`compareGroups` Web User Interface (WUI) Tutorial

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

April 15, 2014

[1]Cardiovascular Epidemiology & Genetics group, Inflammatory and Cardiovascular Disease Programme, IMIM, Barcelona

[2]Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), Barcelona, Spain

[3]CIBER Epidemiology and Public Health (CIBERESP), Spain

[4]Statistics Department, University of Barcelona, Spain

isubirana@imim.es, jvila@imim.es, hsrodenas@gmail.com

Contents

1 Introduction

The `compareGropus` package (version ≥ 2.1) incorporates a Web User Interface (WUI) based on `shiny` `R` package [2] (http://shiny.rstudio.com) to use all functionality of the package very easily and quickly for non `R` users.

This application includes almost all the options included in 'type on' version. Also, thanks to the power of `shiny` package, the user can see the results when setting the included variable, the groups, number of decimals, etc almost instantaneously ('reactivity' -see `shiny` manual and examples-). This is very useful to modify and customize the descriptive table before saving it in the desired format saving a lot of time.

In the following sections, we list and describe all the options available in the `Shiny-` `compareGroups` application, and we illustrate how it works with a real example.

2 Menus and options

Once the WUI is called, a web browser is launched with two main parts: in the left hand side there are all the control menu options to read the data frame from different format files, select the variables list to be analysed, specify the number of digits of the descriptives or p-values displayed in the bivariate table, and much more other aspects; while on the right hand side contains the results, i.e, a table

containing a basic summary of all variables, the bivariate table or some plots.

Following, an exhaustive list describing all the options are presented:

2.1 Control menu (left side)

- **START** A panel with all the options to select the data base and to select the analysed variables. Different file formats are allowed; in the present version these are: SPSS, plain text (txt or csv), EXCEL (2000 or 2007) or R (.rda or RData).

- **SETTINGS** A panel with different issues to control for different aspects of the bivariate table:

  - **Type** Each variable can be analysed as normal or non-normal or as categorical. In the first case, mean and standard deviation are displayed; for the second case, medians and quantiles; and for the latest, frequencies and proportions.

  - **Response** Control menu to select which variable indicates the group. Three choices are possible:

    1) **None:** descriptives are performed for the entire data set with no groups.
    2) **Group:** a variable must be selected. By default, only factor variables with five or less categories can be selected. The descriptives will be performed by these groups. When the grouping variable has only 2 categories, Odds Ratio (OR) can be computed.
    3) **Survival:** to perform survival analysis where the response is a right censored variable,i.e., with time variable and a variable indicating the censoring (i.e., whether or not the individual suffered the disease). Also, the disease or case status must be indicated.

  - **Hide** For categorical variables, select the category you want to be hidden. Additionally, in the 'hide no' input text windows you can type the category which represents 'no' in the sense that the category named as indicated for all binary variables are hidden.

  - **Subset** A global subset can be typed affecting all analysed variables. Also you can select a subgroup of individuals for each variable. In any case, a logical expression in R language must be typed:

Operator	R
=	==
or	—
and	&
nor	!

    Note that to indicate the category, you must type the number that appears in the 'VALUES' table on right side of the application instead of its name. For example, for gender, type 1 instead of 'male' and 2 instead of 'female', if male is the first category and female is the second.

  - **OR/HR** Options for Odds Ratio (OR) when response is binary or Hazard Ratios (HR) for right censored response. You can specify the reference category when computing the OR or HR for the categorical variables. For continuous variables, the scale can be changed. This is very useful for variables with a wide range such as total cholesterol, where it makes more sense to display the OR/HR for each 10 units of change rather than a single unit.

- **DISPLAY** In this panel, the user can select which information must be displayed in the resulting bivariate table, and in which format, etc.

  - **Show** What to be displayed in the bivariate table:
    * **ALL:** Descriptives for the entire data set.
    * **p-overall:** p-value comparing the means, medians, proportions or incidence among groups.
    * **Descriptives:** Descriptives (means, medians, proportions, etc.) by groups.

