`compareGroups`: Descriptives by groups

Isaac Subirana [1,2,3], Héctor Sanz [2,4],Joan Vila [2,1]

February 3, 2012

[1]CIBER Epidemiology and Public Health (CIBERESP)
[2]IMIM-Institut de Recerca Hospital del Mar
[3]Statistics Department, University of Barcelona
[4]Consorcio de Apoyo a la Investigación Biomédica en Red (CAIBER)

isubirana@imim.es, hsanz@imim.es, jvila@imim.es

Contents

1 Introduction

The `compareGroups` package allows users to create tables displaying results of univariate analyses, stratified or not by categorical variable groupings.

Tables can easily be exported to CSV, LaTeX or HTML.

This package can be used from the ℝ prompt or from a user-friendly GUI.

This document provides an overview of the usage of the `compareGroups` package.

To load the package using the ℝ prompt, enter:

```
> library(compareGroups)
```

Once the package is loaded, non-R users can follow the GUI instructions in Section 5.

2 Design: classes and methods

The `compareGroups` package has three functions:

- `compareGroups` creates an object of class `compareGroups`. This object can be:

  - printed
  - summarized
  - plotted
  - updated

- `createTable` creates an object of class `createTable`. This object can be:

  - printed
  - summarized

- `export2latex`, `export2csv`, `export2html` will export results to CSV, LaTeX or HTML, respectively.

Figure 1 shows the diagram of the package.

Figure 1: Diagram of the `compareGroups` package

3 Data used as example

To illustrate how this package works we used data from a cohort study aimed to establish risk factors related to coronary heart disease. Some variables have been partially modified to provide a better example for the use of this package.

```
> data(regicor)
> head(regicor)
```
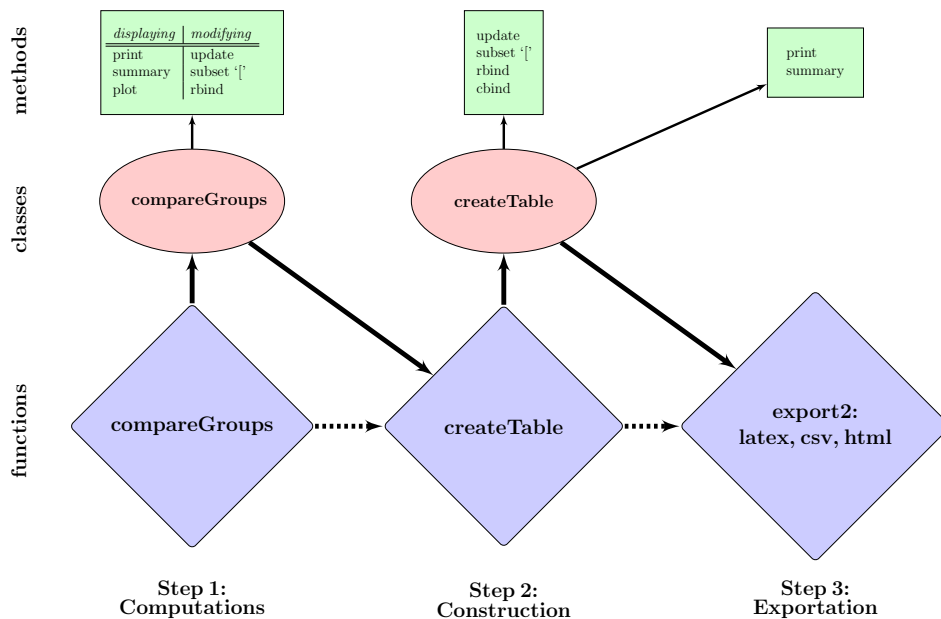
```
            id year age    sex                smoker sbp dbp histhtn txhtn chol      hdl triglyc      ldl histchol txchol he
6101      2265 2005  70 Female          Never smoker 138  75      No    No  294 57.00000      93 218.40000       No      No
5762      1882 2005  56 Female          Never smoker 139  89      No    No  220 50.00000     160 138.00000       No      No
2992 3000105616 2000  37   Male Current or former < 1y 132  82      No    No  245 59.80429      89 167.39571       No      No
2611 3000103485 2000  69 Female          Never smoker 168  97      No    No  168 53.17571     116  91.62429       No      No
2762 3000103963 2000  70 Female                  <NA>  NA  NA      No    No   NA       NA      NA        NA     <NA>    <NA>
1516 3000100883 2000  40 Female Current or former < 1y 108  70      No    No   NA 68.90000      94        NA       No      No
       phyact    pcs    mcs   cv     tocv death   todeath
6101 304.2000 54.455 58.918   No 1024.882   Yes 1299.16343
5762 160.3000 58.165 47.995   No 2756.849    No   39.32629
2992 552.7912 43.429 62.585   No 1905.969    No  858.42203
2611 522.0000 54.325 57.900   No 1055.380    No 1833.07619
2762       NA     NA     NA <NA>       NA  <NA>         NA
1516 386.9505 57.315 47.869   No 3239.241    No  877.61155
```

Variables and labels in this data frame are:

Name	Label	Codes
id	Individual id	
year	Recruitment year	1995; 2000; 2005
age	Age	
sex	Sex	Male; Female
smoker	Smoking status	Never smoker; Current or former < 1y; Former ≥ 1y
sbp	Systolic blood pressure	
dbp	Diastolic blood pressure	
histhtn	History of hypertension	Yes; No
txhtn	Hypertension treatment	No; Yes
chol	Total cholesterol	
hdl	HDL cholesterol	
triglyc	Triglycerides	
ldl	LDL cholesterol	
histchol	History of hyperchol.	Yes; No
txchol	Cholesterol treatment	No; Yes
height	Height (cm)	
weight	Weight (Kg)	
bmi	Body mass index	
phyact	Physical activity (Kcal/week)	
pcs	Physical component	
mcs	Mental component	
cv	Cardiovascular event	No; Yes
tocv	Days to cardiovascular event or end of follow-up	
death	Overall death	No; Yes
todeath	Days to overall death or end of follow-up	

OBSERVATIONS:

1. It is important to note that `compareGroups` is not aimed to perform quality control of the data.

Other useful packages such as `r2lh` [2] are available for this purpose.

2. It is strongly recommended that the *data.frame* contain only the variables to be analyzed; the ones not needed in the present analysis should be removed from the list.

3. The nature of variables to be analyzed should be known, or at least which variables are to be used as categorical. It is important to code categorical variables as factors and the order of their levels is meaningful in this package.

4. The function `label` from the `Hmisc` package could be used to label the variables properly. The tables of results will contain the variable labels (by default).

3.1 Time-to-event variables

A variable of class `Surv` must be created to deal with time-to-event variables (i.e., time to Cardiovascular event/censored in our example):

```
> regicor$tdeath <- with(regicor, Surv(todeath, death == 'Yes'))
> label(regicor$tdeath) <- "Mortality"
> regicor$tcv <- with(regicor, Surv(tocv, cv == 'Yes'))
> label(regicor$tcv) <- "Cardiovascular"
```

Note that variables *tdeath* and *tcv* are created as time-to-death and time-to-cardiovascular event, respectively; both take censoring into account (i.e. they are of class `Surv`).

4 Using R syntax

4.1 compareGroups

This is the main function. It does all the calculus. It is needed to store results in an object. Later, applying the function `createTable` (Section 4.2) to this object will create tables of the analysis results.

For example, to perform a univariate analysis with the *regicor* data between *year* ("response" variable) and all other variables ("explanatory" variables), this formula is required:

```
> compareGroups(year ~ . , data=regicor)
```

4.1.1 Selecting response variables

If only a dot occurs on the right side of the "~" all variables in the data frame will be used.

To remove the variable *id* from the analysis:

```
> compareGroups(year ~ . -id, data=regicor)
```

To select some explanatory variables (e.g., *age*, *sex*, *smoker* and *triglyc*) and store results in an object of class `compareGroups`:

```
> res<-compareGroups(year ~ age + sex + smoker + triglyc , data=regicor)
> res
```

```
-------- Summary of results by groups of 'Recruitment year'---------


  var             N    p.value  method             selection
1 Age             2294 0.078*   continuous normal ALL
2 Sex             2294 0.506    categorical        ALL
3 Smoking status  2233 <0.001** categorical        ALL
4 Triglycerides   2231 0.582    continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Note: Although we have full data (n= 2294) for Age and Sex there are some missing data in Smoking status and Triglycerides.

Year groups have some differences in Age although these don't reach statistical significance (p-value=0.078); however, Smoking status is clearly different.

Age & Triglycerides has been used as continuous and normal distributed. Sex & Smoking status as categorical.

No filters have been used (e.g., selecting only treated patients); therefore, the *selection* column lists "ALL" (for all variables).

4.1.2 Subsetting

To perform the analysis in a subset of participants (e.g., "male" participants):

```
> compareGroups(year ~ age + smoker + triglyc, data=regicor, subset = sex=='Male')


-------- Summary of results by groups of 'Recruitment year'---------


  var             N    p.value  method             selection
1 Age             1101 0.212    continuous normal sex == "Male"
2 Smoking status  1071 <0.001** categorical        sex == "Male"
3 Triglycerides   1075 0.817    continuous normal sex == "Male"
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Note that only results for male participants are shown.

To subset specific variable/s (e.g., *triglyc* and *sbp*):

```
> compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+           selec = list(triglyc=txchol=='No', sbp=txhtn=='No'))


-------- Summary of results by groups of 'Recruitment year'---------


  var                     N    p.value  method             selection
1 Age                     2294 0.078*   continuous normal ALL
2 Smoking status          2233 <0.001** categorical        ALL
3 Triglycerides           1963 0.882    continuous normal txchol == "No"
4 Systolic blood pressure 1810 <0.001** continuous normal txhtn == "No"
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Combinations are also allowed, e.g.:

```
> compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+     selec = list(triglyc=txchol=='No', sbp=txhtn=='No'),
+     subset = sex=='Male')
```

```
-------- Summary of results by groups of 'Recruitment year'---------


  var                     N    p.value  method
1 Age                     1101 0.212    continuous normal
2 Smoking status          1071 <0.001** categorical
3 Triglycerides            943 0.993    continuous normal
4 Systolic blood pressure  886 <0.001** continuous normal
  selection
1 sex == "Male"
2 sex == "Male"
3 (sex == "Male") & (txchol == "No")
4 (sex == "Male") & (txhtn == "No")
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

A variable can appear twice in the formula, e.g.:

```
> compareGroups(year ~ age + smoker + triglyc + triglyc, data=regicor,
+         selec = list(triglyc.1=txchol=='No'))
```

```
-------- Summary of results by groups of 'Recruitment year'---------


  var           N    p.value  method             selection
1 Age           2294 0.078*   continuous normal  ALL
2 Smoking status 2233 <0.001** categorical        ALL
3 Triglycerides 2231 0.582    continuous normal  ALL
4 Triglycerides 1963 0.882    continuous normal  txchol == "No"
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

In this case results for *triglyc* will be reported for all participants (n= 2231) and also for only those not treated with anti-cholesterol drugs (txchol=='No'). Note that "triglyc.1" in the **selec** statement refers to the second time that *triglyc* appears in the formula.

4.1.3 Methods for continuous variables

By default continuous variables are analyzed as normal-distributed. When a table is built (see `createTable` function, Section 4.2), continuous variables will be described with mean and standard deviation. To change default options, e.g., Triglycerides used as non-normal distributed:

```
> compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor, method = c(triglyc=2))
```

```
-------- Summary of results by groups of 'Recruitment year'---------


  var                     N    p.value  method               selection
1 Age                     2294 0.078*   continuous normal    ALL
2 Smoking status          2233 <0.001** categorical          ALL
3 Triglycerides           2231 0.762    continuous non-normal ALL
4 Systolic blood pressure 2280 <0.001** continuous normal    ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Note that "continuous non-normal" is shown in the *method* column for the variable Triglycerides.

