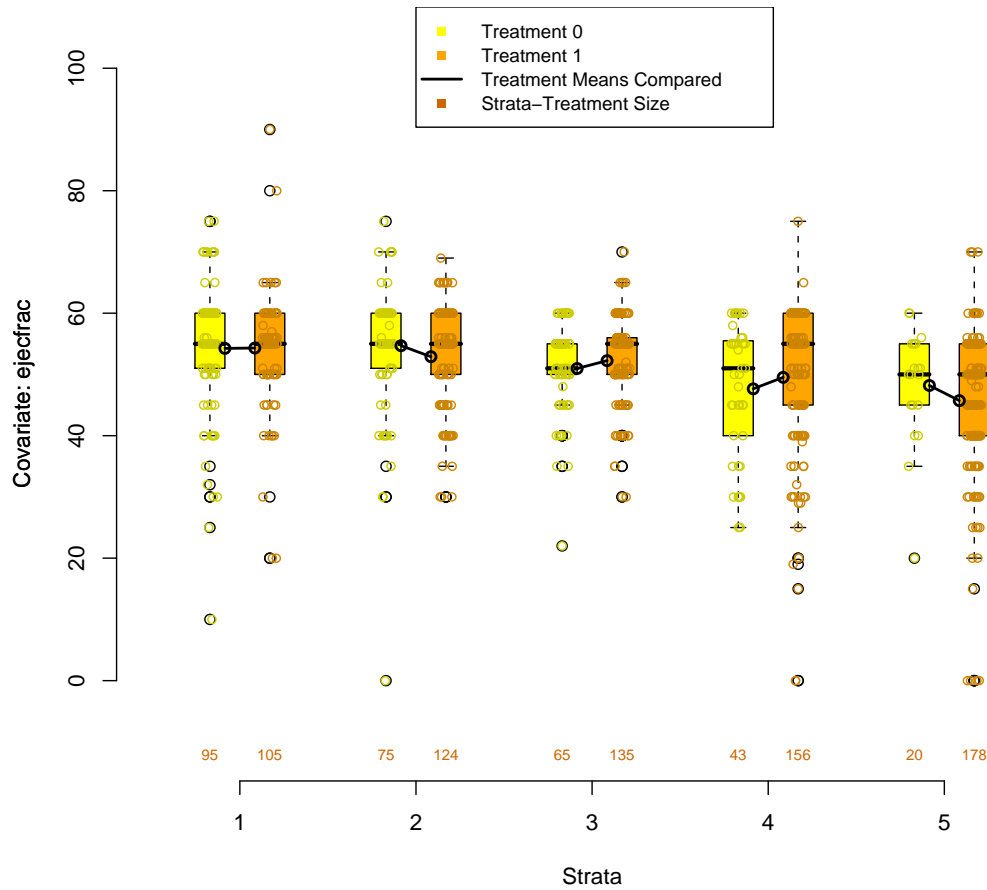


Figure 1: Side-by-side boxplots, 5 strata, for covariate `ejecfrac` produced by `box.pdf`.

shows that one or two patients with large values are left in the treatment groups of strata four and six; further examination of results for these patients might be warranted.

It is generally useful to have numerical as well as graphical measures of differences in treatment distributions. The `balance` option for `box.psa` provides two such numerical measures. First, `bal.ks.psa` uses the Kolmogorov-Smirnov test of equivalence of two distributions to provide a statistic that may be useful for comparing a covariate's distributions within any stratum. Using this test,  $p$  values for a test of the equivalence are shown in red below each pair of boxplots (see Figure 2).<sup>1</sup>

Additionally, it may be of interest to assess whether strata defined in other ways might balance the covariate better than the given stratification. For an assignment  $\alpha$  of patients or cases to  $K$  strata, we define the statistic  $\hat{\delta}_\alpha$ . Let  $T_{ik}$ ,  $i = 0, 1$ ,  $k = 1, \dots, K$  be the set of cases in the  $i$ th

<sup>1</sup>The Kolmogorov-Smirnov test assumes continuous distributions, and thus no common values in the two samples. Exact  $p$  values are not possible with ties.

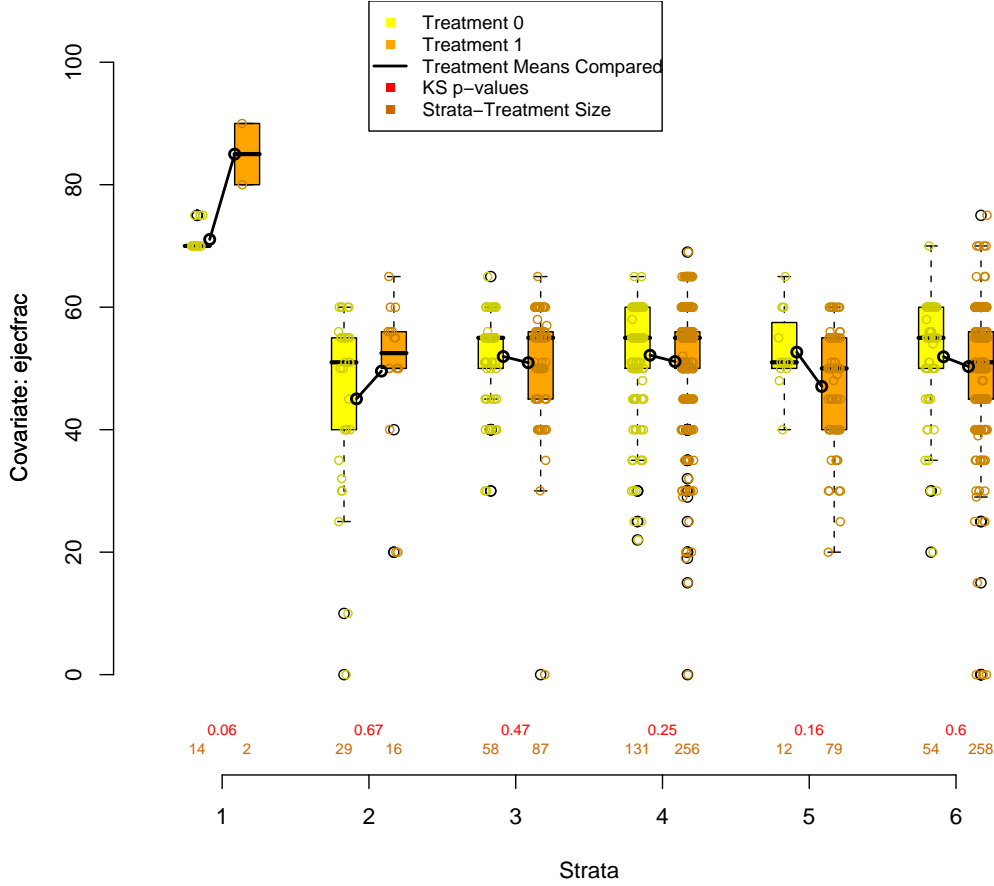


Figure 2: Side-by-side boxplots, 6 `rpart` derived strata, covariate `ejecfrac`.

