

The Analysis of Data from Countries with HRT Reduction and Decrease in Breast Cancer

June 10, 2015

Introduction

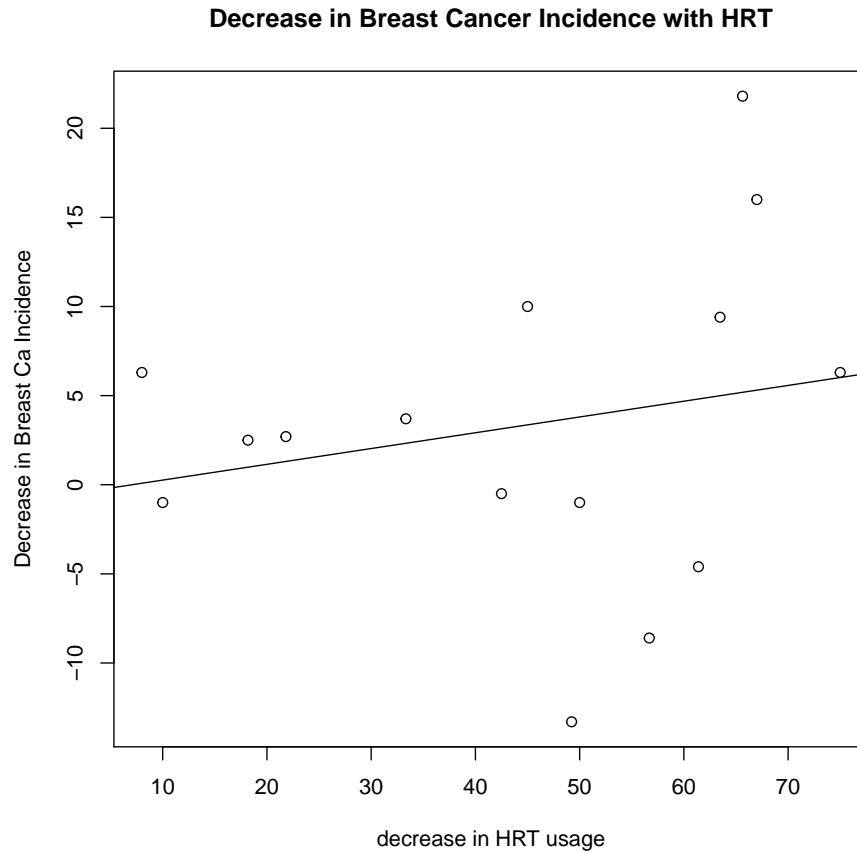
This set of analyses contains the R codes for the breast cancer study. The data were combined with the tables containing HRT usage data and the cancer incidence data were abstracted from the IARC databases corresponding to the decline in the incidence rates three years after the reduction in the usage of HRT in the countries. In the following set of analyses, we present both the R codes that were used to generate the plots and analyses and the graphs and tables.

The Plot of Decrease in Breast Cancer with HRT Usage Reduction

Figure 1 shows the plot of decrease in Ca incidence in the different countries (y axis) and decrease in the use of HRT in the corresponding countries as well.

The R Code to Produce the Plot

```
plot(mydata$declinehrtuse, mydata$cadecrease,  
     main = "Decrease in Breast Cancer Incidence with HRT",  
     ylab = "Decrease in Breast Ca Incidence",  
     xlab = "decrease in HRT usage")  
  
abline(hrtuseca <- lm(cadecrease ~ declinehrtuse,  
                      data = mydata))
```



```
## let's take a look at the simple linear regression model,

library(knitr)
kable(summary(hrtuseca)$coef, digits=2)
```

	Estimate	Std. Error	t value	p-value
(Intercept)	-0.62	5.53	-0.11	0.91
Decline in HRT use	0.09	0.11	0.79	0.45

As can be seen in the above scatterplot, in general, with decrease in HRT usage, there was a corresponding reduction in Breast Cancer incidence three years after the reduction in HRT usage. However, this effect was not uniform as In six countries, there were also an increase in Breast Cancer incidence following reduction in HRT usage.

