

# Evaluating hospital performance

## **RiskStandard** package

Arnout Van Messem and Machteld Varewyck  
Ghent University

February 26, 2015

## Introduction

This R-package implements statistical methods for benchmarking clinical care centers based on a binary quality indicator such as 30-day mortality. For each center we provide directly or indirectly standardized risks based on fixed center effects outcome models that incorporate patient-specific baseline covariates to adjust for differential case-mix. The user can choose to apply the Firth correction (Firth, 1993) in the outcome model to maintain convergence in the presence of very small centers.

The package includes three example datasets: `smallCaseMix`, with small differences in patient mix across centers, `largeCaseMix` with large differences in patient mix across centers, and `largeCaseMix_missing` which is based on `largeCaseMix` but where the consciousness level is missing for some patients. Input data must contain for each patient (1) patient-specific covariates to adjust for in the analysis e.g. age, baseline severity, (2) a hospital code where the patient was treated and (3) a binary quality outcome e.g. 30-day mortality. In this document we will illustrate how center performance can be assessed using the implemented R-functions.

We refer the user to Varewyck et al. (2014) for the theory behind the implemented R-functions.

## Use in R

In this section we illustrate the two summary functions and three plot functions of the package. First, install and load the package ‘RiskStandard’:

```
> install.packages("./RiskStandard_0.0.5.tar.gz",  
+                  repos = NULL, type = "source")  
>  
  
> library(RiskStandard)
```

The dataset `largeCaseMix` contains for the  $n = 50\,000$  patients treated in one of the  $m = 50$  centers:

- patient-specific covariates: age (continuous), sex (binary) and consciousness level at admission (1=alert, 2=drowsy, 3=unconscious)
- hospital code: center (1 to  $m$ )
- binary quality outcome: outcome (0=alive, 1=dead)

Before assessing center performance, we recommend to make some descriptives of the dataset to get an impression of the distribution of patient characteristics across centers (Figure 1).

```
> str(largeCaseMix)

'data.frame':      50000 obs. of  5 variables:
 $ age      : int  65 74 64 69 76 87 61 75 51 71 ...
 $ sex      : int   0 0 0 1 0 0 0 1 1 1 ...
 $ cons     : Factor w/ 3 levels "1","2","3": 1 1 1 1 2 1 1 1 1 1 ...
 $ center   : int   25 50 24 1 25 24 16 24 50 32 ...
 $ outcome  : int   0 0 1 0 1 0 0 0 0 0 ...

> m <- length(unique(largeCaseMix$center))
> n <- dim(largeCaseMix)[1]
> centerSize <- as.vector(table(largeCaseMix$center))

> layout(matrix(1:4))
> # Age
> with(largeCaseMix,
+      plot(1:m, tapply(age, center, mean), pch = 19,
+            cex = centerSize/n*m, cex.lab = 1.2, ylim = c(60,80),
+            xlab = "Center", ylab = "", main = "Mean age per center"))
> with(largeCaseMix,
+      boxplot(age ~ center, xlab = "Center", ylab = "Age distribution",
+              cex.lab = 1.2))
> # Sex
> with(largeCaseMix,
+      plot(1:m, tapply(sex, center, mean), pch = 19, cex=centerSize/n*m,
+            cex.lab = 1.2, ylim = c(0,1), xlab = 'Center', ylab = "",
+            main= "Percentage women per center"))
> # Consciousness
> with(largeCaseMix,
+      plot(1:m, tapply((cons==1), center, mean), pch = 21,
+            cex = centerSize/n*m, cex.lab = 1.2, ylim = c(0,1.1), xlab='Center',
+            ylab = "", main="Distribution of consciousness level per center"))
> with(largeCaseMix,
+      points(c(1:m), tapply((cons %in% c(1,2)), center, mean), pch = 19,
+              cex = centerSize/n*m))
> legend("bottomleft", pch = c(21,19), bty='n',
+        legend = c("Alert", "Alert or drowsy"), cex=1.2)
```

## standardizeRisks()

This function estimates the standardized mortality risks. Necessary parameters are

- patientCovariates: data frame of patient-specific covariates. Please make sure that categorical covariates are passed as factor, otherwise a linear effect will be assumed in the fitted outcome model ( $p$  covariate values for each of the  $n$  patients, giving an  $n \times p$  data frame)
- center: hospital code (1 value out of  $m$  for each of the  $n$  patients)
- Y: binary quality outcome (0 or 1 for each of the  $n$  patients)