treatment and  $k$ th stratum of  $\alpha$ . Let  $\hat{\mu}_{ik}$  be the covariate mean (possibly trimmed) for cases in  $T_{ik}$ . Then  $\hat{\delta}_\alpha = \sum_{k=1}^K |\hat{\mu}_{0k} - \hat{\mu}_{1k}|$ . In particular, let  $\alpha$  be the stratification we are assessing for adequate covariate balance. The function `bal.ms.psa` finds  $\hat{\delta}_\alpha$  (termed the original balance measure) and generates a permutation distribution by randomly assigning cases to strata, thus generating stratifications  $\alpha_b$ ,  $b = 1, \dots, B$ , and the associated permutation statistics  $\hat{\delta}_{\alpha_b}$ . The  $\alpha_b$  are generated under a fixed marginals model to retain comparability with  $\alpha$ ; the distribution of cases between treatments within strata are allowed to differ. The permutation distribution for the statistic  $\hat{\delta}_\alpha$  is plotted as a histogram. To the extent to which the observed or original  $\hat{\delta}_\alpha$  is smaller than the mass of the permutation distribution one has evidence that  $\alpha$  has balanced the given covariate distribution. Finally, we note that if the user specifies a nonzero `trim` argument to the function, then all calculated statistics  $\hat{\delta}_\alpha$  will be based on differences in trimmed means.

The function `bal.ms.psa` is usually called through the `balance = TRUE` option of `box.psa`, which will first produce the histogram of the permutation distribution of  $\hat{\delta}_\alpha$ , before the boxplot

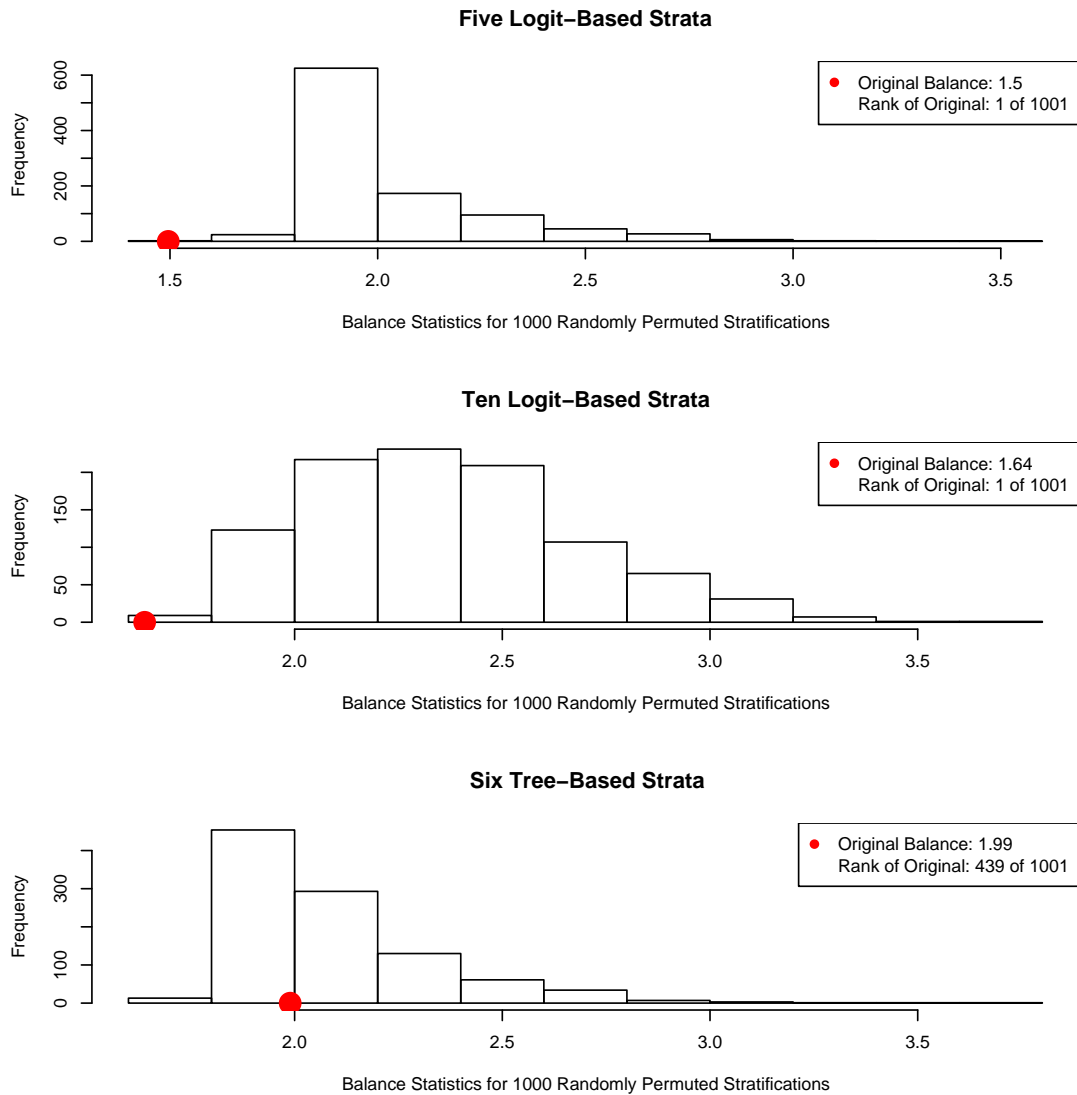


Figure 3: Comparison for covariate `ejecfrac` of the original balance statistics  $\hat{\delta}_\alpha$  and associated permutation distributions of  $\hat{\delta}_{\alpha_b}$  for stratifications of 5 and 10 strata using logit-based propensity scores and one of 6 strata based on a recursively partitioned tree.

graphic. For comparison purposes, we use `bal.ms.psa` directly to calculate  $\hat{\delta}_\alpha$ , and respective distributions associated with the three distinct stratifications defined above (see Figure 3).

```
R> bal.ms.psa(ejecfrac, abcix, lindner.s5, B = 1000, main = "Five
+   Logit-Based Strata")
```

```
$balance.orig
[1] 1.495801
```

```
$rank.orig
```

```
[1] 1
```

```
R> bal.ms.psa(ejecfrac, abcix, lindner.s10, B = 1000, main = "Ten
+   Logit-Based Strata")
```

```
$balance.orig
```

```
[1] 1.638961
```

```
$rank.orig
```

```
[1] 1
```

```
R> bal.ms.psa(ejecfrac, abcix, lindner.tree, B = 1000, main = "Six
+   Tree-Based Strata")
```

```
$balance.orig
```

```
[1] 1.989346
```

```
$rank.orig
```

```
[1] 439
```

The statistics  $\hat{\delta}_\alpha$  generated in the five, ten and tree-derived six strata cases are not directly comparable (see Figure 3). However in both the five and ten strata cases,  $\hat{\delta}_\alpha$  is lower than the entire randomly generated comparison distribution. In the tree-derived strata case,  $\hat{\delta}_\alpha$  is toward the middle (439 of 1001) of the comparison distribution. This indicates that the logit-based score stratifications balance the `ejecfrac` covariate better than the tree-derived strata.