Possible values in methods statement are:

- 1: forces analysis as normal-distributed
- 2: forces analysis as continuous non-normal
- 3: forces analysis as categorical
- NA: performs a Shapiro-Wilks test to decide between normal or non-normal

If the **method** for a variable is stated as **= NA**, then a Shapiro-Wilk test for normality is used to decide if the variable is normal or non-normal distributed. To change the significance threshold:

```
> compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+                method = c(triglyc=NA), alpha= 0.01)


-------- Summary of results by groups of 'Recruitment year'---------


  var                     N    p.value  method               selection
1 Age                     2294 0.078*   continuous normal     ALL
2 Smoking status          2233 <0.001** categorical           ALL
3 Triglycerides           2231 0.762    continuous non-normal ALL
4 Systolic blood pressure 2280 <0.001** continuous normal     ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

According to Shapiro-Wilk test, stating the cutpoint at 0.01 level, Triglycerides departed significantly from the normal distribution and therefore the method for this variable will be "continuous non-normal".

All non factor variables are considered as continuous. Exception is made (by default) for those that have fewer than 5 different values. This threshold can be changed in the **min.dis** statement:

```
> cuts<-"lo:40=1; 41:45=2; 46:50=3; 51:55=4; 56:60=5; 61:65=6; 66:hi=7"
> regicor$age7gr<-car::recode(regicor$age, cuts)
> compareGroups(year ~ age7gr, data=regicor, method = c(age7gr=NA))


-------- Summary of results by groups of 'Recruitment year'---------


  var N     p.value method                selection
1 Age 2294 0.049** continuous non-normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

```
> compareGroups(year ~ age7gr, data=regicor, method = c(age7gr=NA), min.dis=8)


-------- Summary of results by groups of 'Recruitment year'---------


  var N     p.value method      selection
1 Age 2294 0.001** categorical ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

To avoid errors the maximum categories for the response variable is set at 5 in this example (default value). If this variable has more than 5 different values, the function `compareGroups` returns an error message. For example:

```
> compareGroups(age7gr ~ smoker + triglyc + sbp, data=regicor)


Error en compareGroups.default(X = X, y = y, include.label = include.label,  :
  number of groups must be less or equal to 5
```

Defaults setting can be changed with the **max.ylev** statement:

```
> compareGroups(age7gr ~ smoker + triglyc + sbp, data=regicor, max.ylev=7)
```

```
-------- Summary of results by groups of 'Age'---------


  var                    N   p.value method         selection
1 Smoking status        2233 <0.001** categorical      ALL
2 Triglycerides         2231 0.021**  continuous normal ALL
3 Systolic blood pressure 2280 <0.001** continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Similarly, by default there is a limit for the maximum number of levels for an explanatory variable. If this level is exceeded, the variable is removed from the analysis and a warning message is printed:

```
> compareGroups(year ~ smoker + age7gr, method= (age7gr=3), data=regicor, max.xlev=5)


-------- Summary of results by groups of 'Recruitment year'---------


  var            N   p.value method      selection
1 Smoking status 2233 <0.001** categorical ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1


Warning in compareGroups.default(X = X, y = y, include.label = include.label,  :
 Variables 'age7gr' have been removed since some errors ocurred
```

4.1.4   Dressing up the output

Although the options described in this section correspond to `compareGroups` function, results of changing/setting them won't be visible until the table is created with the `createTable` function (explained later).

include.label   By default the variable labels are shown in the output (if there is no label the name will be printed). Changing the statement include.label from "= TRUE" (default) to "= FALSE" will cause variable names to be printed instead.

```
> compareGroups(year ~ smoker + triglyc + sbp, data=regicor, include.label= FALSE)


-------- Summary of results by groups of 'year'---------


  var    N   p.value method         selection
1 smoker 2233 <0.001** categorical      ALL
2 triglyc 2231 0.582    continuous normal ALL
3 sbp    2280 <0.001** continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Q1, Q3   When the method for a variable is stated as "2" (i.e., to be analyzed as continuous non-normal; see section 4.1.3), by default the median and quartiles 1 and 3 will be shown in the final results, after applying the function `createTable` (see Section 4.2).

```
> resu1<-compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+                      method = c(triglyc=2))
> createTable(resu1)
```

```
--------Summary descriptives table by 'Recruitment year'---------


------------------------------------------------------------------------------------------
                                 1995            2000              2005          p.overall
                                N=431           N=786            N=1077
------------------------------------------------------------------------------------------
Age                         54.1 (11.7)     54.3 (11.2)       55.3 (10.6)         0.078
Smoking status:                                                                  <0.001
    Never smoker            234 (56.4%)     414 (54.6%)       553 (52.2%)
    Current or former < 1y  109 (26.3%)     267 (35.2%)       217 (20.5%)
    Former >= 1y             72 (17.3%)      77 (10.2%)       290 (27.4%)
Triglycerides              94.0 [71.0; 136] 98.0 [72.0; 133] 98.0 [72.0; 139]    0.762
Systolic blood pressure     133 (19.2)      133 (21.3)        129 (19.8)        <0.001
------------------------------------------------------------------------------------------
```

Note: percentiles 25 and 75 are calculated for *Triglycerides*.

To get instead percentile 2.5 and 97.5:

```
> resu2<-compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+                      method = c(triglyc=2), Q1=0.025, Q3=0.975)
> createTable(resu2)
```

```
--------Summary descriptives table by 'Recruitment year'---------


------------------------------------------------------------------------------------------
                                 1995            2000              2005          p.overall
                                N=431           N=786            N=1077
------------------------------------------------------------------------------------------
Age                         54.1 (11.7)     54.3 (11.2)       55.3 (10.6)         0.078
Smoking status:                                                                  <0.001
    Never smoker            234 (56.4%)     414 (54.6%)       553 (52.2%)
    Current or former < 1y  109 (26.3%)     267 (35.2%)       217 (20.5%)
    Former >= 1y             72 (17.3%)      77 (10.2%)       290 (27.4%)
Triglycerides              94.0 [47.0; 292] 98.0 [47.0; 278] 98.0 [42.0; 293]    0.762
Systolic blood pressure     133 (19.2)      133 (21.3)        129 (19.8)        <0.001
------------------------------------------------------------------------------------------
```

Note: percentiles 2.5 and 97.5 are calculated for Triglycerides.

To get minimum and maximum:

```
> compareGroups(year ~ age + smoker + triglyc + sbp, data=regicor,
+               method = c(triglyc=2), Q1=0, Q3=1)
```

simplify Sometimes a categorical variable has no individuals for a specific group. For example, *smoker* has 3 levels. As an example and to illustrate this problem, we have created a new variable *smk* with a new category ("Unknown"):

```
> regicor$smk<-regicor$smoker
> levels(regicor$smk)<- c("Never smoker", "Current or former < 1y",
+                         "Never or former >= 1y", "Unknown")
> cbind(table(regicor$smk))
```

```
                        [,1]
Never smoker            1201
Current or former < 1y   593
Never or former >= 1y    439
Unknown                    0
```

Note that this new category ("unknown") has no individuals:

```
> compareGroups(year ~ age + smk + sbp, data=regicor)
```

```
Error in fisher.test(obj) : FEXACT error 6.
LDKEY is too small for this problem.
Try increasing the size of the workspace.


-------- Summary of results by groups of 'Recruitment year'---------


  var                    N    p.value  method          selection
1 Age                    2294 0.078*   continuous normal ALL
2 Smoking status         2233 .        categorical      ALL
3 Systolic blood pressure 2280 <0.001** continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Note that an "Error" message is printed and no p-values for *smk* are calculated.

To avoid using empty categories, `simplify` must be stated as `TRUE`

```
> compareGroups(year ~ age + smk + sbp, data=regicor, simplify=TRUE)


-------- Summary of results by groups of 'Recruitment year'---------


  var                    N    p.value  method          selection
1 Age                    2294 0.078*   continuous normal ALL
2 Smoking status         2233 <0.001** categorical      ALL
3 Systolic blood pressure 2280 <0.001** continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

4.1.5 Summary

Applying the **summary** function to an object of class `createTable` will obtain a more detailed output:

```
> res<-compareGroups(year ~ age + sex + smoker + txchol + triglyc + sbp,
+                     method = c(triglyc=2), data=regicor)
> summary(res[c(1, 3, 4)])


 --- Descriptives of each row-variable by groups of 'Recruitment year' ---

-------------------
row-variable: Age

      N    mean      sd       p.overall p.trend  p.1995 vs 2000 p.1995 vs 2005 p.2000 vs 2005
[ALL] 2294 54.73627 11.04926
1995  431  54.09745 11.7172  0.077837  0.031665 0.930249       0.143499       0.161195
2000  786  54.33715 11.21814
2005  1077 55.28319 10.62606


-------------------
row-variable: Smoking status

      Never smoker Current or former < 1y Former >= 1y Never smoker (row%) Current or former < 1y (row%)
[ALL] 1201         593                    439          53.78415            26.5562
1995  234          109                    72           56.38554            26.26506
2000  414          267                    77           54.61741            35.22427
2005  553          217                    290          52.16981            20.4717
      Former >= 1y (row%) p.overall p.trend p.1995 vs 2000 p.1995 vs 2005 p.2000 vs 2005
[ALL] 19.65965
1995  17.3494             0         2.4e-05 0.000144       0.000144       0
2000  10.15831
2005  27.35849


-------------------
row-variable: Cholesterol treatment
```

```
        No   Yes No (row%) Yes (row%) p.overall p.trend p.1995 vs 2000 p.1995 vs 2005 p.2000 vs 2005
[ALL] 2011 228  89.81688   10.18312
1995   403  28  93.50348    6.49652  0.000429    1e-04       0.193023        0.00196        0.014913
2000   705  68   91.2031    8.796895
2005   903 132  87.24638   12.75362
```

Note that because only variables 1, 3 & 4 are selected, only results for Age, Cholesterol treatment & Triglycerides are shown. Age is summarized by the mean and the standard deviation, Cholesterol treatment by frequencies and percentage, and Triglycerides (method =2) by the median and quartiles.

4.1.6   Plotting

Variables can be plotted to see their distribution. Plots differ according to whether the variable is continuous or categorical. Plots can be seen on-screen or saved as PDF.

```
> plot(res[c(1,2)], file="./univar/")
```
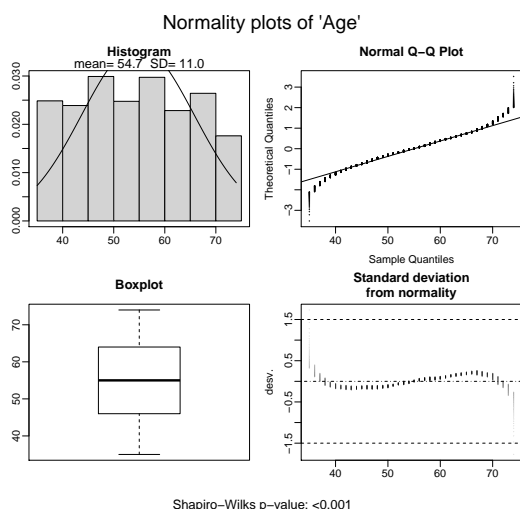

Figure 2: Plot of Age
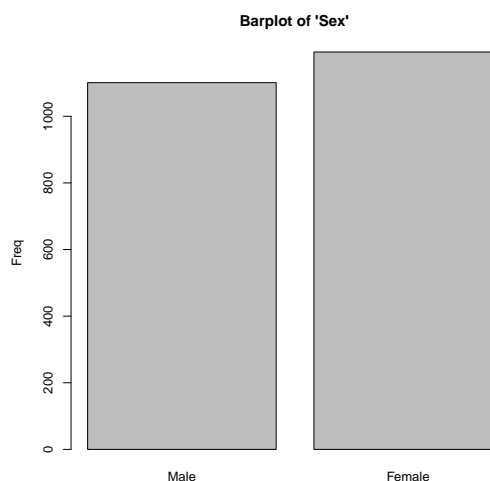
Figure 3: Plot of Sex

Plots also can be done according to grouping variable. In this case only a boxplot is shown for continuous variables:

```
> plot(res[c(1,2)], bivar=TRUE, file="./bivar/")
```

4.1.7   Updating

The object from `compareGroups` can later be updated. For example:

```
> res<-compareGroups(year ~ age + sex + triglyc, data=regicor)
> res
```

Figure 4: Plot of Age

Figure 5: Plot of Sex

```
-------- Summary of results by groups of 'Recruitment year'---------


  var           N    p.value method           selection
1 Age           2294 0.078*  continuous normal ALL
2 Sex           2294 0.506   categorical       ALL
3 Triglycerides 2231 0.582   continuous normal ALL
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

The object *res* is updated using:

```
> res<-update(res, . ~. - sex + smoker +  chol, subset = sex=='Male',
+                 method = c(triglyc=2, chol = 2),
+                 selec = list(triglyc=txchol=='No', sbp=txhtn=='No'))


> res
```

```
-------- Summary of results by groups of 'Recruitment year'---------


  var               N    p.value method                selection
1 Age               1101 0.212   continuous normal     sex == "Male"
2 Triglycerides      943 0.426   continuous non-normal (sex == "Male") & (txchol == "No")
3 Smoking status    1071 <0.001** categorical          sex == "Male"
4 Total cholesterol 1054 <0.001** continuous non-normal sex == "Male"
-----
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

Note that "Sex" is removed as an explanatory variable but used as a filter, subsetting only "Male" participants. Variable "Triglycerides" has been changed to "continuous non-normal" and to show results only for non-treated male participants. Two new variables have been added: Smoking status and Total cholesterol.

4.1.8   Odds Ratios & Hazard Ratios

When the response variable is binary, the Odds Ratio (OR) can be printed in the final table. If the response variable is time-to-event (see Section 3.1), the Hazard Ratio (HR) can be printed instead.

ref This statement can be used to change the reference category:

```
> res1<-compareGroups(sex ~ age + smoker + txchol + sbp, data=regicor, ref=1)
> createTable(res1, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------


-----------------------------------------------------------------------------------
                            Male          Female          OR          p.overall
                            N=1101        N=1193
-----------------------------------------------------------------------------------
Age                       54.8 (11.1)    54.7 (11.0)   1.00 [0.99;1.01]     0.840
Smoking status:                                                            <0.001
   Never smoker           301 (28.1%)    900 (77.5%)        Ref.
   Current or former < 1y 410 (38.3%)    183 (15.7%)   0.15 [0.12;0.19]
   Former >= 1y           360 (33.6%)     79 (6.80%)   0.07 [0.06;0.10]
Cholesterol treatment:                                                      0.583
   No                     962 (89.4%)   1049 (90.2%)        Ref.
   Yes                    114 (10.6%)    114 (9.80%)   0.92 [0.70;1.21]
Systolic blood pressure    134 (18.9)     129 (21.2)   0.99 [0.98;0.99]    <0.001
-----------------------------------------------------------------------------------
```

Note that for categorical response variables the reference category is the first one in the statement:

```
> res2<-compareGroups(sex ~ age + smoker + txchol + sbp, data=regicor,
+                     ref=c(smoker=1, txchol=2))
> createTable(res2, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------