In general, the fit line of the scatter plot indicate that with increase in decrease in HRT usage there was corresponding increase in reduction in Breast Cancer incidence as well. However it also suggests that the relationship between reduction in HRT usage and the corresponding reduction in Breast Cancer incidence followed a pattern, that the reduction in Breast Cancer was higher in those countries where the decrease in HRT usage was also very high, and this did not follow an exactly linear trend. A quadratic trend was explored by developing a quadratic model using a squared term of the HRT Reduction (that is, HRT Reduction Squared and putting this into the regression equation). The resulting summary of the coefficients of regression are shown as follows.

Table 2: Association between HRT Usage Reduction and Reduction in Br Ca incidence three years later – quadratic model

	Estimate	Std. Error	t value	p-value
Intercept	9.35	8.76	1.07	0.31
Decline in HRT use	-0.61	0.50	-1.22	0.25
Decline in HRT Use Squared	0.01	0.01	1.43	0.18

Removal of Data Points with Zero or Increased Incidence and Examination

Next, information of the countries that had registered an increase in Breast Cancer three years following the reduction in HRT usage drop were removed from the data frame of analyses and information from only those countries were retained that had registered a drop in Breast Cancer incidence three years after drop in HRT usage. Figure 2 shows the relationship between the drop in HRT usage and the corresponding drop in Breast Cancer incidence three years after the HRT usage drop.

The Code to remove the countries with zero or increase in Breast Cancer Incidence

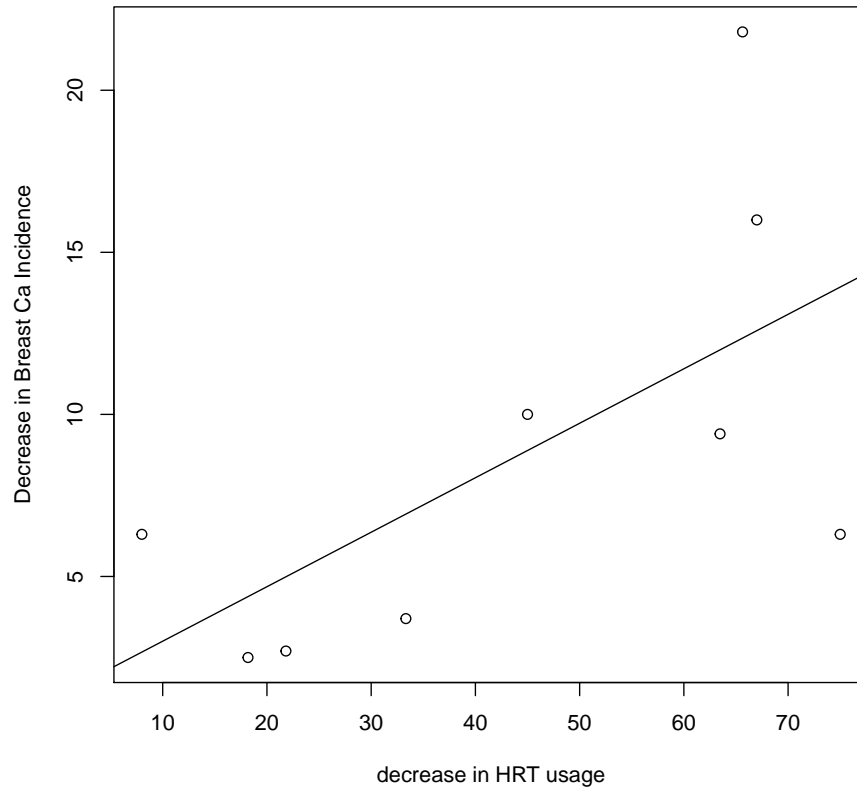
```
mydata2 <- subset(mydata, cadecrease >= 0)

# print(summarize(mydata2))

plot(mydata2$declinehrtuse, mydata2$cadecrease,
      main = "Decrease in Breast Cancer Incidence with HRT",
      ylab = "Decrease in Breast Ca Incidence",
```

```
xlab = "decrease in HRT usage")
abline(hrtca2 <- lm(cadecrease ~ declinehrtuse, data = mydata2))
```