## 5. Balance graphics for categorical covariates

The function `cat.psa` can be used to depict graphically the balance across treatments of a categorical covariate. Side-by-side segmented bars comparing proportions of cases in each category are constructed vertically for treatment groups within each stratum. As in the `box.psa` function, the pairs of segmented bars for each stratum are graphed on the same set of axes. This is intended to facilitate the ready comparison of a covariate's distributions not simply within strata, but across strata as well. For instance, if subsequent analysis indicates large differences in size or direction of effects for different strata, then comparing covariate distributions across strata may give an initial indication of potential causes (see Figure 4).

```
R> cat.psa(stent, abcix, lindner.s5, xlab = "Strata",
+   ylab = "Proportion for 'stent'", catnames = c("No Stent", "Stent"),
+   barnames = c("Standard PCI", "Abciximab"), rtmar = 2)
```

	C	T	C	T	C	T	C	T	C	T
Levels	1.000	1.000	2.000	2.00	3.000	3.000	4.000	4.000	5.00	5.000
0	0.851	0.771	0.263	0.22	0.138	0.169	0.233	0.269	0.25	0.185
1	0.149	0.229	0.737	0.78	0.862	0.831	0.767	0.731	0.75	0.815



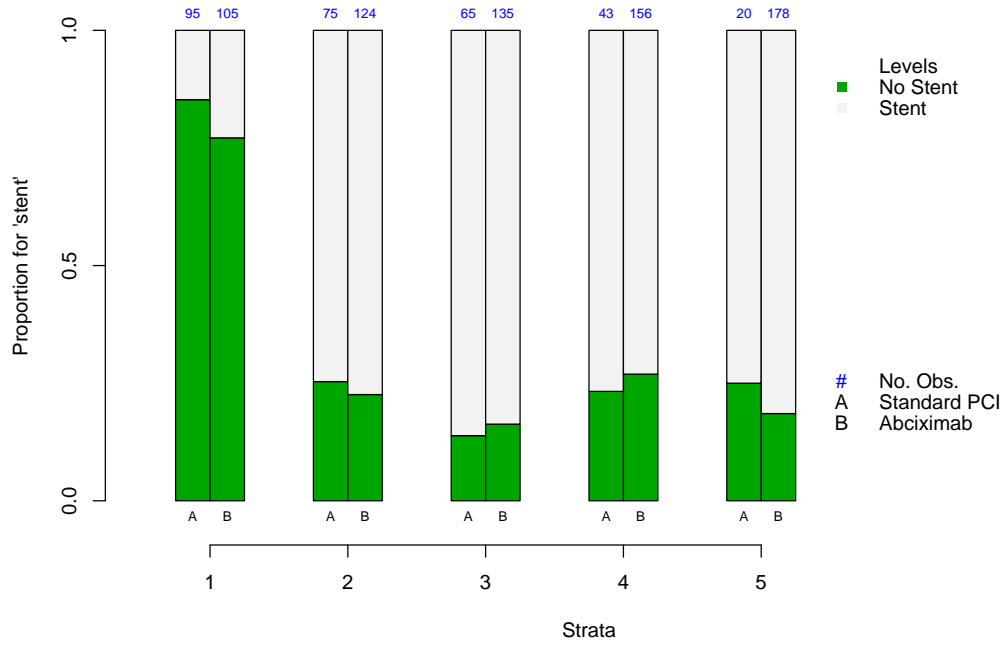


Figure 4: Side by side barplots for covariate `stent`, 5 strata, created using `cat.psa`.

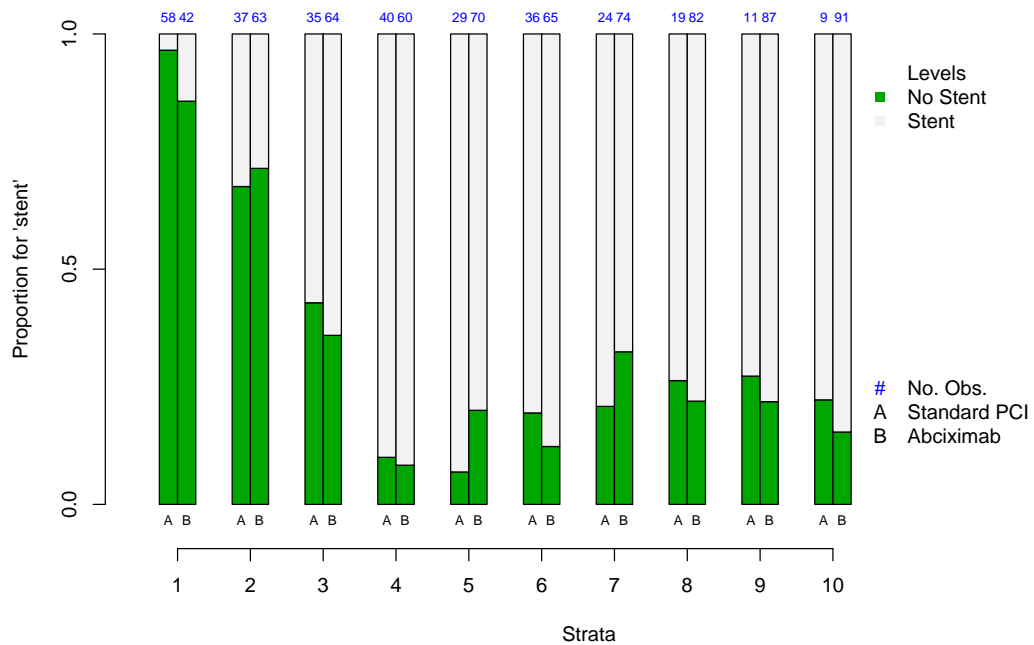


Figure 5: Side by side barplots for covariate `stent`, 10 strata, created using `cat.psa`.

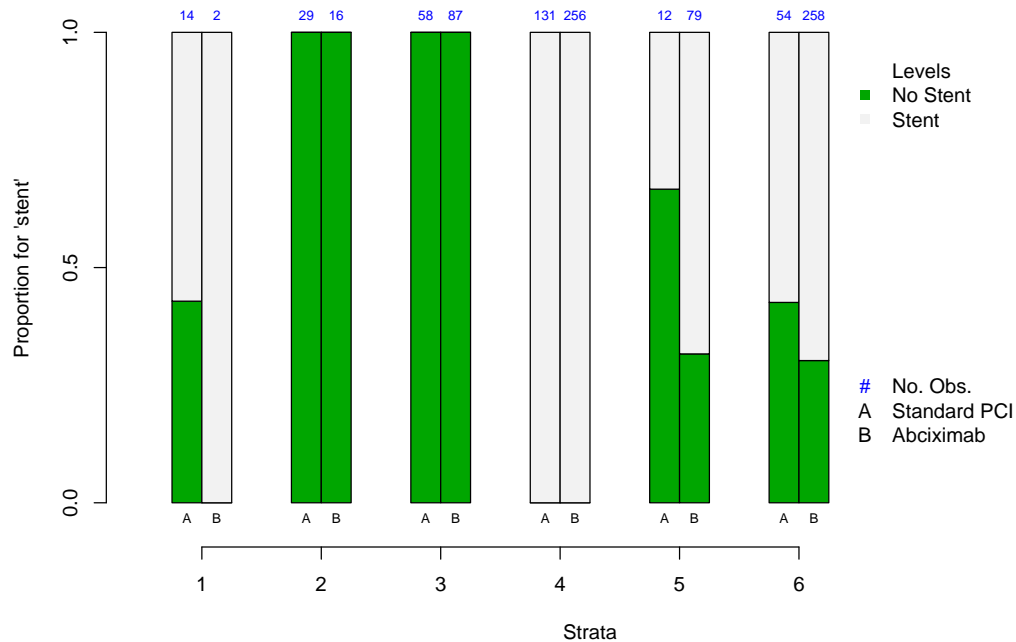


Figure 6: Side by side barplots for covariate `stent`, 6 tree-derived strata, created using `cat.psa`.