The input for the parameter 'center' can be a character vector with the hospital names which will automatically be used in the output. The other function parameters have default values, but can be changed by the user (see R-documentation). By default a Firth corrected outcome regression model is fitted. By default no summary of the fitted model is printed, but it can be asked by adding the argument `trace = TRUE`. This summary can be useful for checking whether each of the covariates was passed in the correct format.

```
> indirectRisks <- standardizeRisks(
+   patientCovariates = largeCaseMix[,c('age', 'sex', 'cons')],
+   center = largeCaseMix[, 'center'],
+   Y = largeCaseMix[, 'outcome'])
> head(indirectRisks)
```

	centerName	centerSize	standardizedRisk	varStandardizedRisk	lowerCI
1	1	386	-0.023367272	0.0003162722	-0.05822335
2	2	61	-0.007509684	0.0019252936	-0.09350931
3	3	984	-0.016308775	0.0001421959	-0.03968055
4	4	448	-0.075252885	0.0002203354	-0.10434600
5	5	2111	0.033703502	0.0001061479	0.01351036
6	6	141	-0.032481875	0.0006909590	-0.08400168

	upperCI	observedRisk
1	0.011488811	0.1658031
2	0.078489944	0.1803279
3	0.007062996	0.1636179
4	-0.046159770	0.1093750
5	0.053896643	0.2676457
6	0.019037931	0.1347518

Similarly for direct standardization:

```
> directRisks <- standardizeRisks(
+   patientCovariates = largeCaseMix[,c('age', 'sex', 'cons')],
+   center = largeCaseMix[, 'center'],
+   Y = largeCaseMix[, 'outcome'],
+   method='direct')
```

When some patients have missing values for a categorical patient covariate such as consciousness level, we offer two ways to handle the missingness. The option `missing='completeCase'` (default) performs a complete case analysis, excluding all patients who have missing consciousness. The option `missing='dummyCategory'` adds a separate category to the fitted outcome model, allowing for a missing value effect. When some patients have missing values for a continuous patient covariate such as age, the function will perform a complete case analysis, excluding all patients who have missing age. Alternatively, multiple imputation can be considered to handle missingness. Although this method is currently not implemented in the `standardizeRisks()` function, the user can pass each of the imputed datasets separately to the function and afterwards average the estimated standardized risks over the different imputations. The variance on the standardized risks can then be obtained by combining the within and between imputation variance as explained in Schafer (1999). The number of observations ( $n$ ) that was used for the analysis can be extracted as an attribute from the function.

```
> indirectRisks2 <-
+   standardizeRisks(patientCovariates = largeCaseMix_missing[,c('age', 'sex', 'cons')],
+   center = largeCaseMix_missing[, 'center'],
+   Y = largeCaseMix_missing[, 'outcome'],
+   method='indirect', missing='completeCase')
> attr(indirectRisks2, "n")
```

## labelCenters()

The output from the `standardizeRisks()` function can then be used to classify the centers as having ‘low’, ‘accepted’ or ‘high’ mortality risk.

```
> labeledCenters <- labelCenters(standardizedRisks = indirectRisks)
> head(labeledCenters)
```

	centerName	centerLabel	lowerCI	upperCI
1	1	A	-0.03536243	-0.011372117
2	2	A	-0.03710506	0.022085690
3	3	A	-0.02435179	-0.008265759
4	4	L	-0.08526481	-0.065240962
5	5	A	0.02675436	0.040652643
6	6	A	-0.05021158	-0.014752171

By default for indirect standardization, centers are classified as having low mortality if the (upper bound of the) 50% confidence interval on the standardized risk is smaller than a consensus value. This clinically relevant boundary is by default set at  $-0.05$ . Analogously, centers are classified as having high mortality if the (lower bound of the) 50% confidence interval on the standardized risk is larger than the clinically relevant boundary of  $0.05$ . For direct standardization the consensus value is by default  $0.8 \hat{E}(Y)$  for low mortality and  $1.2 \hat{E}(Y)$  for high mortality. Note that the function `labelCenters()` returns by default 50% confidence intervals in the output, while the function `standardizeRisks()` gives 95% confidence intervals by default.