-----------------------------------------------------------------------------------
                            Male          Female          OR          p.overall
                            N=1101        N=1193
-----------------------------------------------------------------------------------
Age                       54.8 (11.1)    54.7 (11.0)   1.00 [0.99;1.01]     0.840
Smoking status:                                                            <0.001
   Never smoker           301 (28.1%)    900 (77.5%)        Ref.
   Current or former < 1y 410 (38.3%)    183 (15.7%)   0.15 [0.12;0.19]
   Former >= 1y           360 (33.6%)     79 (6.80%)   0.07 [0.06;0.10]
Cholesterol treatment:                                                      0.583
   No                     962 (89.4%)   1049 (90.2%)   1.09 [0.83;1.43]
   Yes                    114 (10.6%)    114 (9.80%)        Ref.
Systolic blood pressure    134 (18.9)     129 (21.2)   0.99 [0.98;0.99]    <0.001
-----------------------------------------------------------------------------------
```

Note that the reference category for Smoking status is the first and for Cholesterol treatment the second.

ref.no Similarly to the "ref" statement, `ref.no` is used to state "no" as the reference category for all variables with this category:

```
> res<-compareGroups(sex ~ age + smoker + txchol + sbp, data=regicor, ref.no='NO')
> createTable(res, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------


-----------------------------------------------------------------------------------
                            Male          Female          OR          p.overall
                            N=1101        N=1193
-----------------------------------------------------------------------------------
Age                       54.8 (11.1)    54.7 (11.0)   1.00 [0.99;1.01]     0.840
Smoking status:                                                            <0.001
   Never smoker           301 (28.1%)    900 (77.5%)        Ref.
   Current or former < 1y 410 (38.3%)    183 (15.7%)   0.15 [0.12;0.19]
   Former >= 1y           360 (33.6%)     79 (6.80%)   0.07 [0.06;0.10]
Cholesterol treatment:                                                      0.583
   No                     962 (89.4%)   1049 (90.2%)        Ref.
   Yes                    114 (10.6%)    114 (9.80%)   0.92 [0.70;1.21]
Systolic blood pressure    134 (18.9)     129 (21.2)   0.99 [0.98;0.99]    <0.001
-----------------------------------------------------------------------------------
```

Note: 'no', 'No' or 'NO' will produce the same results; the coding is not case sensitive.

fact.ratio By default OR or HR for continuous variables are calculated for each unit increase. It can be changed by the `fact.or` statement:

```
> res<-compareGroups(sex ~ age + txchol + sbp, data=regicor)
> createTable(res, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------

_____
                          Male          Female           OR          p.overall
                         N=1101          N=1193
_____
Age                     54.8 (11.1)     54.7 (11.0)    1.00 [0.99;1.01]     0.840
Cholesterol treatment:                                                      0.583
    No                  962 (89.4%)     1049 (90.2%)        Ref.
    Yes                 114 (10.6%)     114 (9.80%)     0.92 [0.70;1.21]
Systolic blood pressure  134 (18.9)      129 (21.2)     0.99 [0.98;0.99]    <0.001
_____
```

Here the OR is for the increase of one unit for Age and Systolic blood pressure.

```
> res<-compareGroups(sex ~ age + txchol + sbp, data=regicor, fact.ratio= c(age=10, sbp=100))
> createTable(res, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------

_____
                          Male          Female           OR          p.overall
                         N=1101          N=1193
_____
Age                     54.8 (11.1)     54.7 (11.0)    0.99 [0.92;1.07]     0.840
Cholesterol treatment:                                                      0.583
    No                  962 (89.4%)     1049 (90.2%)        Ref.
    Yes                 114 (10.6%)     114 (9.80%)     0.92 [0.70;1.21]
Systolic blood pressure  134 (18.9)      129 (21.2)     0.26 [0.17;0.39]    <0.001
_____
```

Here the OR is for the increase of 10 years for Age and 100 mmHg for Systolic blood pressure.

ref.y By default when OR or HR are calculated, the reference category for the response variable is the first. The reference category could be changed using the `ref.y` statement:

```
> res<-compareGroups(sex ~ histchol + txchol + sbp, data=regicor)
> createTable(res, show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Sex'---------

_____
                          Male          Female           OR          p.overall
                         N=1098          N=1182
_____
History of hyperchol.:                                                      0.308
    Yes                 353 (32.3%)     356 (30.2%)         Ref.
    No                  741 (67.7%)     823 (69.8%)     1.10 [0.92;1.32]
Cholesterol treatment:                                                      0.583
    No                  962 (89.4%)     1049 (90.2%)        Ref.
    Yes                 114 (10.6%)     114 (9.80%)     0.92 [0.70;1.21]
Systolic blood pressure  134 (18.9)      129 (21.2)     0.99 [0.98;0.99]    <0.001
_____
```

Note: This output shows the OR of being female. Therefore, 'male' is the reference category.

```
> res<-compareGroups(sex ~ histchol + txchol + sbp, data=regicor, ref.y=2)
> createTable(res, show.ratio=TRUE)


--------Summary descriptives table by 'Sex'---------


_____
                             Male           Female           OR         p.overall
                            N=1098          N=1182
_____
History of hyperchol.:                                                    0.308
    Yes                  353 (32.3%)     356 (30.2%)        Ref.
    No                   741 (67.7%)     823 (69.8%)   0.91 [0.76;1.08]
Cholesterol treatment:                                                    0.583
    No                   962 (89.4%)    1049 (90.2%)        Ref.
    Yes                  114 (10.6%)     114 (9.80%)   1.09 [0.83;1.43]
Systolic blood pressure   134 (18.9)      129 (21.2)   1.01 [1.02;1.01]    <0.001
_____
```

Note: This output shows the OR of being male.

  Note: This output shows the OR of being male, and 'female' is now the reference category.

  When the response variable is of class `Surv`, the bivariate `plot` function returns a Kaplan-Meier figure if the explanatory variable is categorical. For continuous variables the function returns a line for each individual, ending with a circle for censored and with a plus sign for uncensored.

```
> plot(compareGroups(tcv ~ txchol, data=regicor), bivar=TRUE, file="./bivar/")
> plot(compareGroups(tcv ~ sbp, data=regicor), bivar=TRUE, file="./bivar/")
```
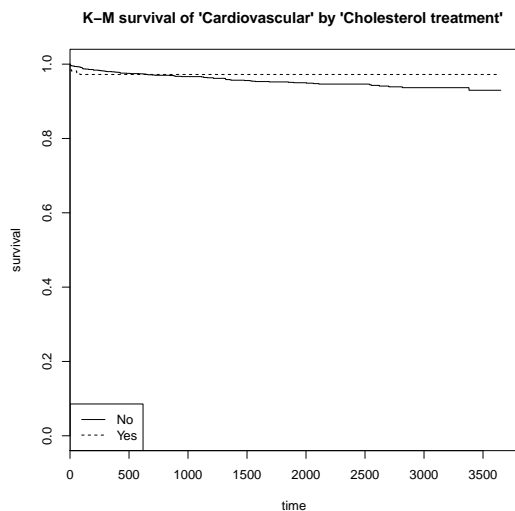

Figure 6: Categorical

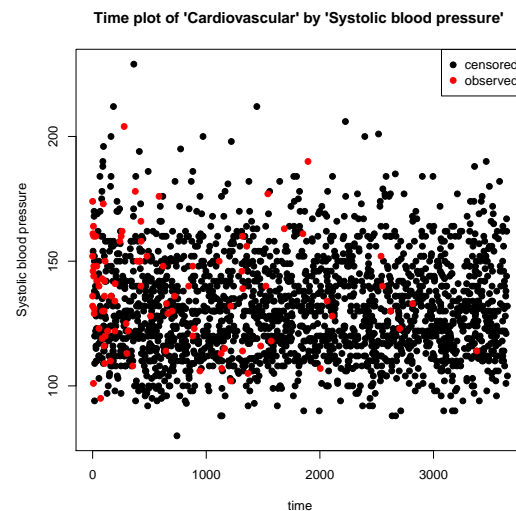
Figure 7: Continuous

4.1.9 Time-to-event explanatory variables

When a variable of class `Surv` (see Section 3.1) is used as explanatory it will be described with the probability of event, computed by Kaplan-Meier, up to a stated time.

15

timemax By default probability is calculated at the median of the follow-up period. `timemax` option allows us to change at what time probability is calculated.

```
> res<-compareGroups(sex ~  age + tdeath + tcv, timemax=c(tcv=365.25*5, tdeath=365.25*3),
+                    data=regicor)
> res
```

Note that *tdeath* is calculated at 5 years and *tcv* at 3 years (see section 3.1).

The `plot` function applied to a variable of class `Surv` returns a Kaplan-Meier figure. The figure can be stratified by the grouping variable.

```
> plot(res[2], file="./univar/")
> plot(res[2], bivar=TRUE, file="./bivar/")
```
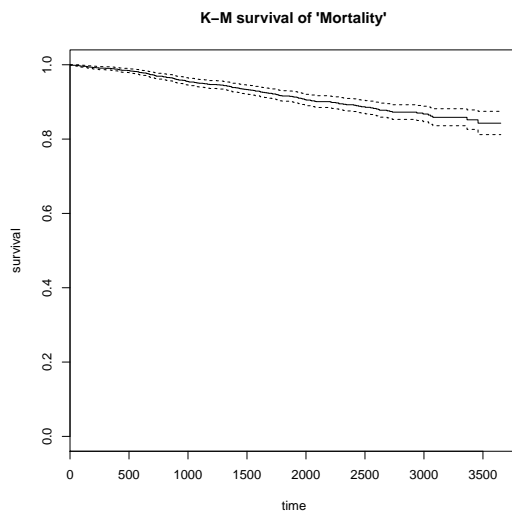

Figure 8: Univariate
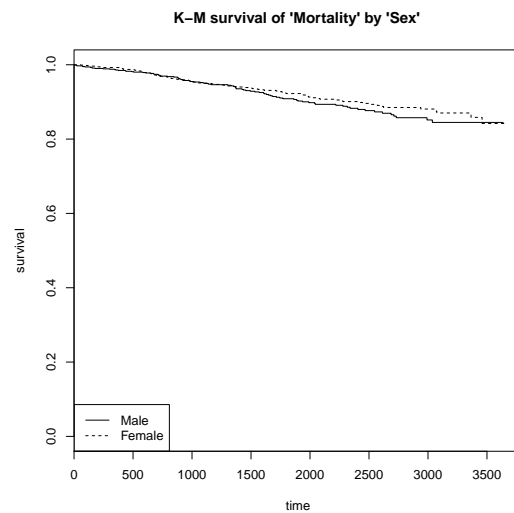
Figure 9: Stratified by *year*

4.2 `createTable`

`createTable` function, applied to an object of `compareGroups` class, returns tables with descriptives that can be displayed on-screen or exported to CSV, LaTeX or HTML.

```
> res<-compareGroups(year ~ age + sex + smoker + triglyc , method= c(triglyc=2),
+        data=regicor)
> restab<-createTable(res)
```

Two tables are created with the `createTable` function: one with the descriptives and the other with the available data. The `print` command print applied to an object of class `createTable` returns one or both tables:

```
> print(restab,which.table='descr')
```

```
--------Summary descriptives table by 'Recruitment year'---------

-----------------------------------------------------------------------------------
                               1995            2000            2005         p.overall
                              N=431           N=786           N=1077
-----------------------------------------------------------------------------------
Age                        54.1 (11.7)      54.3 (11.2)      55.3 (10.6)      0.078
Sex:                                                                          0.506
    Male                   206 (47.8%)      390 (49.6%)      505 (46.9%)
    Female                 225 (52.2%)      396 (50.4%)      572 (53.1%)
Smoking status:                                                             <0.001
    Never smoker           234 (56.4%)      414 (54.6%)      553 (52.2%)
    Current or former < 1y 109 (26.3%)      267 (35.2%)      217 (20.5%)
    Former >= 1y            72 (17.3%)       77 (10.2%)      290 (27.4%)
Triglycerides            94.0 [71.0; 136] 98.0 [72.0; 133] 98.0 [72.0; 139]   0.762
-----------------------------------------------------------------------------------
```

Note that the option "descr" returns descriptives.