The sizes of treatment groups within strata are printed (in blue) at the top of the bars. The proportions used for the categorical plot are output as a table. In the case of the `stent` covariate it is seen that the strata are reasonably well balanced. However the first stratum has a much larger proportion of patients who did not receive a stent than the other four strata. Recall that for the set of 5 strata, the `ejecfrac` covariate in the first stratum was not substantially different from the other strata. The issue is further illuminated by the graph based on 10 strata, in that the proportion of patients without stents declines throughout the first several strata - that is, as the estimated propensity for receiving treatment with abciximab increases (see Figure 5). Numerical output has been omitted.

```
R> cat.psa(stent, abcix, lindner.s10, xlab = "Strata",
+   ylab = "Proportion for 'stent'", catnames = c("No Stent", "Stent"),
+   barnames = c("Standard PCI", "Abciximab"), rtmar = 2.8)
```

The tree-derived strata show much less homogenization for the `stent` covariate, though much of the difference appears in the smaller strata (see Figure 6). Again, numerical output has been omitted.

```
R> cat.psa(stent, abcix, lindner.tree, xlab = "Strata",
+   ylab = "Proportion for 'stent'", catnames = c("No Stent", "Stent"),
+   barnames = c("Standard PCI", "Abciximab"), rtmar = 2)
```

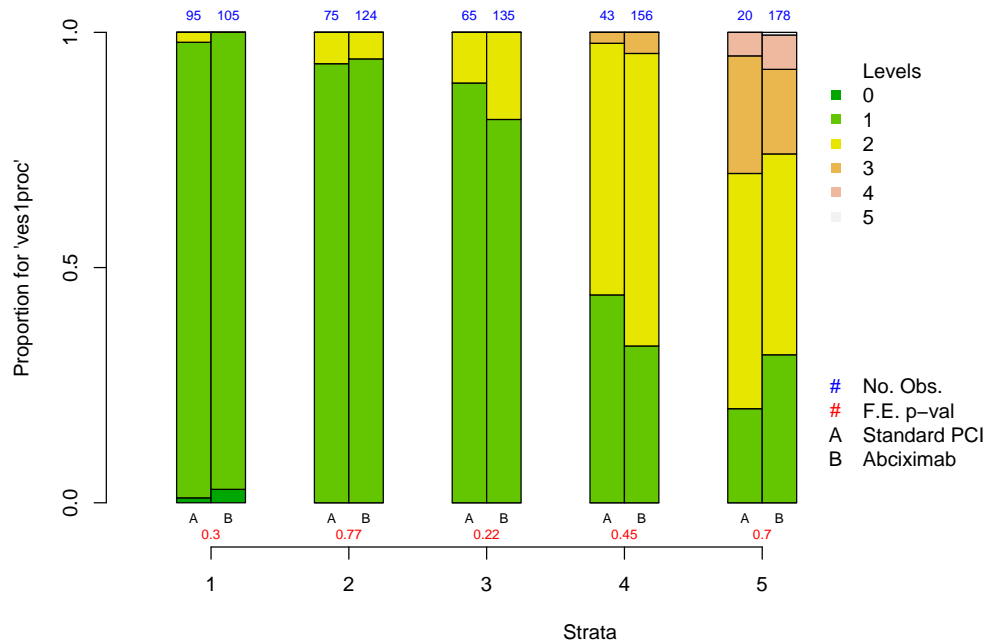


Figure 7: Side by side barplots for covariate `ves1proc`, 5 strata, created using `cat.psa`.

Continuing with the categorical covariate `ves1proc` and using five logit-based strata, we see a pattern of patients with higher numbers of vessels involved in the initial PCI tending to be in higher strata (see Figure 7).

```
R> cat.psa(ves1proc, abcix, lindner.s5, xlab = "Strata",
+   ylab = "Proportion for 'ves1proc'", catnames = 0:5,
+   barnames = c("Standard PCI", "Abciximab"),
+   rtmar = 2, balance = TRUE, B = 1000)
```

Histogram of Random Strata Balance. Press <enter> for next chart...

```
$`treatment:stratum.proportions`
  0:1  1:1  0:2  1:2  0:3  1:3  0:4  1:4  0:5  1:5
0 0.011 0.029 0.000 0.000 0.000 0.000 0.000 0.000 0.00 0.000
1 0.968 0.971 0.933 0.944 0.892 0.815 0.442 0.333 0.20 0.315
2 0.021 0.000 0.067 0.056 0.108 0.185 0.535 0.622 0.50 0.427
3 0.000 0.000 0.000 0.000 0.000 0.000 0.023 0.045 0.25 0.180
4 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.05 0.073
5 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.00 0.006
```

The inclusion of the logical `balance = TRUE` in `cat.psa` makes a call to two balance functions, similar in spirit to the `bal.ks.psa` and `bal.ms.psa` functions used for continuous covariates.

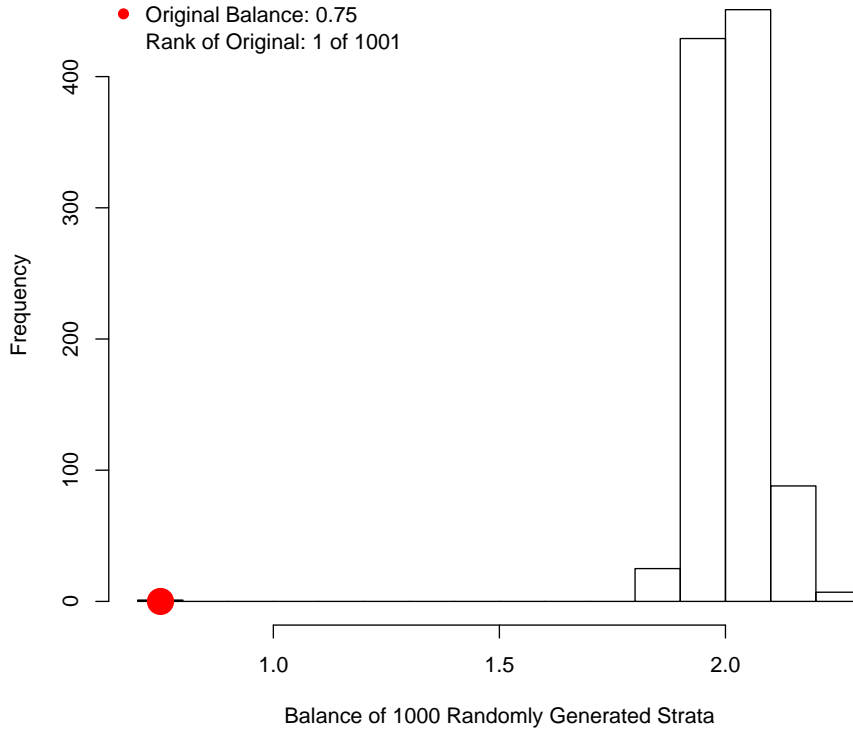


Figure 8: Balance statistic  $\hat{\gamma}_\alpha$  for covariate `ves1proc`, 5 strata, created using the balance option of `cat.psa`.