Decrease in Breast Cancer Incidence with HRT



```
kable(summary(hrtca2)$coef, digits = 2)
```

	Estimate	Std. Error	t value	p-value
(Intercept)	1.32	3.81	0.35	0.74
Decline in HRT use	0.17	0.08	2.21	0.06

A very similar pattern of association can be observed between reduction in HRT usage and decrease in Breast Ca incidence, that is, as HRT usage has reduced in these countries, a corresponding decrease in Breast Cancer can be observed. Further, with higher levels of decreases in HRT usage the corresponding drop in Breast Cancer have been higher.

Further Exploration with Categorising the Breast Cancer Incidence Drop and HRT Usage

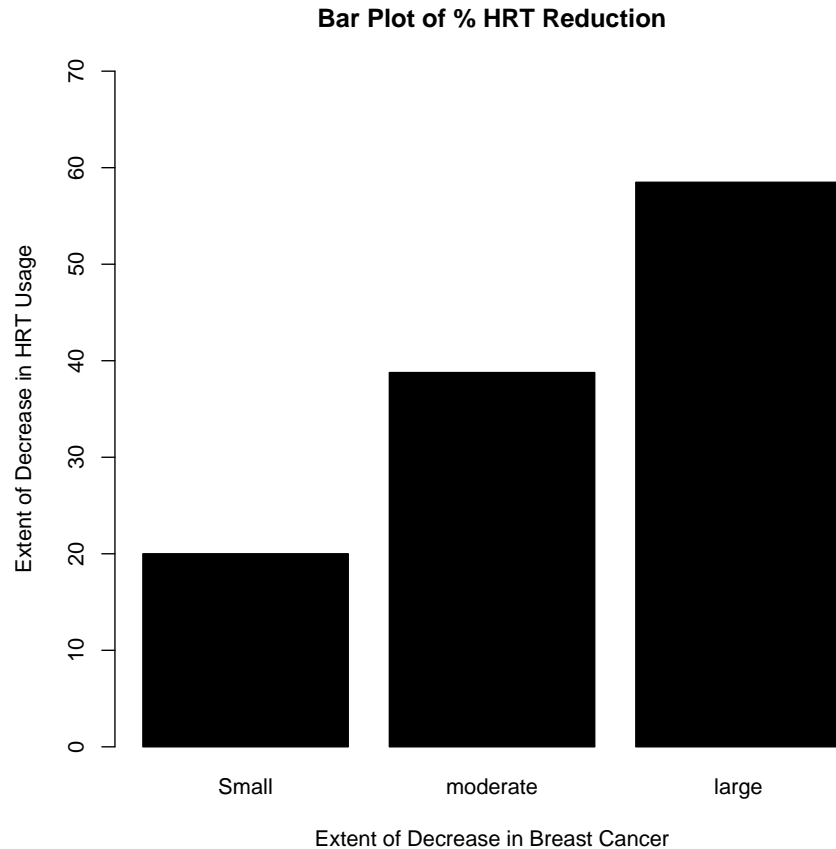
This was further examined in a subset analysis where the decrease in breast cancer incidence is divided into three equal groups:

1. No to mild decrease in Breast Ca incidence (0-3.2)
2. Moderate decrease in Breast Ca incidence (3.2-8.6)
3. Large decrease in Breast Ca incidence (> 8.6)

We examine graphically the reduction in HRT usage in the three groups of countries.