The clinically relevant boundaries can be adapted by providing specific values for the parameter `lambda` in the `labelCenters()` function. When `lambda` is a vector of two elements, the first factor determines the consensus value before classifying centers as having low mortality risk, while for high mortality it is the second factor. For example, we can implement a consensus value of  $-0.06$  for low mortality and  $0.02$  for high mortality as follows:

```
> labeledCenters2 <- labelCenters(standardizedRisks = indirectRisks,
+                                lambda=c(low = 0.06, high = 0.02))
> head(labeledCenters2)
```

	centerName	centerLabel	lowerCI	upperCI
1	1	A	-0.03536243	-0.011372117
2	2	A	-0.03710506	0.022085690
3	3	A	-0.02435179	-0.008265759
4	4	L	-0.08526481	-0.065240962
5	5	H	0.02675436	0.040652643
6	6	A	-0.05021158	-0.014752171

## plotRisks()

For indirectly standardized outcomes, this function generates a descriptive scatterplot of the observed against the expected risk under the average care level for patients of that center (Figure 2).

```
> plotRisks(standardizedRisks = indirectRisks,
+           labeledCenters = labeledCenters)
```

It illustrates how much the observed risk in each center deviates from the expected risk under the average care level for patients of that center. Large deviation is expected when differences in patient mix are large among centers and is visualised by large deviations from the first bisector. The character ‘M’ denotes the estimated overall mortality risk  $\hat{E}(Y)$ .

For directly standardized outcomes, this function generates a descriptive scatterplot of the observed against the estimated directly standardized risk for each center.

## plotCenterLabels()

Center performance classification can be visualised using the estimated standardized risk and variance per center from the output of `standardizeRisks()` and classification labels from `labelCenters()` (Figure 3).

```
> plotCenterLabels(standardizedRisks = indirectRisks,
+                  labeledCenters = labeledCenters)
```

Centers that are classified as having low mortality risk are indicated by green triangles (on the left) and their confidence interval (by default 50%) lies completely below the clinically relevant boundary. Similarly, centers classified with high mortality risk are indicated by red triangles (on the right) and have their confidence interval lying completely above the clinically relevant boundary. Of course, larger centers have narrower estimated confidence intervals than smaller centers.

## funnelPlot()

The funnelplot is an internationally recommended plot for comparing institutional performance (Spiegelhalter, 2005) (Figure 4). The estimated standardized risks are plotted against a measure of precision (default is center size). Care centers with an estimated standardized risk lying outside the 95% control limits are flagged as outlying centers. A horizontal line is drawn (at the displayed value), for indirect standardization this is at the average of the indirectly standardized excess risks over all centers while for direct standardization this is at the overall mortality risk  $\hat{E}(Y)$ .

```
> funnelPlot(standardizedRisks = indirectRisks)
```

## References

- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.
- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15.
- Spiegelhalter, D. (2005). Funnel plots for comparing institutional performance. *Statistics in medicine*, 24(8):1185–1202.
- Varewyck, M., Goetghebeur, E., Eriksson, M., and Vansteelandt, S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance. *Biostatistics*, 15(4):651–664.