First, `bal.fe.psa` is used to measure the balance achieved within the defined strata by obtaining a  $p$  value using Fisher’s exact test for each strata. This  $p$  value is printed below each pair of barplots. In Figure 6, they agree with a visual impression of rough comparability of control and treatment distributions for each stratum.

The second balance function `bal.cs.psa` (the analogue of `bal.ms.psa`) measures the balance achieved across strata using a statistic  $\hat{\gamma}_\alpha$ . As before, let  $\alpha$  be a stratification of cases into  $K$  strata. Let the categorical covariate of interest have  $J$  categories. Define  $p_{ijk}$  to be the proportion of cases in stratum  $k$ ,  $k = 1, \dots, K$ , category  $j$ ,  $j = 1, \dots, J$ , and treatment  $i$ ,  $i = 0, 1$ . Then

$$\hat{\gamma}_\alpha = \sum_{k=1}^K \sum_{j=1}^J |(p_{0jk} + \epsilon)^{\frac{1}{2}} - (p_{1jk} + \epsilon)^{\frac{1}{2}}|.$$

Note that for a fixed category disparity  $\hat{\Delta} = |p_{0jk} - p_{1jk}|$ , the contribution to  $\hat{\gamma}_\alpha$  from the  $jk$ th category will be inversely related to the size of the category. The function of adding a small constant  $\epsilon$  is to help ensure when the  $p_{ijk}$  are especially small that this weight is still reasonable (we have found a value of  $\epsilon = 0.02$  seems to work well). As in the continuous

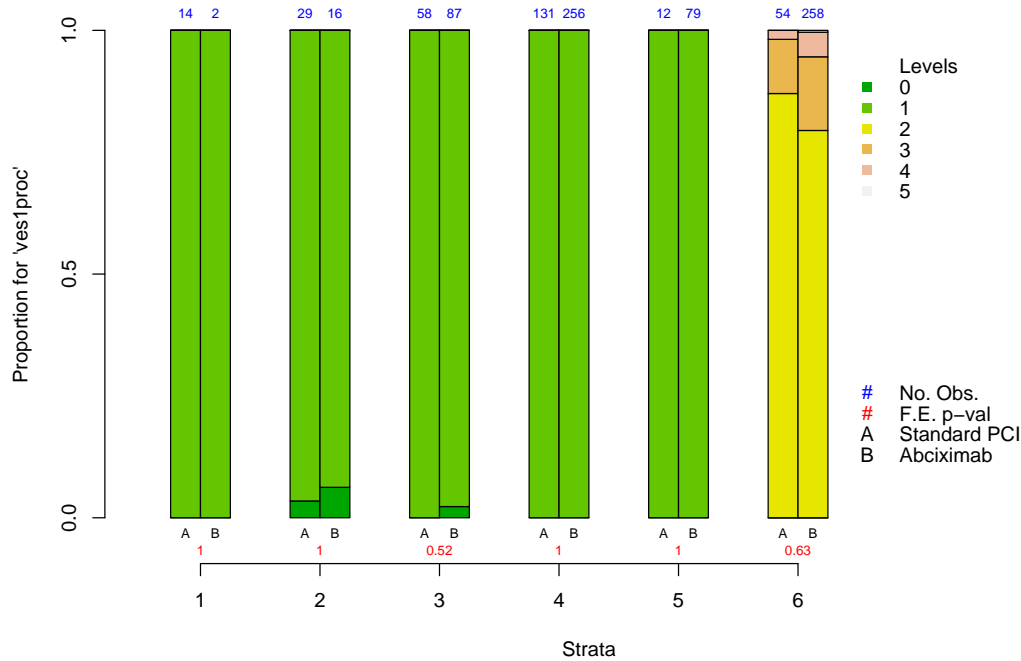


Figure 9: Barplots for covariate `ves1proc`, 6 strata, created using of `cat.psa`.

covariate case, if  $\alpha$  represents a particular assignment of cases to strata, then `bal.cs.psa` generates a permutation distribution by randomly assigning cases to strata, thus generating assignments  $\alpha_b$ ,  $b = 1, \dots, B$ , and the associated permutation statistics  $\hat{\gamma}_{\alpha_b}$ . Again, the  $\alpha_b$  are generated under a fixed marginals model to retain comparability with  $\alpha$ ; the distribution of  $\hat{\gamma}_{\alpha_b}$  is plotted as a histogram along with the particular  $\hat{\gamma}_{\alpha}$  referenced as a red dot (see Figure 8).

For comparison purposes, consider the tree-derived strata (see Figure 9).

```
R> cat.psa(ves1proc, abcix, lindner.tree, xlab = "Strata",
+   ylab = "Proportion for 'ves1proc'", catnames = 0:5,
+   barnames = c("Standard PCI", "Abciximab"),
+   rtmar = 2, balance = TRUE, B = 1000)
```

Histogram of Random Strata Balance. Press <enter> for next chart...

```
$`treatment:stratum.proportions`
```

	0:1	1:1	0:2	1:2	0:3	1:3	0:4	1:4	0:5	1:5	0:6	1:6
0	0	0	0.034	0.062	0	0.023	0	0	0	0	0.000	0.000
1	1	1	0.966	0.938	1	0.977	1	1	1	1	0.000	0.000
2	0	0	0.000	0.000	0	0.000	0	0	0	0	0.870	0.795
3	0	0	0.000	0.000	0	0.000	0	0	0	0	0.111	0.151
4	0	0	0.000	0.000	0	0.000	0	0	0	0	0.019	0.050
5	0	0	0.000	0.000	0	0.000	0	0	0	0	0.000	0.004

The `rpart` tree has placed all patients who had two or more vessels involved in the original PCI into one stratum. Given the imbalance in the sizes of the two treatment groups, with patients who received additional treatment with abciximab much more numerous in this group, it appears the `ves1proc` covariate may well be used as a primary indicator of necessity of treatment with the new drug. The graph highlights the issue associated with an unadjusted comparison between the two treatment groups. It clearly leads one to be interested in the differences in outcomes for each of the groups within the specific strata, not simply across-the-board averages of the differences.