```
> print(restab,which.table='avail')
```

```
---Available data----

----------------------------------------------------------------
               [ALL] 1995 2000 2005       method        select
----------------------------------------------------------------
Age            2294   431  786  1077   continuous-normal    ALL
Sex            2294   431  786  1077      categorical       ALL
Smoking status 2233   415  758  1060      categorical       ALL
Triglycerides  2231   403  752  1076 continuous-non-normal  ALL
----------------------------------------------------------------
```

Note that the option "avail" returns the available data, as well as methods and selections.

By default only the descriptives table is shown. Stating "both" in `which.table` options returns both tables.

4.2.1 Dressing up tables

hide If the explanatory variable is dichotomous, one of the categories often is hidden in the results displayed (i.e., if 47.8% are male, obviously 52.2% are female). To hide some category, e.g., Male:

```
> update(restab, hide = c(sex='Male'))
```

```
--------Summary descriptives table by 'Recruitment year'---------

-----------------------------------------------------------------------------------
                               1995            2000            2005         p.overall
                              N=431           N=786           N=1077
-----------------------------------------------------------------------------------
Age                        54.1 (11.7)      54.3 (11.2)      55.3 (10.6)      0.078
Sex: Female                225 (52.2%)      396 (50.4%)      572 (53.1%)      0.506
Smoking status:                                                             <0.001
    Never smoker           234 (56.4%)      414 (54.6%)      553 (52.2%)
    Current or former < 1y 109 (26.3%)      267 (35.2%)      217 (20.5%)
    Former >= 1y            72 (17.3%)       77 (10.2%)      290 (27.4%)
Triglycerides            94.0 [71.0; 136] 98.0 [72.0; 133] 98.0 [72.0; 139]   0.762
-----------------------------------------------------------------------------------
```

Note that the percentage of males is hidden.

hide.no Similarly, as explained above, if the category "no" is to be hidden for all variables:

```
> res<-compareGroups(year ~ age + sex + histhtn + txhtn, data=regicor)
> createTable(res, hide.no='no')
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                          1995       2000       2005     p.overall
                         N=431      N=786      N=1077
_____
Age                   54.1 (11.7) 54.3 (11.2) 55.3 (10.6)  0.078
Sex:                                                        0.506
    Male              206 (47.8%) 390 (49.6%) 505 (46.9%)
    Female            225 (52.2%) 396 (50.4%) 572 (53.1%)
History of hypertension 111 (25.8%) 233 (29.6%) 379 (35.5%) <0.001
Hypertension treatment  71 (16.5%) 127 (16.2%) 230 (22.2%)  0.002
_____
```

Note: 'no', 'No' or 'NO' will produce the same results; the coding is not case sensitive.

digits The number of digits that appear in the results can be changed, e.g:

```
> createTable(res, digits= c(age=2, sex = 3))
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                          1995         2000         2005      p.overall
                         N=431        N=786        N=1077
_____
Age                   54.10 (11.72) 54.34 (11.22) 55.28 (10.63)  0.078
Sex:                                                              0.506
    Male              206 (47.796%) 390 (49.618%) 505 (46.890%)
    Female            225 (52.204%) 396 (50.382%) 572 (53.110%)
History of hypertension:                                         <0.001
    Yes               111 (25.8%)   233 (29.6%)   379 (35.5%)
    No                320 (74.2%)   553 (70.4%)   690 (64.5%)
Hypertension treatment:                                           0.002
    No                360 (83.5%)   659 (83.8%)   804 (77.8%)
    Yes                71 (16.5%)   127 (16.2%)   230 (22.2%)
_____
```

Note that mean and standard deviation has two decimal places for age, while percentage in sex has been set to three decimal places.

type By default categorical variables are summarized by frequencies and percentages. This can be changed by the `type` command:

```
> createTable(res, type=1)
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                          1995       2000       2005     p.overall
                         N=431      N=786      N=1077
_____
Age                   54.1 (11.7) 54.3 (11.2) 55.3 (10.6)  0.078
Sex:                                                        0.506
    Male                 47.8%       49.6%       46.9%
    Female               52.2%       50.4%       53.1%
History of hypertension:                                   <0.001
    Yes                  25.8%       29.6%       35.5%
    No                   74.2%       70.4%       64.5%
Hypertension treatment:                                     0.002
    No                   83.5%       83.8%       77.8%
    Yes                  16.5%       16.2%       22.2%
_____
```

Note that only percentages are displayed.

```
> createTable(res, type=3)
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                          1995        2000        2005     p.overall
                          N=431       N=786       N=1077
_____
Age                     54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Sex:                                                            0.506
    Male                   206         390         505
    Female                 225         396         572
History of hypertension:                                      <0.001
    Yes                    111         233         379
    No                     320         553         690
Hypertension treatment:                                        0.002
    No                     360         659         804
    Yes                     71         127         230
_____
```

Note that only frequencies are displayed.

Values 2 or "NA" return the same results, i.e., the default option.

show.n   If option `show.n` is set to "TRUE" a column with available data for each variable appears in the results:

```
> createTable(res, show.n=TRUE)
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                          1995        2000        2005     p.overall    N
                          N=431       N=786       N=1077
_____
Age                     54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078    2294
Sex:                                                            0.506    2294
    Male                206 (47.8%) 390 (49.6%) 505 (46.9%)
    Female              225 (52.2%) 396 (50.4%) 572 (53.1%)
History of hypertension:                                      <0.001    2286
    Yes                 111 (25.8%) 233 (29.6%) 379 (35.5%)
    No                  320 (74.2%) 553 (70.4%) 690 (64.5%)
Hypertension treatment:                                        0.002    2251
    No                  360 (83.5%) 659 (83.8%) 804 (77.8%)
    Yes                  71 (16.5%) 127 (16.2%) 230 (22.2%)
_____
```

show.descr   If option `show.descr` is set to "FALSE" only p-values are displayed:

```
> createTable(res, show.descr=FALSE)
```

```
--------Summary descriptives table by 'Recruitment year'---------

_____
                       p.overall
_____
Age                      0.078
Sex:
    Male                 0.506
    Female
History of hypertension:
    Yes                 <0.001
    No
Hypertension treatment:
    No                   0.002
    Yes
_____
```

show.all   If `show.all` option is set to "TRUE" a column is displayed with descriptives for all data:

```
> createTable(res, show.all=TRUE)
```

--------Summary descriptives table by 'Recruitment year'---------

	[ALL] N=2294	1995 N=431	2000 N=786	2005 N=1077	p.overall
Age	54.7 (11.0)	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)	0.078
Sex:					0.506
Male	1101 (48.0%)	206 (47.8%)	390 (49.6%)	505 (46.9%)	
Female	1193 (52.0%)	225 (52.2%)	396 (50.4%)	572 (53.1%)	
History of hypertension:					<0.001
Yes	723 (31.6%)	111 (25.8%)	233 (29.6%)	379 (35.5%)	
No	1563 (68.4%)	320 (74.2%)	553 (70.4%)	690 (64.5%)	
Hypertension treatment:					0.002
No	1823 (81.0%)	360 (83.5%)	659 (83.8%)	804 (77.8%)	
Yes	428 (19.0%)	71 (16.5%)	127 (16.2%)	230 (22.2%)	

show.p.overall   If option `show.p.overall` is set to "FALSE" p-values are omitted from the table:

```
> createTable(res, show.p.overall=FALSE)
```

--------Summary descriptives table by 'Recruitment year'---------

	1995 N=431	2000 N=786	2005 N=1077
Age	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)
Sex:			
Male	206 (47.8%)	390 (49.6%)	505 (46.9%)
Female	225 (52.2%)	396 (50.4%)	572 (53.1%)
History of hypertension:			
Yes	111 (25.8%)	233 (29.6%)	379 (35.5%)
No	320 (74.2%)	553 (70.4%)	690 (64.5%)
Hypertension treatment:			
No	360 (83.5%)	659 (83.8%)	804 (77.8%)
Yes	71 (16.5%)	127 (16.2%)	230 (22.2%)

show.p.trend   If the response variable has more than two categories a p-value for trend can be calculated. Results are displayed if the `show.p.trend` option is set to "TRUE":

```
> createTable(res, show.p.trend=TRUE)
```

--------Summary descriptives table by 'Recruitment year'---------

	1995 N=431	2000 N=786	2005 N=1077	p.overall	p.trend
Age	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)	0.078	0.032
Sex:				0.506	0.544
Male	206 (47.8%)	390 (49.6%)	505 (46.9%)		
Female	225 (52.2%)	396 (50.4%)	572 (53.1%)		
History of hypertension:				<0.001	<0.001
Yes	111 (25.8%)	233 (29.6%)	379 (35.5%)		
No	320 (74.2%)	553 (70.4%)	690 (64.5%)		
Hypertension treatment:				0.002	0.002
No	360 (83.5%)	659 (83.8%)	804 (77.8%)		
Yes	71 (16.5%)	127 (16.2%)	230 (22.2%)		

Note: The p-value for trend is computed from the Pearson test when row-variable is normal and from the Spearman test when it is continuous non-normal. If row-variable is of class `Surv`, the test score is computed from a Cox model where the grouping variable is introduced as an integer variable predictor. If the row-variable is categorical, the p-value for trend is computed as $1 - pchisq(cor(as.integer(x), as.integer(y))^2 * (length(x) - 1), 1)$

show.p.mul For a response variable with more than two categories a pairwise comparison of p-values, corrected for multiple comparisons, can be calculated. Results are displayed if the `show.p.mul` option is set to "TRUE":

```
> createTable(res, show.p.mul=TRUE)
```

```
--------Summary descriptives table by 'Recruitment year'---------
```

	1995 N=431	2000 N=786	2005 N=1077	p.overall	p.1995 vs 2000	p.1995 vs 2005	p.2000 vs 20(
Age	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)	0.078	0.930	0.143	0.161
Sex:				0.506	0.794	0.794	0.792
Male	206 (47.8%)	390 (49.6%)	505 (46.9%)				
Female	225 (52.2%)	396 (50.4%)	572 (53.1%)				
History of hypertension:				<0.001	0.169	0.001	0.015
Yes	111 (25.8%)	233 (29.6%)	379 (35.5%)				
No	320 (74.2%)	553 (70.4%)	690 (64.5%)				
Hypertension treatment:				0.002	0.951	0.023	0.004
No	360 (83.5%)	659 (83.8%)	804 (77.8%)				
Yes	71 (16.5%)	127 (16.2%)	230 (22.2%)				

Note: Tukey method is used when explanatory variable is normal-distributed and Benjamini & Hochberg [1] method otherwise.

show.ratio If response variable is dichotomous or has been defined as class `survival` (see Section 3.1), Odds Ratios and Hazard Ratios can be displayed in the results by stating "TRUE" at the show.ratio option:

```
> createTable(update(res, subset= year!=2005), show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Recruitment year'---------
```

	1995 N=431	2000 N=786	OR	p.overall
Age	54.1 (11.7)	54.3 (11.2)	1.00 [0.99;1.01]	0.729
Sex:				0.583
Male	206 (47.8%)	390 (49.6%)	Ref.	
Female	225 (52.2%)	396 (50.4%)	0.93 [0.73;1.18]	
History of hypertension:				0.169
Yes	111 (25.8%)	233 (29.6%)	Ref.	
No	320 (74.2%)	553 (70.4%)	0.82 [0.63;1.07]	
Hypertension treatment:				0.951
No	360 (83.5%)	659 (83.8%)	Ref.	
Yes	71 (16.5%)	127 (16.2%)	0.98 [0.71;1.35]	

Note that category "2005" of the response variable has been omitted in order to have only two categories (i.e., a dichotomous variable). No Odds Ratios would be calculated if response variable has more than two categories.

```
> createTable(compareGroups(tcv ~  age + sex, data=regicor), show.ratio=TRUE)
```

```
--------Summary descriptives table by 'Cardiovascular'---------


----------------------------------------------------------------------
              No event        Event          HR         p.overall
              N=2071          N=92
----------------------------------------------------------------------
Age         54.6 (11.1)    57.5 (11.0)    1.02 [1.00;1.04]     0.011
Sex:                                                           0.696
    Male    996 (48.1%)    46 (50.0%)         Ref.
    Female  1075 (51.9%)   46 (50.0%)     0.92 [0.61;1.39]
----------------------------------------------------------------------
```

Note that when response variable is of class `Surv`, Hazard Ratios are calculated instead of Odds Ratios.

digits.ratio   The number of decimal places for Odds/Hazard ratios can be changed by the `digits.ratio` option:

```
> createTable(compareGroups(tcv ~  age + sex, data=regicor), show.ratio=TRUE,
+              digits.ratio= 3)
```

```
--------Summary descriptives table by 'Cardiovascular'---------


----------------------------------------------------------------------------
              No event         Event            HR           p.overall
              N=2071           N=92
----------------------------------------------------------------------------
Age         54.6 (11.1)     57.5 (11.0)    1.022 [1.003;1.041]     0.011
Sex:                                                               0.696
    Male    996 (48.1%)     46 (50.0%)          Ref.
    Female  1075 (51.9%)    46 (50.0%)     0.922 [0.613;1.387]
----------------------------------------------------------------------------
```

4.2.2   Combining tables by row

Tables made with the same response variable can be combined by row:

```
> restab1 <- createTable(compareGroups(year~ age + sex, data=regicor))
> restab2 <- createTable(compareGroups(year~ smoker + triglyc, data=regicor))
> rbind("Epidemiological"=restab1,"History"=restab2)
```

```
--------Summary descriptives table by 'Recruitment year'---------


-------------------------------------------------------------------------
                              1995        2000        2005      p.overall
                              N=431       N=786       N=1077
-------------------------------------------------------------------------
Epidemiological:
    Age                    54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
    Sex:                                                          0.506
        Male               206 (47.8%) 390 (49.6%) 505 (46.9%)
        Female             225 (52.2%) 396 (50.4%) 572 (53.1%)
History:
    Smoking status:                                              <0.001
        Never smoker       234 (56.4%) 414 (54.6%) 553 (52.2%)
        Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%)
        Former >= 1y       72 (17.3%)  77 (10.2%)  290 (27.4%)
    Triglycerides          114 (74.4)  114 (70.7)  117 (76.0)     0.582
-------------------------------------------------------------------------
```

Note how variables are grouped under "Epidemiological" and "History" because of an epigraph defined in the `rbind` command in the example.

The resulting object is of class `rbind.createTable`, which can be subset but not updated. It inherits the class 'createTable'. Therefore, columns and other arguments from the `createTable` function cannot be modified:

To select only Age and Triglycerides:

```
> rbind("Epidemiological"=restab1,"History"=restab2)[c(1,4)]
```

```
--------Summary descriptives table by 'Recruitment year'---------


_____
                 1995        2000        2005      p.overall
                N=431       N=786       N=1077
_____
Epidemiological:
    Age          54.1 (11.7) 54.3 (11.2) 55.3 (10.6)   0.078
History:
    Triglycerides 114 (74.4)  114 (70.7)  117 (76.0)    0.582
_____
```

To change the order:

```
> rbind("Epidemiological"=restab1,"History"=restab2)[c(4,3,2,1)]
```

```
--------Summary descriptives table by 'Recruitment year'---------


_____
                              1995        2000        2005    p.overall
                             N=431       N=786       N=1077
_____
History:
    Triglycerides             114 (74.4)  114 (70.7)  117 (76.0)   0.582
    Smoking status:                                                <0.001
        Never smoker          234 (56.4%) 414 (54.6%) 553 (52.2%)
        Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%)
        Former >= 1y          72 (17.3%)  77 (10.2%)  290 (27.4%)
Epidemiological:
    Sex:                                                           0.506
        Male                  206 (47.8%) 390 (49.6%) 505 (46.9%)
        Female                225 (52.2%) 396 (50.4%) 572 (53.1%)
    Age                       54.1 (11.7) 54.3 (11.2) 55.3 (10.6)  0.078
_____
```

4.2.3 Combining tables by column

Columns from tables built with the same explanatory and response variables but done with a different subset (i.e. ALL, Male and Female) can be combined:

```
> res<-compareGroups(year ~ age + smoker + sbp + histhtn +triglyc , data=regicor)
> alltab <- createTable(res,  show.p.overall = FALSE)
> femaletab <- createTable(update(res,subset=sex=='Female'), show.p.overall = FALSE)
> maletab <- createTable(update(res,subset=sex=='Male'), show.p.overall = FALSE)
> cbind("ALL"=alltab,"FEMALE"=femaletab,"MALE"=maletab)

-------Summary descriptives table ---------

_____
                                            ALL                               FEMALE                           MALE
                            _____  _____  _____
                              1995        2000        2005      1995        2000        2005      1995        2000        2005
                             N=431       N=786       N=1077     N=225       N=396       N=572      N=206       N=390       N=505
_____
Age                        54.1 (11.7) 54.3 (11.2) 55.3 (10.6) 54.1 (11.7) 54.4 (11.2) 55.2 (10.6) 54.1 (11.8) 54.3 (11.2) 55.4 (10.7)
Smoking status:
    Never smoker           234 (56.4%) 414 (54.6%) 553 (52.2%) 182 (83.1%) 302 (79.3%) 416 (74.0%) 52 (26.5%) 112 (29.7%) 137 (27.5%)
    Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%) 32 (14.6%) 68 (17.8%) 83 (14.8%) 77 (39.3%) 199 (52.8%) 134 (26.9%)
    Former >= 1y            72 (17.3%) 77 (10.2%) 290 (27.4%)  5 (2.28%) 11 (2.89%) 63 (11.2%)  67 (34.2%) 66 (17.5%) 227 (45.6%)
Systolic blood pressure    133 (19.2) 133 (21.3) 129 (19.8)  132 (19.8) 129 (22.6) 127 (20.5)  134 (18.4) 137 (19.3) 132 (18.7)
History of hypertension:
    Yes                    111 (25.8%) 233 (29.6%) 379 (35.5%) 61 (27.1%) 123 (31.1%) 198 (34.8%) 50 (24.3%) 110 (28.2%) 181 (36.2%)
    No                     320 (74.2%) 553 (70.4%) 690 (64.5%) 164 (72.9%) 273 (68.9%) 371 (65.2%) 156 (75.7%) 280 (71.8%) 319 (63.8%)
Triglycerides              114 (74.4) 114 (70.7) 117 (76.0)  97.8 (47.9) 99.7 (55.1) 104 (57.6)  131 (91.5) 128 (81.1) 132 (90.3)
_____
```

With the argument *caption* set to NULL no name is displayed for columns.

```
> cbind(alltab,femaletab,maletab,caption=NULL)

-------Summary descriptives table ---------

_____
                              By Recruitment year               By Recruitment year              By Recruitment year
                            _____  _____  _____
                              1995        2000        2005      1995        2000        2005      1995        2000        2005
                             N=431       N=786       N=1077     N=225       N=396       N=572      N=206       N=390       N=505
_____
Age                        54.1 (11.7) 54.3 (11.2) 55.3 (10.6) 54.1 (11.7) 54.4 (11.2) 55.2 (10.6) 54.1 (11.8) 54.3 (11.2) 55.4 (10.7)
Smoking status:
    Never smoker           234 (56.4%) 414 (54.6%) 553 (52.2%) 182 (83.1%) 302 (79.3%) 416 (74.0%) 52 (26.5%) 112 (29.7%) 137 (27.5%)
    Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%) 32 (14.6%) 68 (17.8%) 83 (14.8%) 77 (39.3%) 199 (52.8%) 134 (26.9%)
    Former >= 1y            72 (17.3%) 77 (10.2%) 290 (27.4%)  5 (2.28%) 11 (2.89%) 63 (11.2%)  67 (34.2%) 66 (17.5%) 227 (45.6%)
```

```
Systolic blood pressure        133 (19.2)  133 (21.3)  129 (19.8)   132 (19.8)  129 (22.6)  127 (20.5)   134 (18.4)  137 (19.3)  132 (18.7)
History of hypertension:
    Yes                       111 (25.8%) 233 (29.6%) 379 (35.5%)   61 (27.1%) 123 (31.1%) 198 (34.8%)   50 (24.3%) 110 (28.2%) 181 (36.2%)
    No                        320 (74.2%) 553 (70.4%) 690 (64.5%)  164 (72.9%) 273 (68.9%) 371 (65.2%)  156 (75.7%) 280 (71.8%) 319 (63.8%)
Triglycerides                 114 (74.4)  114 (70.7)  117 (76.0)   97.8 (47.9) 99.7 (55.1) 104 (57.6)   131 (91.5)  128 (81.1)  132 (90.3)
--------------------------------------------------------------------------------------------------------------------------------------------
```

By default the name of the table is displayed for each set of columns.

> cbind(alltab,femaletab,maletab)

--------Summary descriptives table --------

	alltab			femaletab			maletab		
	1995 N=431	2000 N=786	2005 N=1077	1995 N=225	2000 N=396	2005 N=572	1995 N=206	2000 N=390	2005 N=505
Age	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)	54.1 (11.7)	54.4 (11.2)	55.2 (10.6)	54.1 (11.8)	54.3 (11.2)	55.4 (10.7)
Smoking status:									
Never smoker	234 (56.4%)	414 (54.6%)	553 (52.2%)	182 (83.1%)	302 (79.3%)	416 (74.0%)	52 (26.5%)	112 (29.7%)	137 (27.5%)
Current or former < 1y	109 (26.3%)	267 (35.2%)	217 (20.5%)	32 (14.6%)	68 (17.8%)	83 (14.8%)	77 (39.3%)	199 (52.8%)	134 (26.9%)
Former >= 1y	72 (17.3%)	77 (10.2%)	290 (27.4%)	5 (2.28%)	11 (2.89%)	63 (11.2%)	67 (34.2%)	66 (17.5%)	227 (45.6%)
Systolic blood pressure	133 (19.2)	133 (21.3)	129 (19.8)	132 (19.8)	129 (22.6)	127 (20.5)	134 (18.4)	137 (19.3)	132 (18.7)
History of hypertension:									
Yes	111 (25.8%)	233 (29.6%)	379 (35.5%)	61 (27.1%)	123 (31.1%)	198 (34.8%)	50 (24.3%)	110 (28.2%)	181 (36.2%)
No	320 (74.2%)	553 (70.4%)	690 (64.5%)	164 (72.9%)	273 (68.9%)	371 (65.2%)	156 (75.7%)	280 (71.8%)	319 (63.8%)
Triglycerides	114 (74.4)	114 (70.7)	117 (76.0)	97.8 (47.9)	99.7 (55.1)	104 (57.6)	131 (91.5)	128 (81.1)	132 (90.3)

NOTE: The resulting object is of class `cbind.createTable` and inherits also the class `createTable`. This cannot be updated. It can be nicely printed on the R console and also exported to LaTeX but it cannot be exported to CSV or HTML.

4.2.4 `createTable` miscellaneous

print By default only the table with the descriptives is printed. With the `which.table` command it can be changed: '`avail`' returns data available and '`both`' returns both tables:

```
> print(createTable(compareGroups(year ~ age +  smoker + sbp, data=regicor)),
+         which.table='both')


--------Summary descriptives table by 'Recruitment year'---------

_____
                           1995         2000         2005     p.overall
                          N=431        N=786        N=1077
_____
Age                      54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Smoking status:                                                 <0.001
    Never smoker          234 (56.4%) 414 (54.6%) 553 (52.2%)
    Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%)
    Former >= 1y           72 (17.3%)  77 (10.2%)  290 (27.4%)
Systolic blood pressure   133 (19.2)  133 (21.3)  129 (19.8)    <0.001
_____



---Available data----

_____
                      [ALL] 1995 2000 2005       method      select
_____
Age                   2294  431  786  1077 continuous-normal  ALL
Smoking status        2233  415  758  1060     categorical    ALL
Systolic blood pressure 2280 428  775  1077 continuous-normal  ALL
_____
```

With the `print` command setting `nmax` option = FALSE, the total maximum "n" in the available data is omitted in the first row.

```
> print(createTable(compareGroups(year ~ age +  smoker + sbp, data=regicor)), nmax=FALSE)


--------Summary descriptives table by 'Recruitment year'---------

_____
                           1995         2000         2005     p.overall
_____
Age                      54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Smoking status:                                                 <0.001
    Never smoker          234 (56.4%) 414 (54.6%) 553 (52.2%)
    Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%)
    Former >= 1y           72 (17.3%)  77 (10.2%)  290 (27.4%)
Systolic blood pressure   133 (19.2)  133 (21.3)  129 (19.8)    <0.001
_____
```

summary returns the same table as that generated with `print` command setting `which.table='avail'`:

```
> summary(createTable(compareGroups(year ~ age +  smoker + sbp, data=regicor)))
```

```
---Available data----


--------------------------------------------------------------
                   [ALL] 1995 2000 2005       method      select
--------------------------------------------------------------
Age                2294  431  786  1077 continuous-normal  ALL
Smoking status     2233  415  758  1060     categorical    ALL
Systolic blood pressure 2280  428  775  1077 continuous-normal  ALL
--------------------------------------------------------------
```

update   An object of class `createTable` can be updated:

```
> res<-compareGroups(year ~ age +  smoker + sbp + txhtn, data=regicor)
> restab<-createTable(res, type=1, show.ratio=TRUE )
> restab
```

```
--------Summary descriptives table by 'Recruitment year'---------


_____
                         1995         2000         2005      p.overall
                        N=431        N=786        N=1077
_____
Age                  54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Smoking status:                                            <0.001
    Never smoker         56.4%        54.6%        52.2%
    Current or former < 1y  26.3%        35.2%        20.5%
    Former >= 1y         17.3%        10.2%        27.4%
Systolic blood pressure  133 (19.2)  133 (21.3)  129 (19.8)   <0.001
Hypertension treatment:                                     0.002
    No                   83.5%        83.8%        77.8%
    Yes                  16.5%        16.2%        22.2%
_____
```

```
> update(restab, show.n=TRUE)
```

```
--------Summary descriptives table by 'Recruitment year'---------


_____
                         1995         2000         2005      p.overall     N
                        N=431        N=786        N=1077
_____
Age                  54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078      2294
Smoking status:                                            <0.001      2233
    Never smoker         56.4%        54.6%        52.2%
    Current or former < 1y  26.3%        35.2%        20.5%
    Former >= 1y         17.3%        10.2%        27.4%
Systolic blood pressure  133 (19.2)  133 (21.3)  129 (19.8)   <0.001      2280
Hypertension treatment:                                     0.002      2251
    No                   83.5%        83.8%        77.8%
    Yes                  16.5%        16.2%        22.2%
_____
```

In just one statement it is possible to update an object of class `compareGroups` and `createTable`:

```
> update(restab, x = update(res, subset=c(sex=='Male')), show.n=TRUE)
```

```
--------Summary descriptives table by 'Recruitment year'---------


_____
                         1995         2000         2005      p.overall     N
                        N=206        N=390        N=505
_____
Age                  54.1 (11.8) 54.3 (11.2) 55.4 (10.7)    0.212      1101
Smoking status:                                            <0.001      1071
    Never smoker         26.5%        29.7%        27.5%
    Current or former < 1y  39.3%        52.8%        26.9%
    Former >= 1y         34.2%        17.5%        45.6%
Systolic blood pressure  134 (18.4)  137 (19.3)  132 (18.7)   0.002      1098
Hypertension treatment:                                     <0.001      1078
    No                   85.0%        87.7%        77.2%
    Yes                  15.0%        12.3%        22.8%
_____
```

Note that the `compareGroups` object (*res*) is updated, selecting only 'Male' participants, and the `createTable` object (*restab*) is updated to add a column with the maximum available data for each explanatory variable.

subsetting   Objects from `createTable` function can also be subsetted using "[":

```
> createTable(compareGroups(year~ age + sex + smoker +sbp, data=regicor))


--------Summary descriptives table by 'Recruitment year'---------

_____
                             1995        2000        2005     p.overall
                            N=431       N=786       N=1077
_____
Age                      54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Sex:                                                             0.506
    Male                 206 (47.8%) 390 (49.6%) 505 (46.9%)
    Female               225 (52.2%) 396 (50.4%) 572 (53.1%)
Smoking status:                                                 <0.001
    Never smoker         234 (56.4%) 414 (54.6%) 553 (52.2%)
    Current or former < 1y 109 (26.3%) 267 (35.2%) 217 (20.5%)
    Former >= 1y          72 (17.3%)  77 (10.2%) 290 (27.4%)
Systolic blood pressure   133 (19.2)  133 (21.3)  129 (19.8)   <0.001
-----------------------------------------------------------------------
```

```
> createTable(compareGroups(year~ age + sex + smoker +sbp, data=regicor))[1:2, ]


--------Summary descriptives table by 'Recruitment year'---------

_____
            1995        2000        2005     p.overall
            N=431       N=786       N=1077
-----------------------------------------------------------------
Age       54.1 (11.7) 54.3 (11.2) 55.3 (10.6)    0.078
Sex:                                             0.506
    Male  206 (47.8%) 390 (49.6%) 505 (46.9%)
    Female 225 (52.2%) 396 (50.4%) 572 (53.1%)
-----------------------------------------------------------------
```

4.3   Exporting tables

Tables can be exported to CSV, HTML or LaTeX:

- export2csv(restab, file="table1"), exports to CSV format
- export2html(restab, file="table1"), exports to HTML format
- export2latex(restab, file="table1"), exports to LaTeX format

File extensions are added automatically. Only the filename is needed (and the path if necessary) without the extension.

4.3.1   General exporting options

which.table   By default only the table with the descriptives is exported. This can be changed with the `which.table` command: `'avail'` exports only available data and `'both'` both tables.

nmax   By default a first row with the maximum "n" for available data (i.e. the number of partici-
pants minus the least missing data) is exported. Stating `nmax = FALSE` this first row is omitted.

sep   Only relevant when table is exported to csv. Stating, for example, `sep = ";"` table will be ex-
ported to csv with columns separated by ";"

4.3.2   Exporting to LaTeX

A special case of exporting is when tables are exported to LaTeX. The function `export2latex` returns
an object with the tex code as a character that can be changed in the ℝ session.

file   If the `file` argument in `export2latex` is missing, the code is printed in the ℝ console. This
can be useful when ℝ code is inserted in a LaTeX document chunk to be processed with `Sweave`.