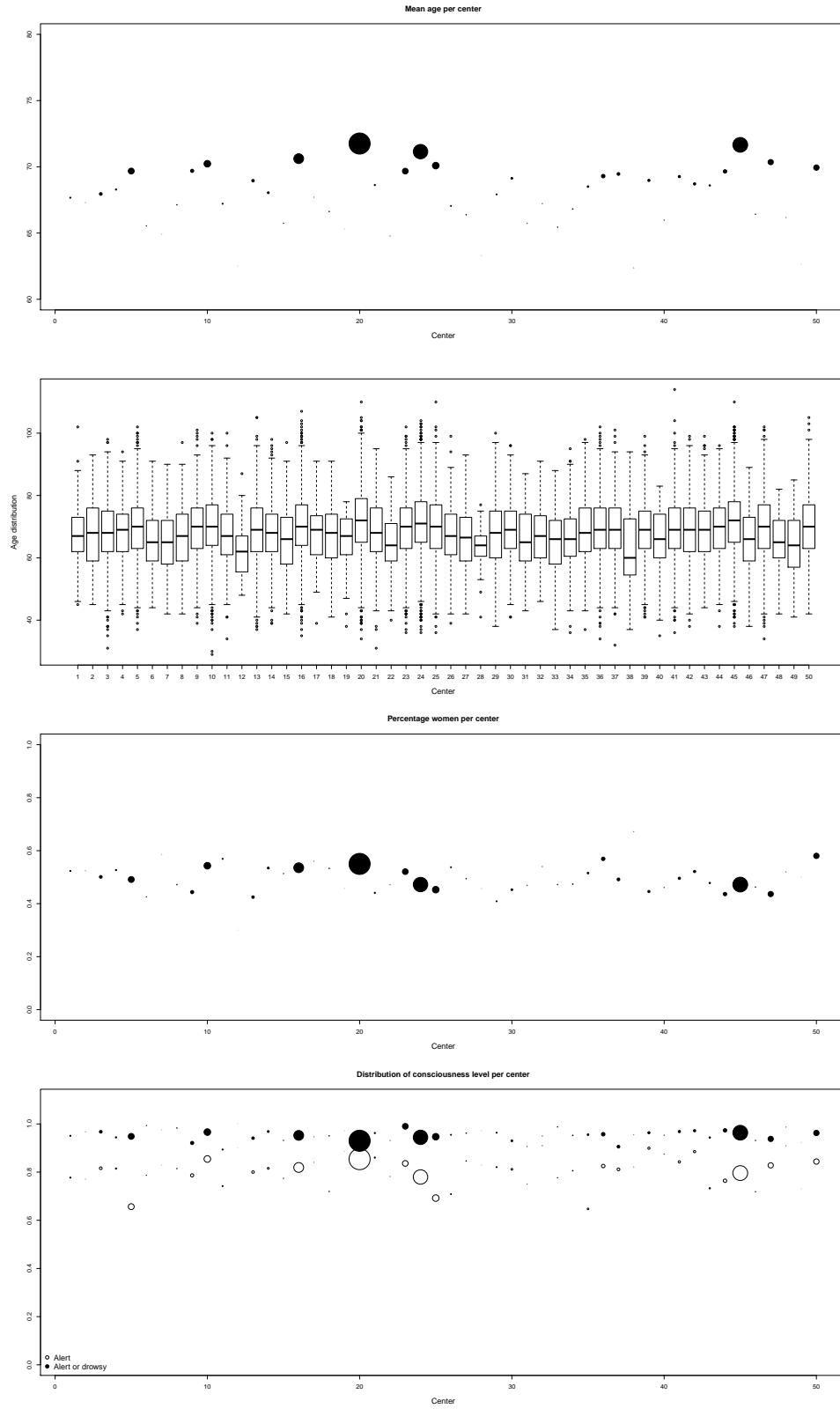


Figure 1: Descriptive plots for patient case-mix across centers. The size of the plot symbol is proportional to the center size.

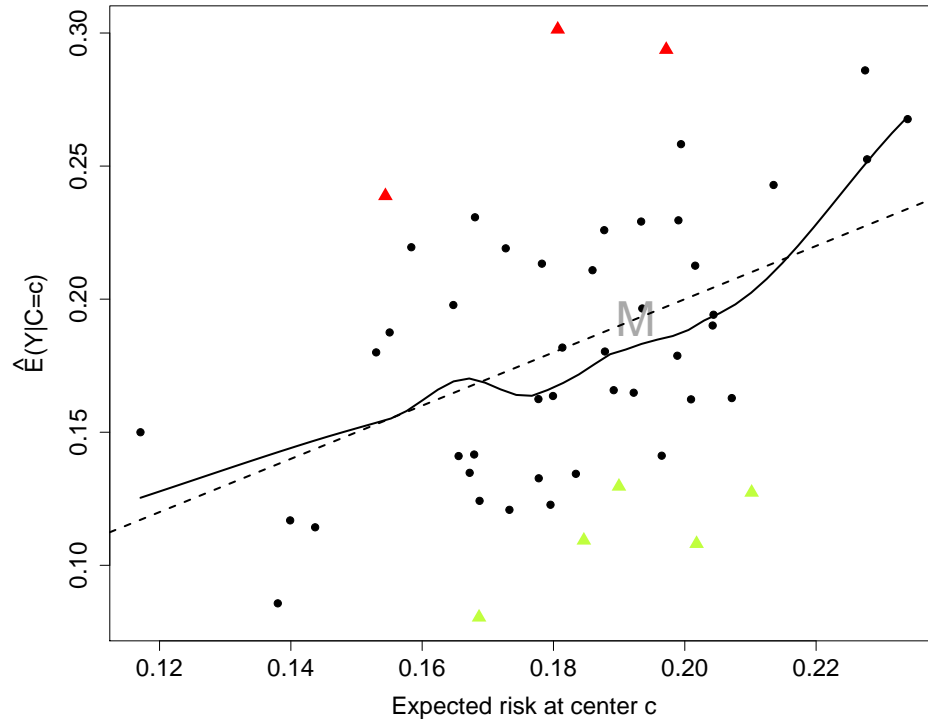


Figure 2: Plot the observed versus the expected risk under the average care level for patients of that center. The full line represents the best local fit through the points (`loess()` function in `R`), the dashed line represents the first bisector and 'M' denotes the estimated overall mortality risk  $\hat{E}(Y)$ .