## 6. Outcome analysis: `circ.psa`

Having examined the covariate differences from stratum to stratum, it is of major interest to compare outcome measures for the treatment groups across strata. By default the mean (possibly trimmed using the `trim` argument) is used, but any univariate statistic may be used as outcome measure via the `statistic` argument. If the direction and size of the effects are similar across strata, then the adjustment provided by PSA is relatively simple to interpret. However, when differences in outcomes between treatments vary across strata, the investigator may want to learn how these differences are related to changes in covariate distributions in the strata. In the case of the Lindner data

```
R> circ.psa(log(cardbill), abcix, lindner.s5, revc = TRUE,
+   xlab = "Treatment with Abciximab", ylab = "Standard Treatment")
```

produces the output below, and the graph in Figure 10.

```
$summary.strata
```

	n.0	n.1	means.0	means.1
1	95	105	9.370104	9.474068
2	75	124	9.338528	9.589840
3	65	135	9.377147	9.577885
4	43	156	9.471588	9.638329
5	20	178	9.665442	9.592309

```
$`1.wtd.Mn`
```

```
[1] 9.574371
```

```
$`0.wtd.Mn`
```

```
[1] 9.444198
```

```
$dae
```

```
[1] -0.1301735
```

```
$se.wtd
```

```
[1] 0.04119358
```

```
$approx.t
```

```
[1] -3.160044
```

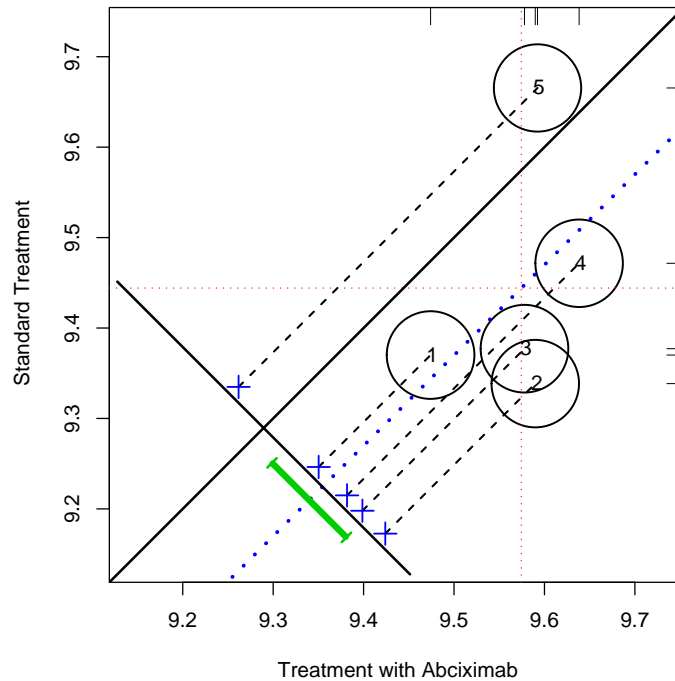


Figure 10: PSA assessment plot of  $\log(\text{cardbill})$ , 5 strata, constructed using `circ.psa`.

```
$df
```

```
[1] 986
```

```
$CI.95
```

```
[1] -0.21101070 -0.04933637
```

Circles are plotted and labeled for each stratum. Sizes of circles vary according to the sizes of the strata. Thus in the case of `lindner.s5` all circles are about the same size. The center of each stratum's circle corresponds to outcome means for the respective control and treatment groups for that stratum. Thus circles below the heavy black  $Y = X$  identity line indicate that for the first stratum (labeled 1) the outcome mean is larger than the control group mean. Indeed, the same thing occurred for four of five strata. Circles above the identity line indicate the reverse. The heavy (blue) dashed line, parallel to the  $Y = X$  identity indicates the (weighted) average of these outcome differences, that is, the overall adjusted outcome. In Chapter 2 of [Rosenbaum \(2002\)](#) this is defined as the *direct adjustment estimator* (DAE) of treatment effect. Centers of circles are projected parallel to the identity line, to the lower left line segment where the crosses show the distribution of strata differences; the heavy dashed blue line corresponds to the mean of this distribution of differences. The vertical and horizontal dashed red lines represent the (weighted) means for the control and treatment groups respectively. The heavy dashed line always passes through the point where control

and treatment group means cross. Rug plots along the sides of the graph depict marginal distributions of control and treatment outcome measures.

The output includes a summary by strata of size and outcome means, as well as the weighted mean for each treatment group, the adjusted difference in outcome between the two treatment groups, and the weighted standard error of the strata estimates for this difference. The weighted standard error follows [Conniffe, Gash, and O'Connell \(2000\)](#). An approximate  $t$  statistic is computed as the ratio of the of adjusted difference to the weighted standard error; a conservative estimate of the degrees of freedom is also given. In addition, the `circ.psa` function produces a 95% confidence interval for the population DAE, shown as a (heavy green) line segment to the left of the distribution of outcome mean differences for the respective strata, centered on the sample DAE line segment (see below). The data input for `circ.psa` can take two forms: viz., three columns that provide responses, treatment coding and strata for all individuals; or summary statistics in the format of `summary.strata`. This is useful for analysis of strata that have been defined using the function `loess.psa` (see below).

Consider the circle plot given in [Figure 10](#). Four of the five strata lie below the identity line in close proximity to one another, indicating concordance of outcome measures in these strata. However the fifth stratum appears in the upper right region of the plane, rather far from the other strata. The abciximab treatment distribution of `log(cardbill)` is shown in the rug plot along the top margin. This information illuminates key features of the results for this outcome measure. Recall that the number of stents was not notably different from the second, third and fourth strata (see [Figure 4](#)), but the number of vessels involved in the initial PCI was larger for the fifth stratum than for the the others. Investigation across other covariates reveals that all are reasonably balanced, and the distribution of covariates in the fifth stratum is not markedly different from the other strata except for `acutemi`. About half of the patients in the fifth stratum had an acute myocardial infarction within the week before their PCI, while the proportion in other strata was negligible. The fifth stratum appears to isolate patients with some of the more severe problems.