```
> restab<-createTable(compareGroups(year~ age + sex, data=regicor))
> export2latex(restab)


   \begin{longtable}{lcccc}
   \caption{Summary descriptives table by groups of 'Recruitment year'}\\
   \hline
    &    1995    &    2000    &    2005    & \multirow{2}{*}{ p.overall }\\
 &    N=431    &    N=786    &    N=1077    &                \\

   \hline
   \hline
   \endfirsthead
   \multicolumn{5}{l}{\tablename\ \thetable{} \textit{-- continued from previous page}}\\
   \hline
    &    1995    &    2000    &    2005    & \multirow{2}{*}{ p.overall }\\
 &    N=431    &    N=786    &    N=1077    &                \\

   \hline
   \hline
   \endhead
   \hline
   \multicolumn{5}{l}{\textit{continued on next page}} \\
   \endfoot
   \multicolumn{5}{l}{}  \\
   \endlastfoot
   Age & 54.1 (11.7) & 54.3 (11.2) & 55.3 (10.6) &    0.078    \\
Sex: &            &            &            &    0.506    \\
$\qquad$Male & 206 (47.8\%) & 390 (49.6\%) & 505 (46.9\%) &                \\
$\qquad$Female & 225 (52.2\%) & 396 (50.4\%) & 572 (53.1\%) &                \\