The Code and the Plot for Reduction in HRT Usage in Y axis and Breast Ca incidence reduction in X axis

```
barplot(hrtcadeclevel,  
        col = "black", main = "Bar Plot of % HRT Reduction",  
        xlab = "Extent of Decrease in Breast Cancer",  
        ylab = "Extent of Decrease in HRT Usage",  
  
        ylim = c(0, 70))
```



Graphical Exploration of Reduction in HRT Usage in X axis (categorised) and reduction in Incidence of Breast Cancer three years after the last registered year of HRT usage drop

This was further explored with drawing bar plots where the reduction in Breast Ca incidence three years after the last known year for drop in HRT usage was plotted against the extent to which the countries showed a drop in HRT usage. This is essentially the above plot turned on the other side. In addition, a series of box plots are also shown here that suggests that at higher level of reduction in HRT usage, the corresponding drop in the Breast Cancer incidence were high. The usage of reduction in HRT was classified as follows:

1. Lowest level of drop, 30% or lower in the drop from peak

2. Middle Level of drop, between 30 to 65.5 %
3. High level of drop in HRT usage, more than 65.5 %

The Code and the Plots are Shown Below

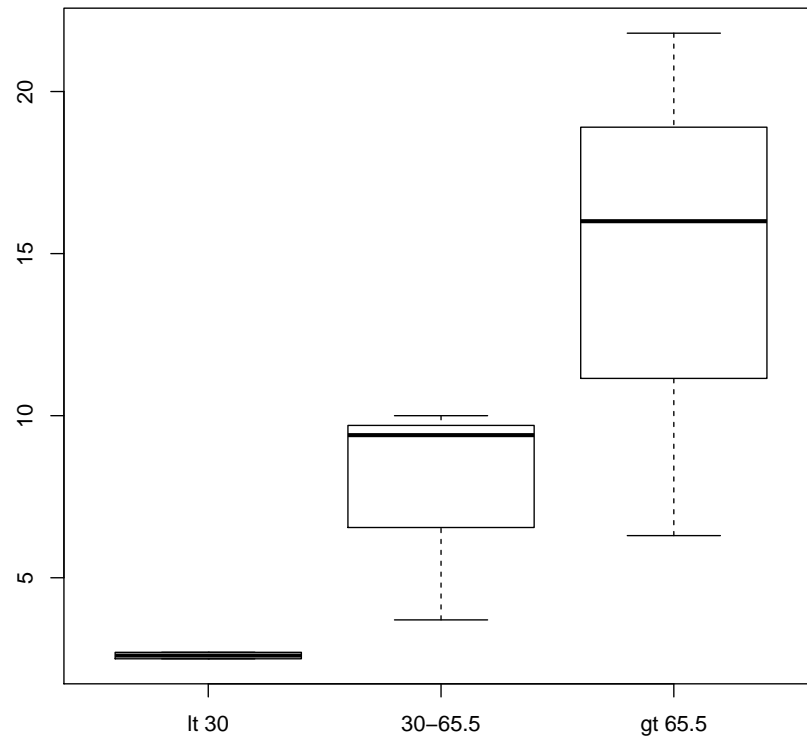
```
## Now, analyse the the extent to reduction in HRT usage
##print(summary(mydata2$declinehrtuse))
## given the minimum 8, q1 21.8, q3 65.5, highest 75
mydata2$hrtdeclinecat <- cut(mydata2$declinehrtuse,
                             breaks = c(8, 30, 65.6, 75))
levels(mydata2$hrtdeclinecat) <- c("lt 30", "30-65.5", "gt 65.5")

cadeclevel <- tapply(mydata2$cadecrease, mydata2$hrtdeclinecat, mean)
```

It can be seen that there is a progressive increase in the reduction in Breast Cancer incidence three years after the reduction in HRT usage

This can be further examined using the boxplots of reduction

```
boxplot(mydata2$cadecrease ~ mydata2$hrtdeclinecat)
```



Also as the barplot shows,

```
barplot(cadeclevel,  
  col = "black", main = "Bar Plot of Breast Ca Reduction",  
  ylab = "Extent of Decrease in Breast Cancer",  
  xlab = "Extent of Decrease in HRT Usage",  
  
  ylim = c(0, 20))
```