For 10 strata, we have: (see [Figure 11](#))

```
R> circ.psa(log(cardbill), abcix, lindner.s10, revc = TRUE,
+   ne = 0.1, sw = 0.15, inc = 0.17, labcex = 0.9,
+   xlab = "Treatment with Abciximab", ylab = "Standard Treatment")
```

```
$summary.strata
  n.0 n.1 means.0 means.1
1   58  42 9.344224 9.400807
2   37  63 9.410674 9.522909
3   35  64 9.494560 9.585320
4   40  60 9.202001 9.594662
5   29  70 9.403091 9.523627
6   36  65 9.356248 9.636317
7   24  74 9.509425 9.656899
8   19  82 9.423793 9.621572
9   11  87 9.528079 9.550337
10   9  91 9.833331 9.632437
```



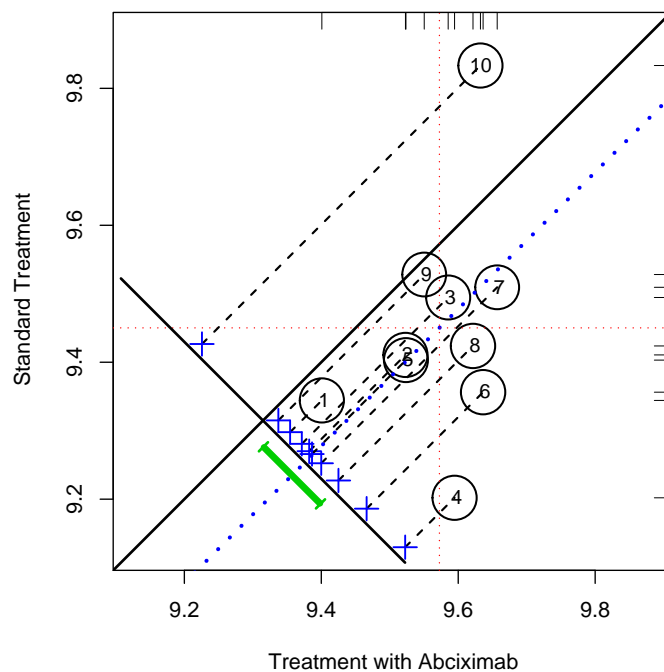


Figure 11: PSA assessment plot of  $\log(\text{cardbill})$ , 10 strata, constructed using `circ.psa`.

```
$`1.wtd.Mn`
[1] 9.572513
```

```
$`0.wtd.Mn`
[1] 9.45015
```

```
$dae
[1] -0.1223628
```

```
$se.wtd
[1] 0.04151743
```

```
$approx.t
[1] -2.947263
```

```
$df
[1] 976
```

```
$CI.95
[1] -0.20383648 -0.04088907
```

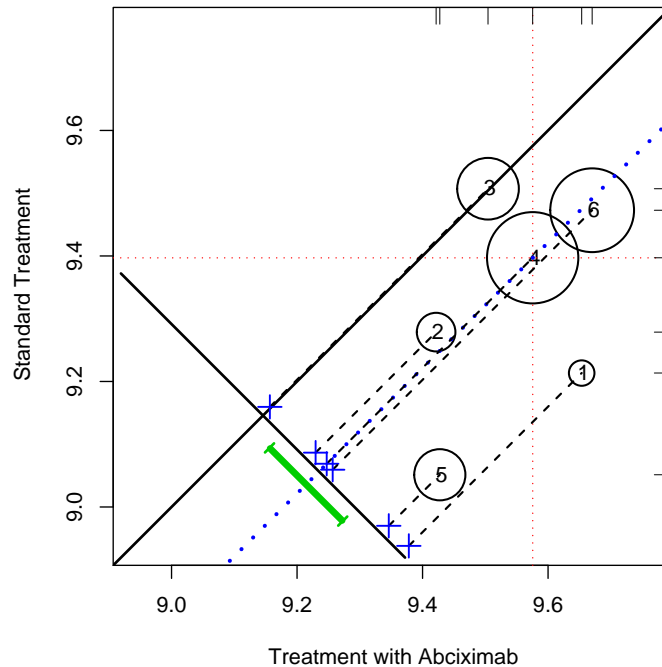


Figure 12: PSA assessment plot of  $\log(\text{cardbill})$ , 6 strata, constructed using `circ.psa`.

All strata except the tenth are uniform in direction of outcome, though stratum four shows a larger effect than the other strata. Though the effect direction is different in the 10th stratum, it is based on only nine patients in the control group.

Now consider the tree-derived strata (see Figure 12).

```
R> circ.psa(log(cardbill), abcix, lindner.tree, ne = .15, sw = 0.3, rev = TRUE,
+   xlab = "Treatment with Abciximab", ylab = "Standard Treatment")
```

```
$summary.strata
```

	n.0	n.1	means.0	means.1
1	14	2	9.213456	9.653571
2	29	16	9.278627	9.421610
3	58	87	9.507137	9.504301
4	131	256	9.397094	9.575469
5	12	79	9.051331	9.427481
6	54	258	9.472841	9.670249

```
$`1.wtd.Mn`
[1] 9.57558
```

```
$`0.wtd.Mn`
[1] 9.39695
```

```
$dae
[1] -0.1786314
```

```
$se.wtd
[1] 0.05838647
```

```
$approx.t
[1] -3.059466
```

```
$df
[1] 984
```

```
$CI.95
[1] -0.29320771 -0.06405511
```

While in the first two instances all strata were about the same size, the tree-based strata have varying sizes which is reflected in the area of the circles. While the third, medium sized stratum shows no difference in  $\log(\text{cardbill})$ , the rest of strata all are below the identity line indicating concordance of effect direction. The estimated effect here is larger than for either the 5 or 10 strata cases. The first and fifth strata are somewhat different than the rest, and would perhaps bear investigation.

## 7. Outcome analysis: `loess.psa`

Another possibility for assessing the outcome effect is to use a loess plot.<sup>2</sup> The loess concept is to plot the bivariate distributions showing estimated propensity scores versus outcomes for each of the treatment groups (using different symbols for each) and to overlay the point-plots with smoothed regression curves showing how outcomes are predicted by the propensity scores for the groups being compared. By averaging the (vertical) differences between the loess regression curves across the range of propensity scores, in principle an overall adjusted effect estimate can be produced without reference to strata (see Figure 13); however, strata in general can still be useful. In particular, if the user specifies a given number of strata that are used to calculate the DAE (possibly in several ways, to get comparisons), then variances of outcomes can be computed to serve as a basis for forming an approximate test statistic. We compute the weighted standard error in this way, again using methods of [Conniffe \*et al.\* \(2000\)](#); and from this we obtain both the approximate  $t$  statistic and a 95% confidence interval for the loess graphic.

```
R> loess.psa(log(cardbill), abcix, ps, ylab = "Response: log(cardbill)")
```

```
$dae
[1] 0.1249827
```

---

<sup>2</sup>The plot was first suggested by Obenchain [Obenchain \(2007\)](#), who implements this plot in his R package `USPS`, although his function and ours are not equivalent.

```
$se.wtd
```

```
[1] 0.03919522
```

```
$CI95
```

```
[1] 0.04659229 0.20337319
```

```
$summary.strata
```

	counts.0	counts.1	means.0	means.0	diff.means
1	43	25	9.320024	9.438872	0.11884817
2	28	37	9.399415	9.486414	0.08699918
3	24	43	9.408554	9.508797	0.10024378
4	21	45	9.399970	9.533743	0.13377315
5	33	41	9.372128	9.562461	0.19033363
6	21	38	9.361383	9.578040	0.21665629
7	19	48	9.367159	9.591070	0.22391164
8	26	40	9.377707	9.598483	0.22077634
9	20	47	9.391453	9.606396	0.21494334
10	13	52	9.415172	9.616330	0.20115833
11	21	45	9.446956	9.616071	0.16911564
12	9	59	9.488429	9.610938	0.12250840
13	8	57	9.536530	9.603450	0.06692042
14	7	59	9.644657	9.601478	-0.04317965
15	5	62	9.742843	9.597939	-0.14490405