   \hline

   \end{longtable}
```

size   The font size of exported tables can be changed by this option. Possible values are 'tiny', 'script-
size', 'footnotesize', 'small', 'normalsize', 'large', 'Large', 'LARGE','huge', 'Huge' or 'same'. Default
is 'same', which means that font size of the table is the same as specified in the main LaTeX document
where the table will be inserted.

caption   The table caption for descriptives table and available data table. If which.table='both' the first element of 'caption' will be assigned to descriptives table and the second to available data table. Default value is `NULL`, which writes 'Summary descriptives table by groups of "y" for descriptives table and 'Available data by groups of "y" for the available data table.

loc.caption   Table caption location. Possible values are 'top' or 'bottom'. Default value is 'top'.

label   Used to cite tables in a LATEX document. If `which.table`='both' the first element of 'label' will be assigned to the descriptives table and the second to the available data table. Default value is `NULL`, which assigns no label to the table/s.

5   Using GUI

Once the `compareGroups` package is loaded, a Graphical User Interface (GUI) is displayed in response to typing `cGroupsGUI`. The GUI is meant to make it feasible for users who are unfamiliar with ℝ to construct bivariate tables.

In this section we illustrate, step by step, how to construct a bivariate table containing descriptives by year from the *regicor* data using the GUI:

Table 2: Summary descriptives table by groups of 'Recruitment year'

|	1995 N=431	2000 N=786	2005 N=1077	p.trend
Age	54.1 (11.7)	54.3 (11.2)	55.3 (10.6)	0.032
Gender: Female	52.2%	50.4%	53.1%	0.544
Smoking status:				<0.001
Never smoker	56.4%	54.6%	52.2%	
Current or former < 1y	26.3%	35.2%	20.5%	
Former ≥ 1y	17.3%	10.2%	27.4%	
Systolic blood pressure	132.6 (19.2)	133.0 (21.3)	129.3 (19.8)	<0.001
Diastolic blood pressure	77.0 (10.5)	80.8 (10.3)	79.9 (10.6)	<0.001
History of hypertension: Yes	25.8%	29.6%	35.5%	<0.001
Hypertension treatment: Yes	16.5%	16.2%	22.2%	0.002
Total cholesterol	223.5 (43.2)	223.6 (44.5)	212.9 (46.4)	<0.001
HDL cholesterol	52.0 (14.5)	52.6 (15.8)	53.3 (14.2)	0.126
Triglycerides	92.0 [70.0; 131.0]	97.0 [72.0; 132.0]	93.0 [70.0; 132.0]	0.723
LDL cholesterol	151.7 (38.4)	149.0 (38.6)	136.3 (39.7)	<0.001
History of hyperchol.: Yes	22.5%	33.2%	33.2%	<0.001
Cholesterol treatment: Yes	6.5%	8.8%	12.8%	<0.001
Height (cm)	163.5 (9.2)	162.0 (9.4)	163.3 (9.0)	0.527
Weight (Kg)	72.3 (12.6)	73.8 (14.0)	73.6 (13.9)	0.185
Body mass index	27.0 (4.1)	28.1 (4.6)	27.6 (4.6)	0.300
Physical activity (Kcal/week)	390.0 [225.5; 617.0]	347.5 [185.1; 573.7]	261.8 [126.6; 442.6]	<0.001
Physical component	49.3 (8.1)	49.0 (9.6)	50.1 (8.9)	0.043
Mental component	49.2 (11.3)	48.9 (11.0)	46.9 (10.8)	<0.001

Step 1. Browse for and select the data to be loaded. Valid file types include SPSS or ℝ format, CSV plain text file or a *data.frame* already existing in the Workspace. By default, the *regicor* example data is loaded when the GUI is opened.
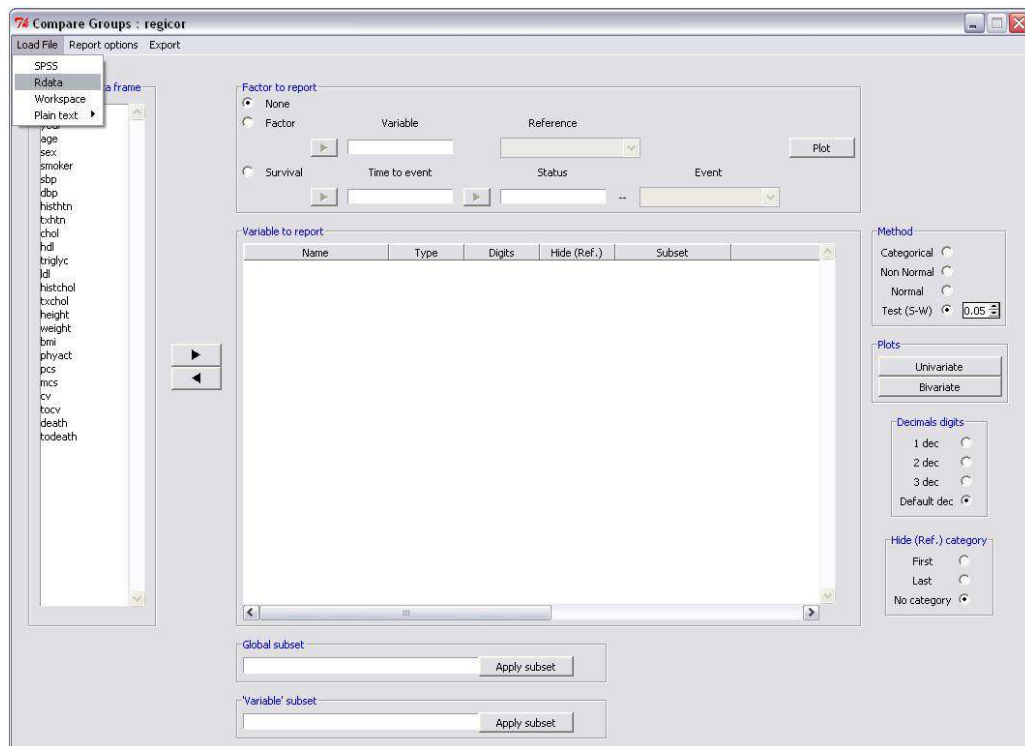

Step 2. Choose the variables to be described (row-variables).
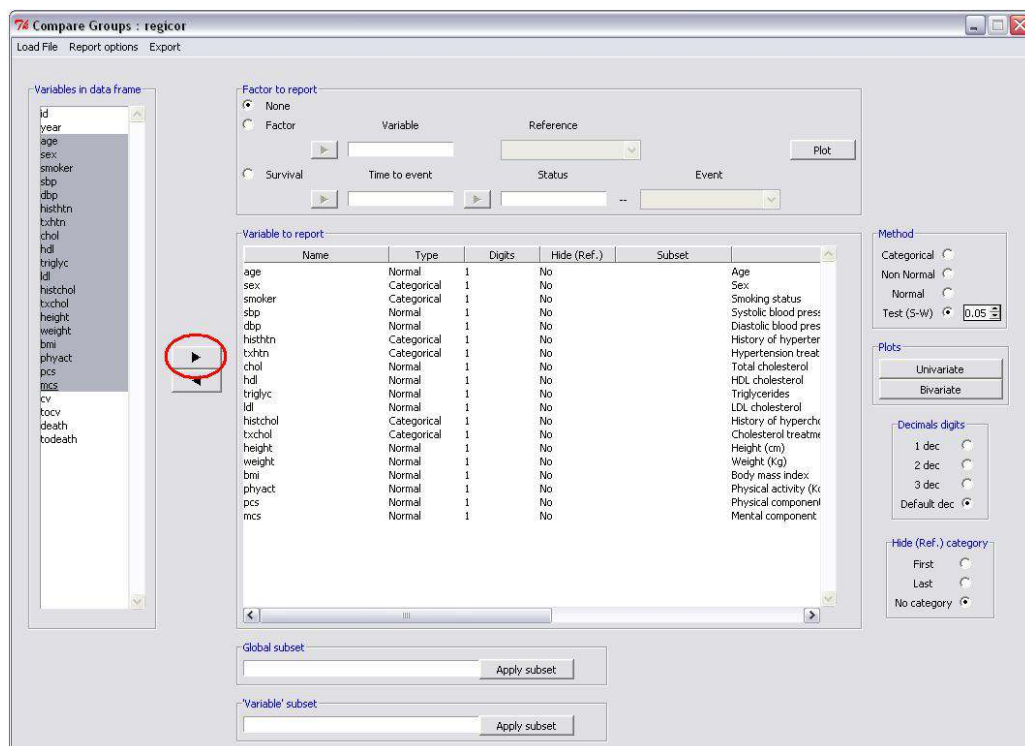
Step 3. If descriptives by year are desired (for example), move the variable *year* to the GUI top frame, making it the factor variable. To report descriptives for the whole sample (i.e., no groups), click on the 'none' button.
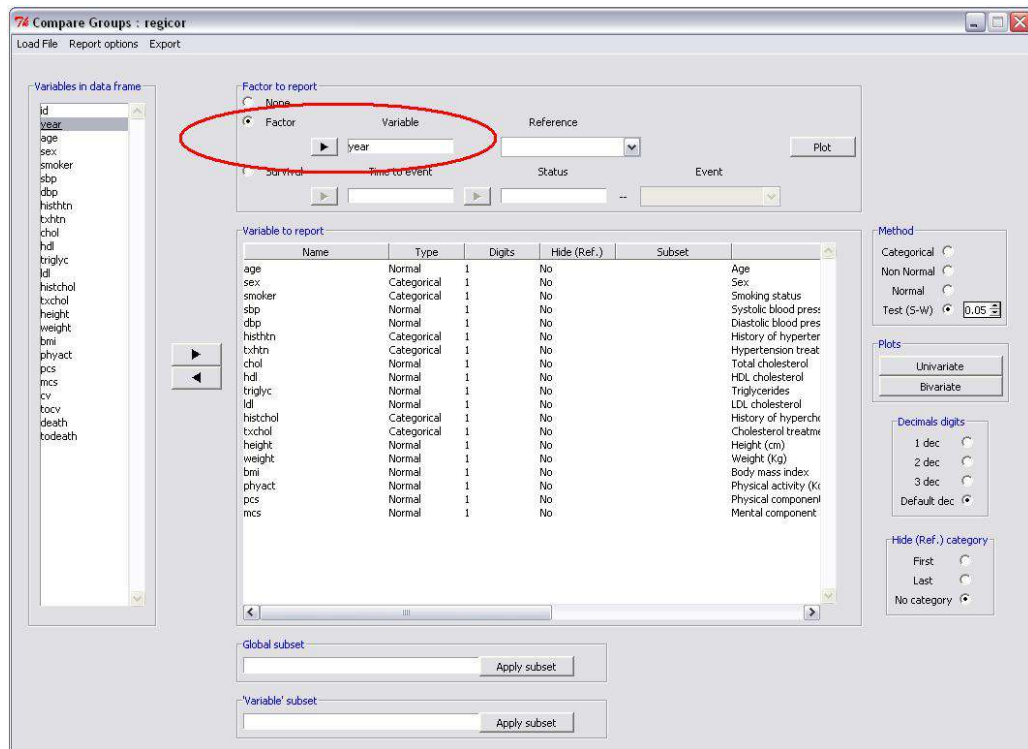
Step 4. It is possible to hide the first, last or no categories of a categorical row-variable. In this example, 'no' levels will be hidden for History of hypertension, Hypertension treatment, History of hypercholesterolemia and Cholesterol treatment; conversely, all categories will be shown for smoking.
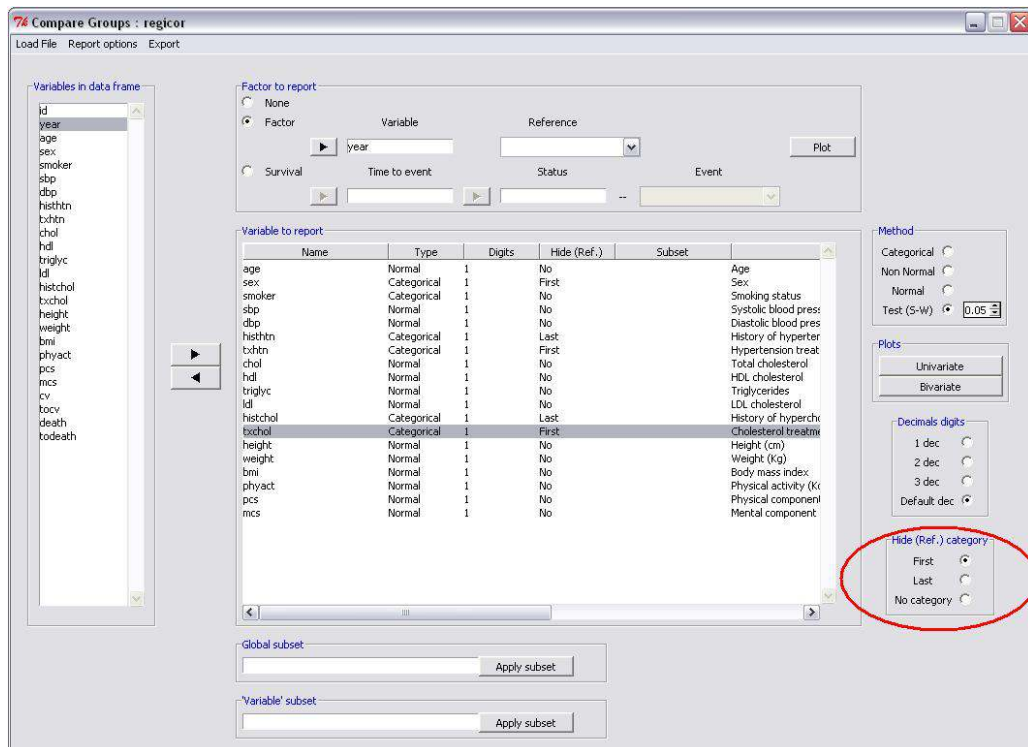
Step 5. For each continuous variable, it is possible to specify whether to treat it as normal or non-normal or to transform a numerical variable into a categorical one. This last option can be interesting if a categorical variable has been coded as numerical. By default, all continuous variables are treated as normal. In this example, triglycerides and physical activity will be treated as non-normal, i.e., median and quartiles will be reported instead of mean and standard deviation.
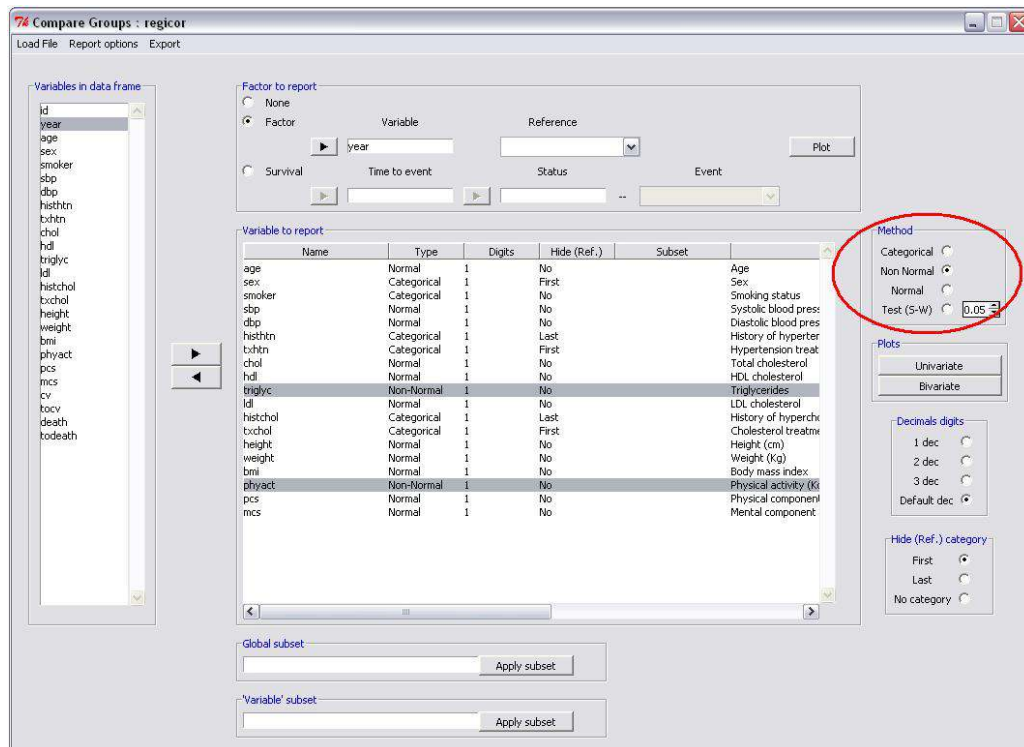