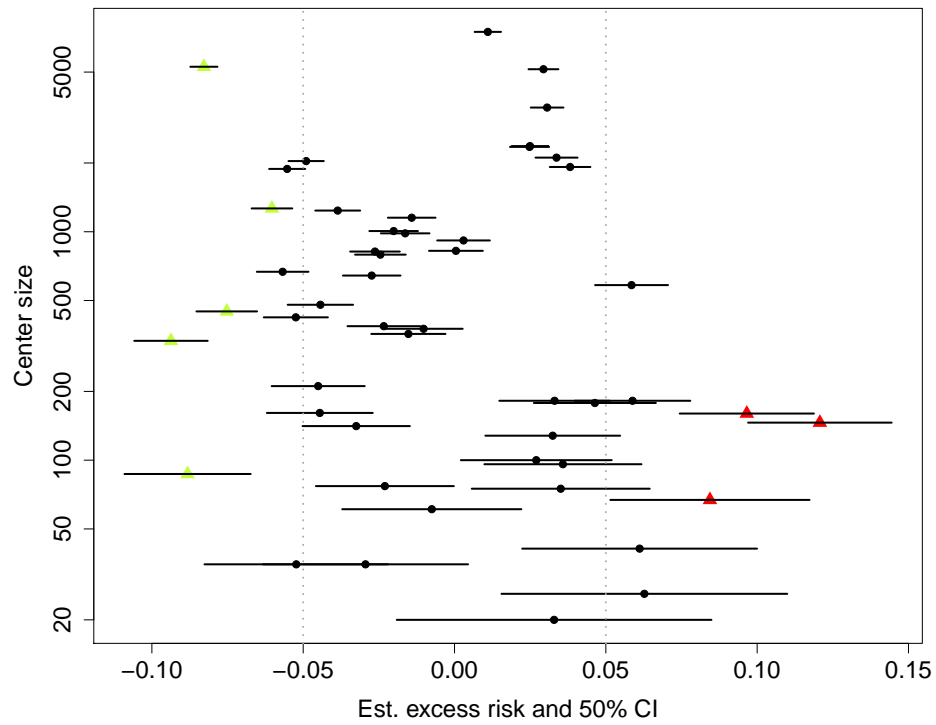


Figure 3: Plot the estimated directly standardized risk with 50% confidence limits per center.

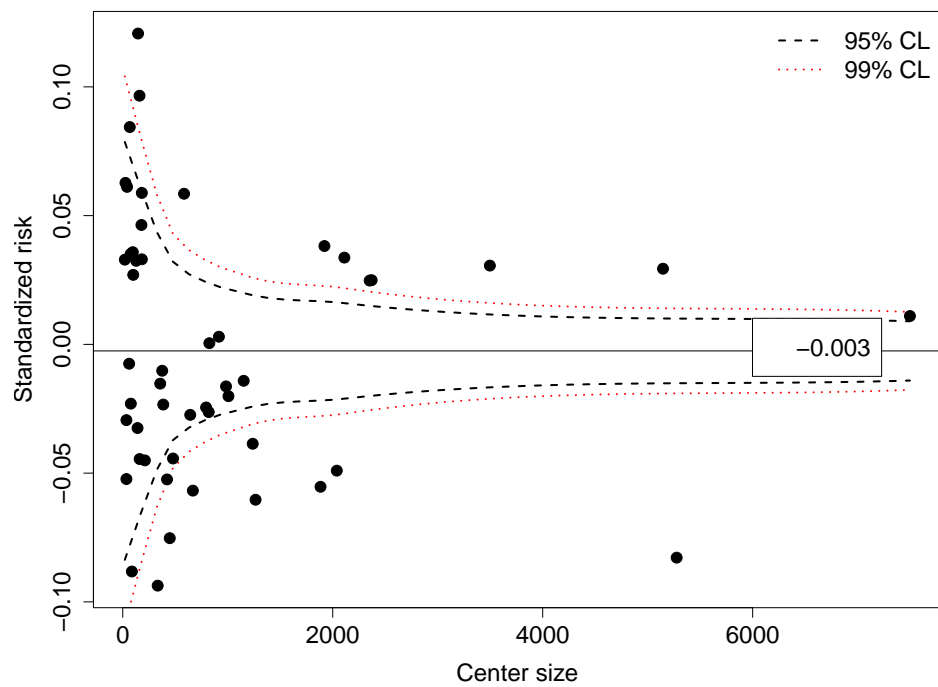


Figure 4: Funnelplot for the estimated indirectly standardized risk per center. The horizontal line represents the average over all centers of the indirectly standardized risks.