By default 15 strata of approximately the same size number of observations are used in the computation of the adjusted effect estimation. Again, by default only strata with at least 5 observations in each treatment are used; other strata are ignored. Internally, the function compares effect size estimates for 10, 15 and 25 strata and warns the user if estimates vary unreasonably. Rug plots of the distributions of propensity scores and outcomes for the two treatments are placed at the sides of the graph. The `int` option may be used to define the number of strata used in the effect size estimate, or to define strata based upon the the loess estimates. For instance, if the overlap of the two treatment groups at the extremes of the propensity scores is not adequate, these propensity scores may be eliminated. If loess estimates for the two treatments cross or are otherwise distinct we may wish to define strata accordingly. Given the plot in Figure 13, we might be interested in the effect size for strata beginning above the few low propensity scores, say at .375, up to where the vertical difference narrows at 0.55, and then to where the loess line cross at about 0.875. Then `int = c(0.375, 0.55, 0.875, 1)` would produce an DAE based upon propensity scores in the three strata (0.375, 0.55], (0.55, 0.875], and (0.875, 1]. The confidence intervals (and  $t$  statistics) are similar, but not identical, for the `circ.psa` and `loess.psa` results. Differences of this kind are to be expected, given the different ways of defining strata for these two functions. Finally, note that the first four columns of `strata.summary` may be graphed using `circ.psa` with the option `summary = TRUE`, however in that case the  $t$  statistic and confidence interval will not be computed.

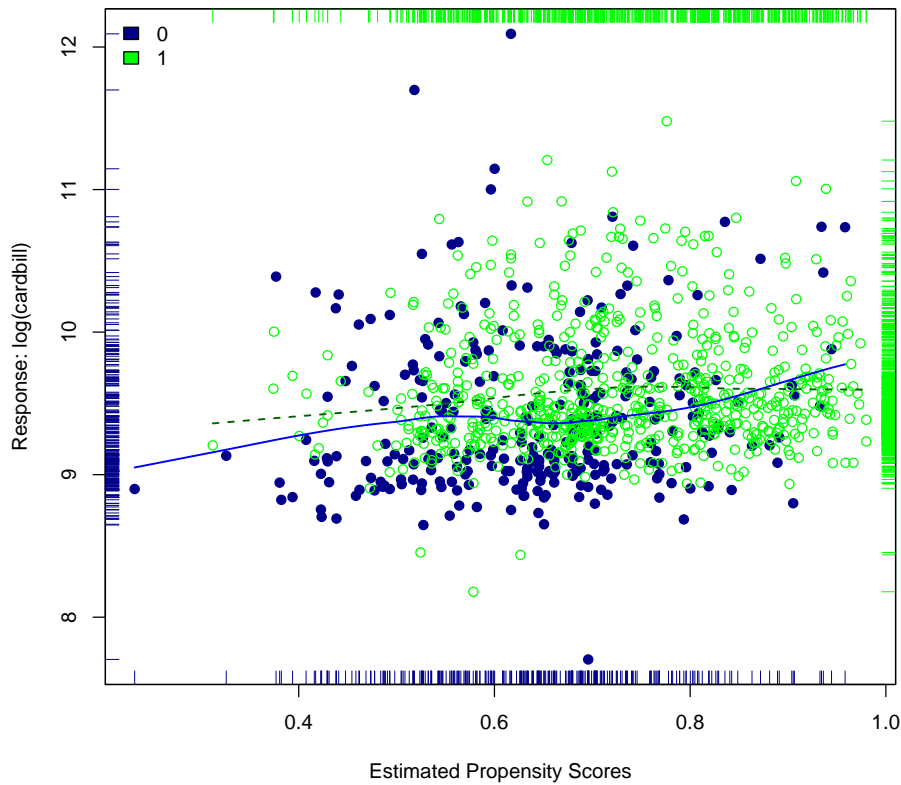


Figure 13: Loess Plot of  $\log(\text{cardbill})$  constructed using `loess.psa`.

## 8. Conclusion

The **PSAgraphics** package has been designed mainly to address questions that arise in observational studies, where propensity scores have been constructed using covariates. The functions have been constructed to be as flexible as possible, in several respects. For example, in case the investigator has defined either propensity scores or strata in two or more ways, the functions provide both numerical and graphical supports for comparisons of covariate balance with respect to those different specifications.

Furthermore, the package can be used for experimental data analysis. For example, when covariate information is available at the termination of a true experiment, one may seek to learn whether randomization has worked as intended to balance extant covariates. Adjusted effect estimators can then be applied in the experimental data analysis in a fashion that parallels methods used for observational data. The two functions, `box.psa` and `cat.psa` aim to facilitate studies of covariate balance, both within and across propensity-defined strata. To ease interpretations, all functions will by default label graphical objects and output value structures in terms of names or characters taken from the data frame or factors to which they are applied.

In the second phase of a PSA study, the DAE is usually assessed numerically using defined strata. Historically, graphical displays of effects are less common than numerical ones, perhaps because effective graphics have not been widely available in the past. The two functions `circ.psa` and `loess.psa` aim to improve this situation, where the first focuses on effect estimation using predefined strata, possibly of an ordered categorical nature as from a classification tree, and the second assumes continuously distributed propensity score estimates and does not require predefined strata. Both functions facilitate visual display of outcomes with respect to estimated propensity scores and both can help to learn how or how much effects differ across the range (or categories) of estimated propensity scores.

## References

- Cochran WG (1968). “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies.” *Biometrics*, **24**, 295–314.
- Conniffe D, Gash V, O’Connell PJ (2000). “Evaluating State Programmes: ‘Natural Experiments’ and Propensity Scores.” *The Economic and Social Review*, **31**(4), 283–308.
- Obenchain RL (2007). *USPS: Unsupervised and Supervised Methods of Propensity Score Adjustment for Bias*. R package version 1.1-0, URL <http://CRAN.R-project.org/package=USPS>.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rosenbaum PR (2002). *Observational Studies*. 2nd edition. Springer-Verlag, New York.
- Rosenbaum PR, Rubin DB (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, **70**(1), 41–55.
- Sarkar D (2008). *lattice: Multivariate Data Visualization with R*. Springer-Verlag, New York.
- Therneau TM, Atkinson B (2002). *rpart: Recursive partitioning and regression trees*. R package version 3.1-42, URL <http://CRAN.R-project.org/package=rpart>.

### Affiliation:

James E. Helmreich  
Department of Mathematics  
Marist College  
Poughkeepsie, NY 12601, United States of America  
E-mail: [James.Helmreich@Marist.edu](mailto:James.Helmreich@Marist.edu)  
URL: <http://foxweb.marist.edu/users/james.helmreich/>

Robert M. Pruzek  
Department of Educational and Counseling Psychology  
1400 Washington Ave.  
State University of New York at Albany  
Albany, NY 12222, United States of America  
E-mails: [rmpruzek@yahoo.com](mailto:rmpruzek@yahoo.com)  
URL: <http://rmpruzek.com/>