Step 6. For each row-variable, it is possible to select a subset of individuals from the data set to be included. In this example, descriptives of cholesterol, HDL-cholesterol, triglycerides and LDL-cholesterol will be reported only for non-treated individuals. Also, it is possible to specify criteria to select a subset of individuals to be included for all row-variables: type the logical condition (selection criteria of individuals) on the 'Global subset' window instead of 'Variable subset'.
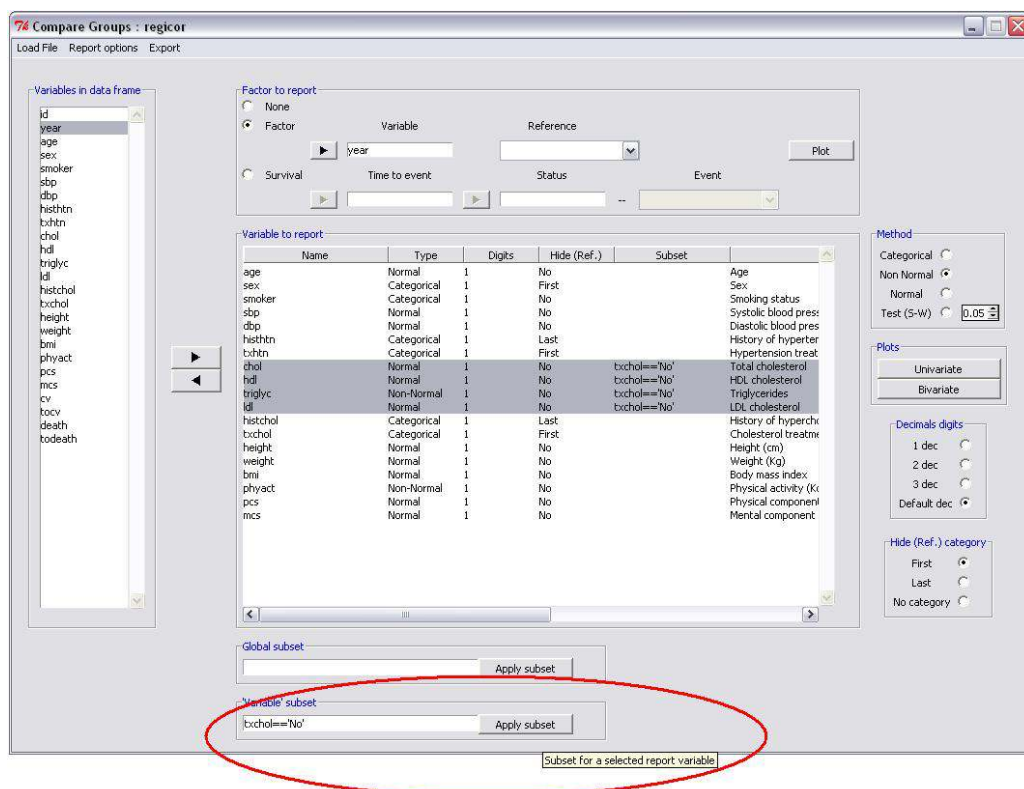
Step 7. Some bivariate table characteristics can be set by clicking on 'Report options' from the main menu, such as to report descriptives (mean, frequencies, medians, etc.), display the p-trend, and show only relative frequencies.
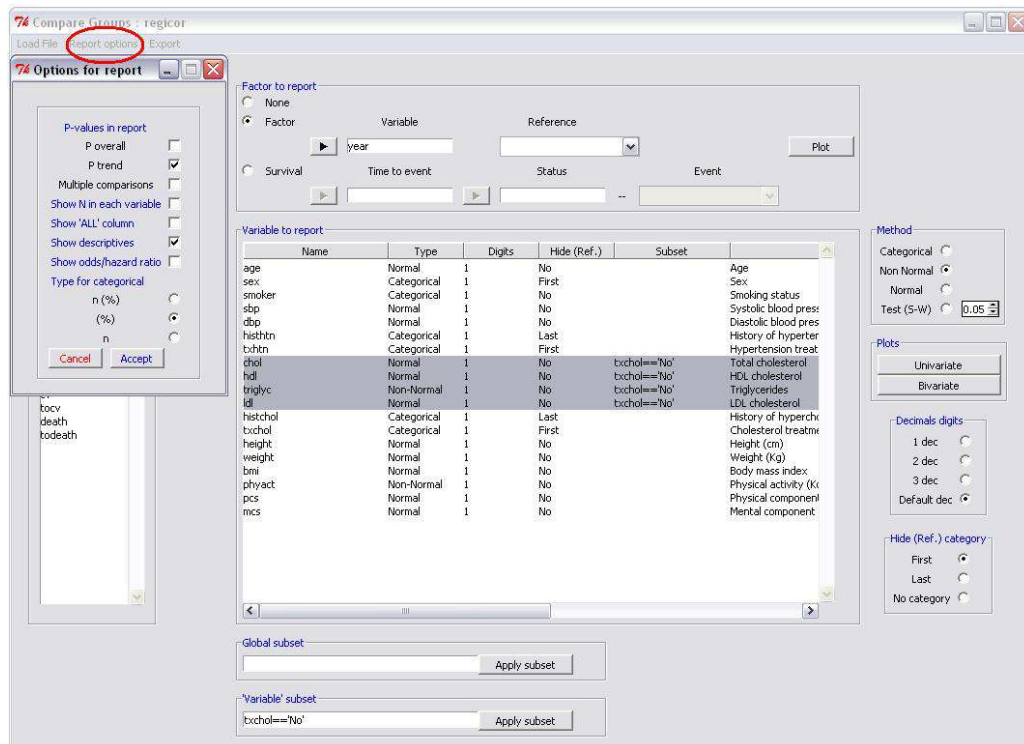
Step 8. Finally, specify the bivariate table format (LaTeX, CVS plain text or HTML). Clicking on 'print' will then display the bivariate table, as well as a summary (available data, etc.), on the ⓡ console. The table can also be exported to the file formats listed.
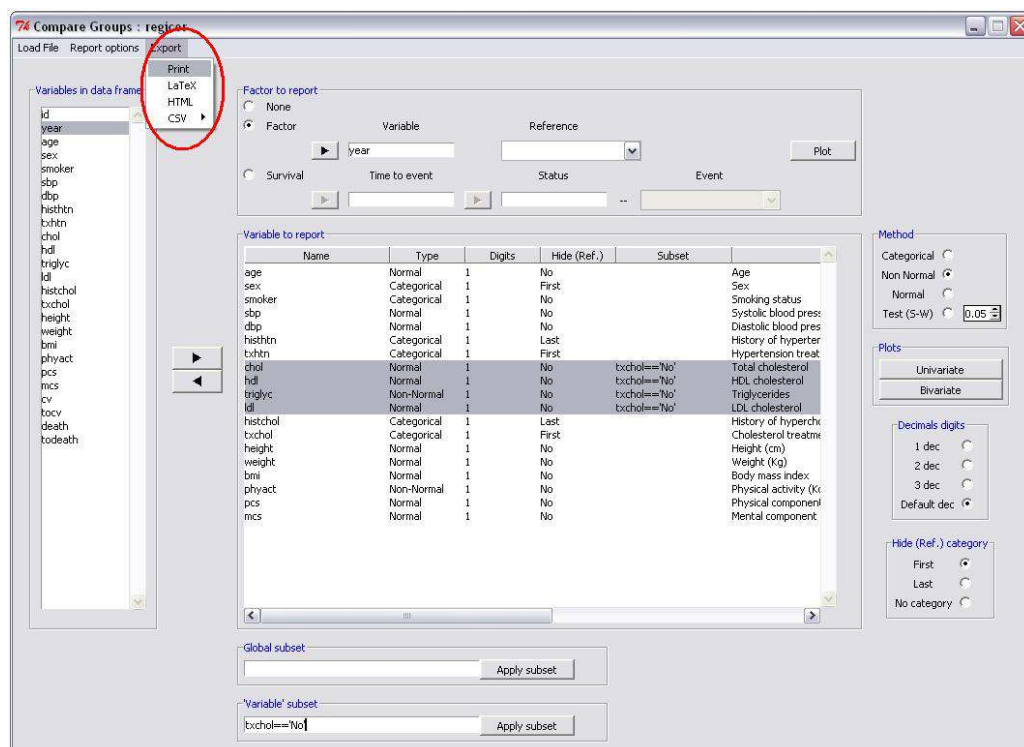
5.1   Computing Odds Ratio

For a case-control study, it may be necessary to report the Odds Ratio between cases and controls for each variable. The table below contains Odds Ratios for each row-variable by cardiovascular disease status.

Table 3: Summary descriptives table by groups of 'Cardiovascular event'

	OR
Recruitment year:	
1995	Ref.
2000	1.90 [0.96;4.12]
2005	1.84 [0.96;3.92]
Age	1.02 [1.00;1.04]
Gender: Female	0.93 [0.61;1.41]
Smoking status:	
Never smoker	Ref.
Current or former < 1y	2.75 [1.77;4.32]
Former ≥ 1y	0.58 [0.25;1.19]
Systolic blood pressure	1.02 [1.01;1.02]
Diastolic blood pressure	1.03 [1.01;1.05]
History of hypertension: Yes	1.54 [1.00;2.36]
Hypertension treatment: Yes	1.37 [0.82;2.21]
Total cholesterol	1.00 [1.00;1.01]
HDL cholesterol	0.99 [0.98;1.01]
Triglycerides	1.00 [1.00;1.00]
LDL cholesterol	1.00 [1.00;1.01]
History of hyperchol.: Yes	0.83 [0.51;1.31]
Cholesterol treatment: Yes	0.61 [0.23;1.30]
Height (cm)	1.00 [0.98;1.03]
Weight (Kg)	1.01 [0.99;1.02]
Body mass index	1.02 [0.98;1.07]
Physical activity (Kcal/week)	1.00 [1.00;1.00]
Physical component	0.97 [0.95;1.00]
Mental component	0.99 [0.97;1.00]

    To build this table, as illustrated in the screens below, you would select *cv* variable (cardiovascular disease status) as the factor variable, indicate 'no' category on the 'reference' pull-down menu, and mark 'Show odds/hazard ratio' in the 'Report Options' menu before exporting the table.

Select *cv* variable and define the reference.
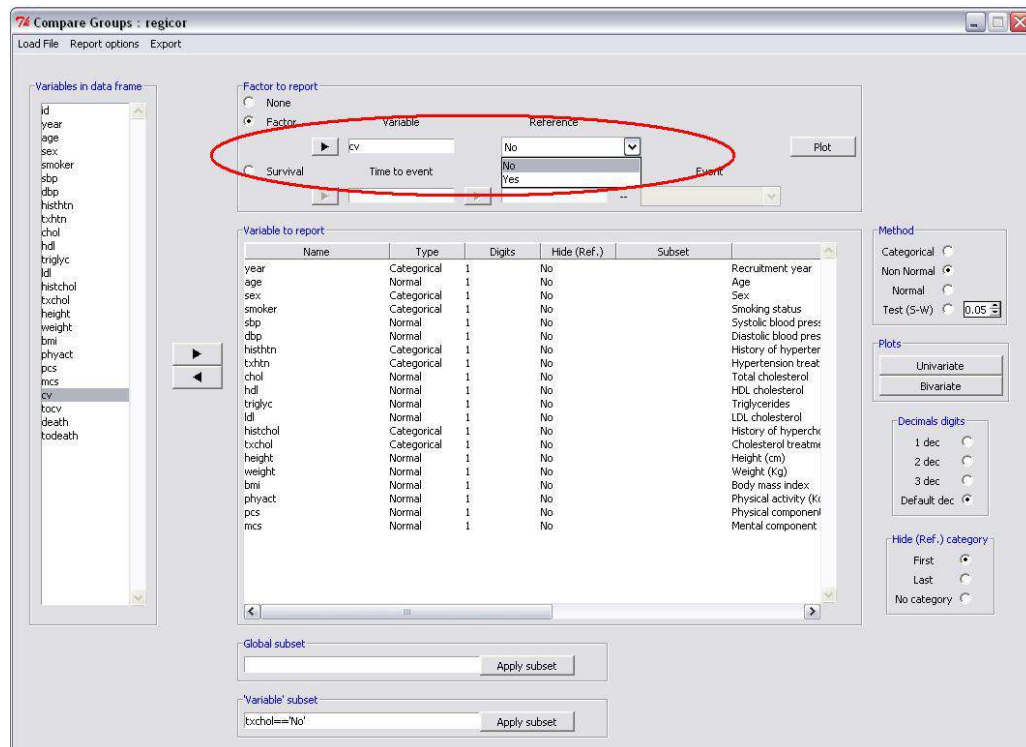
Mark 'Show odds/hazard ratio' in the 'Report options' menu and click on 'Accept'.
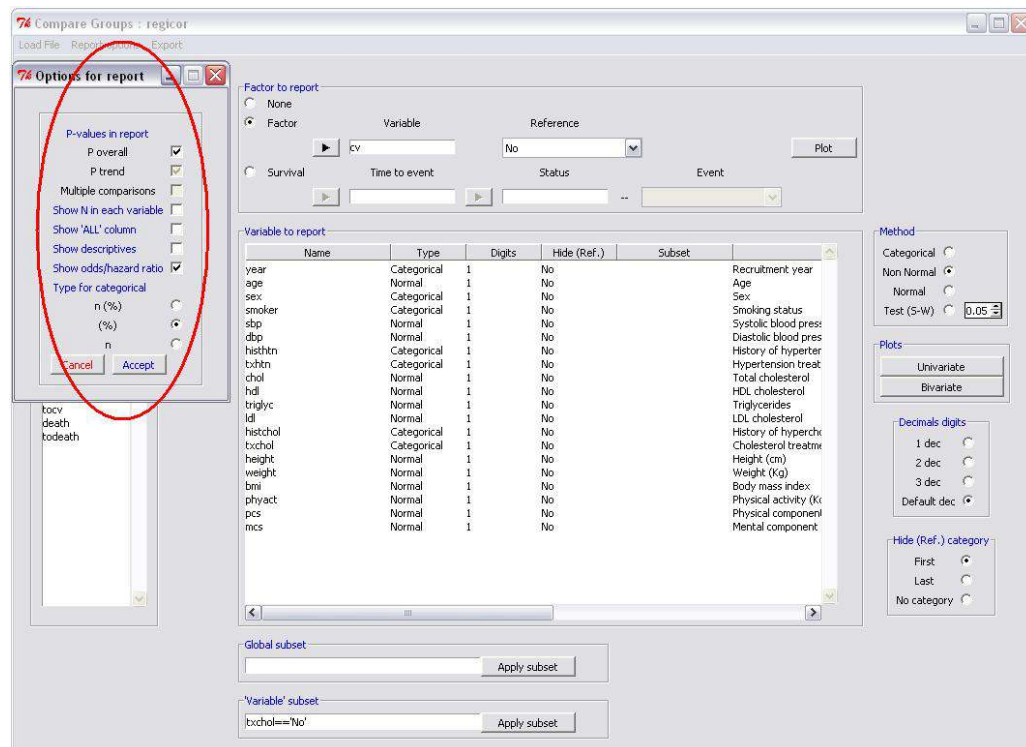
Finally, choose the desired output mode from the 'Export' options.

## 5.2 Computing Hazard Ratio

In a cohort study, it may be more informative to compute Hazard Ratios taking into account time-to-event.

Table 4: Summary descriptives table by groups of 'Cardiovascular event'

	HR
Recruitment year:	
1995	Ref.
2000	1.95 [0.96;3.93]
2005	1.82 [0.92;3.59]
Age	1.02 [1.00;1.04]
Gender: Female	0.92 [0.61;1.39]
Smoking status:	
Never smoker	Ref.
Current or former < 1y	2.67 [1.74;4.11]
Former ≥ 1y	0.55 [0.26;1.18]
Systolic blood pressure	1.02 [1.01;1.02]
Diastolic blood pressure	1.03 [1.01;1.05]
History of hypertension: Yes	1.52 [1.01;2.31]
Hypertension treatment: Yes	1.37 [0.85;2.22]
Total cholesterol	1.00 [1.00;1.01]
HDL cholesterol	0.99 [0.98;1.00]
Triglycerides	1.00 [1.00;1.00]
LDL cholesterol	1.00 [1.00;1.01]
History of hyperchol.: Yes	0.82 [0.52;1.30]
Cholesterol treatment: Yes	0.61 [0.27;1.39]
Height (cm)	1.00 [0.98;1.03]
Weight (Kg)	1.01 [0.99;1.02]
Body mass index	1.02 [0.98;1.07]
Physical activity (Kcal/week)	1.00 [1.00;1.00]
Physical component	0.98 [0.96;1.00]
Mental component	0.99 [0.97;1.00]

To generate this table, select *tocv* variable and *cv*, indicating the time-to-event and the status, respectively, and select the event category for the status variable. Finally, as for Odds Ratios, mark 'Show odds/hazard ratio' in the 'Report Options' menu before exporting the table.

Select *cv* variable and *tocv*.
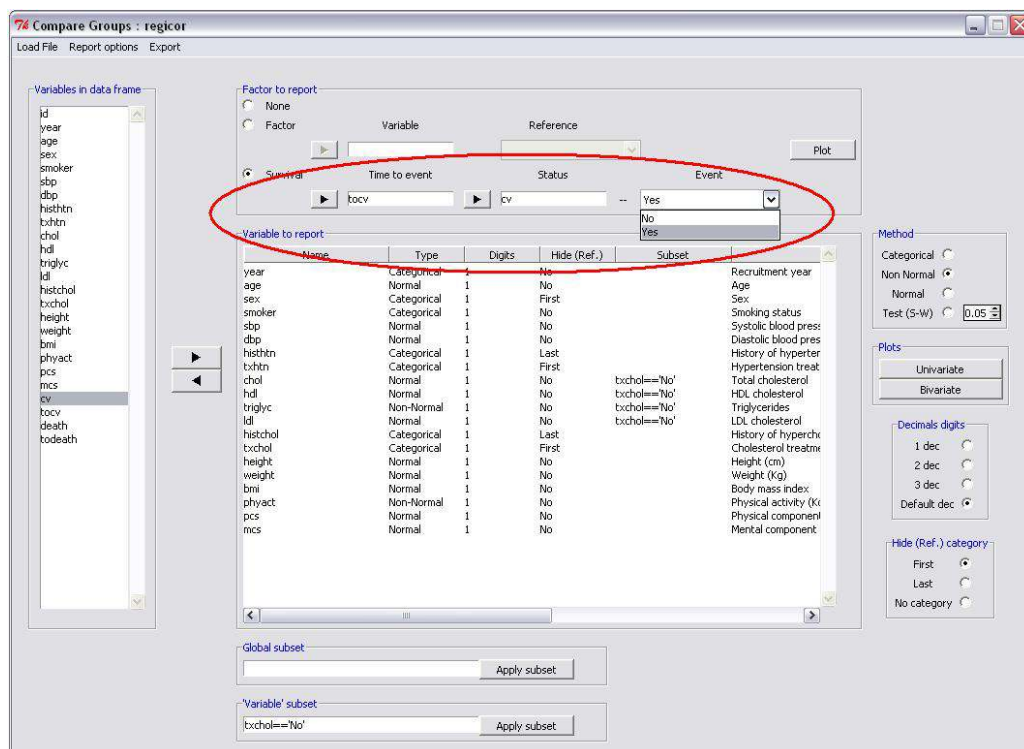
Mark 'Show odds/hazard ratio' in the 'Report options' menu.
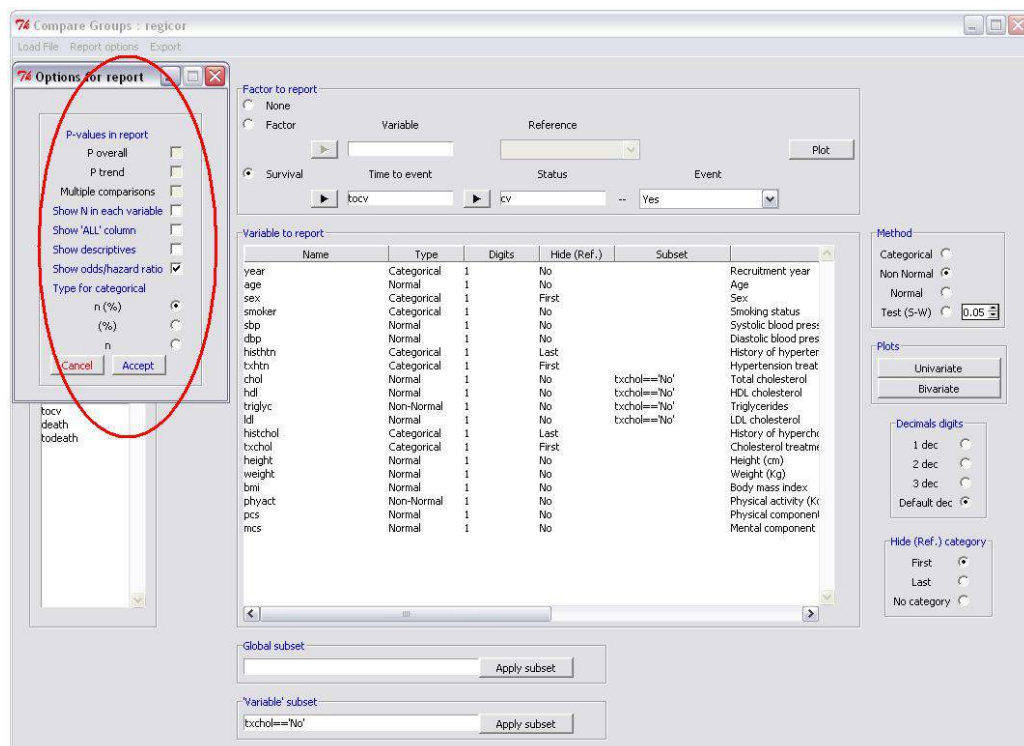
Finally, choose the Export option.

To return to the ⓡ console, just close the GUI window.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57:289–300, 1995.

[2] C. Genolini, B Desgraupes, and Lionel-Riou Franca. *r2lh: R to LaTeX and HTML*, 2011. R package version 0.7.