* **p-trend:** p-value for trend. It is supposed that the groups are ordered.
* **OR/HR:** Odds Ratio or Hazard Ratio for binary or time-to-event response (grouping variable), respectively.
* **Available:** Number of individuals with valid data for each variable and each groups.
* **NA category:** For categorical variables, the non-available data is considered as a new category.
* **Pairwise p-value:** When more than two groups are considered, p-values corresponding to 2 by 2 comparisons are performed taking into account multiple testing.
* **Simplify:** Empty categories (with no available data) are removed from the analyses.
    - Format The user can specify how to display the mean and standard deviation, the quantiles and frequencies.
    - Decimals The number of decimals to be displayed in the bivariate table, for descriptives, p-values and for OR / HR.
    - Labels This tab allows to change the 'key' headers of the descriptive table such as 'ALL', 'p-value', etc.

- SAVE A panel with the options to save both the bivariate table or the plots. Depending on the active results tab (right side), i.e., tables or plots, different options are displayed.

2.2 Results panels (right side)

- VALUES A table displaying the name, the label and a basic summary of all the variables contained in the data set.

- INFO A table showing the available data for each value and possibly each group, as well as some other information such as the selection criteria, etc.

- TABLE The bivariate table. It can be displayed in three different formats: HTML, PDF or as it is printed in the R console. Note that, despite of this is the less nice format, it is the faster to be displayed and may be the most useful to see how the table looks like while we modify some options (like the grouping variable) 'on line'.

- PLOT Univariate or bivariate (taking into account the groups if proceeds) plots of selected variable is displayed.

- SNPs Descriptives of SNPs (Single Nucleotide Polymorphims) with appropriate statistics and tests for these genetic variants are displayed. To perform this analysis, only SNPs variables can be selected. A factor response to display genotype frequencies by groups is permitted but not a time-to-event response variable.

3 Example

In this section we illustrate how to analyse a data set.

- **Step 0. Launch the WUI application on a web browser**. To use the WUI locally (and not on a remote server), first load the `compareGroups` package and call the `cGroupsWUI` function:

  ```
  > library(compareGroups)
  > cGroupsWUI()
  ```

- **Step 1. Load the data.** In this example we will load the 'predimed' data set from the PREDIMED study [1] stored in SPSS format (*.sav). These data is already available in the `compareGroups` package.
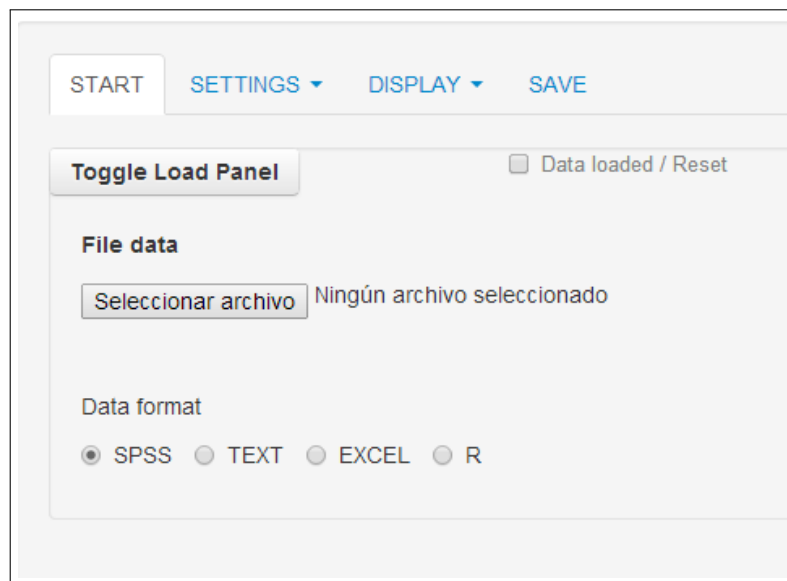
- **Step 2. Select the variables to be analysed.** We will select all variables except the event and the time-to-event variable ('toevent'). They will represent the response and must be removed from the row-variables.

- **Step 3. Select the grouping variable.** Since PREDIMED is a longitudinal study where the main goal is to check whether the mediterranean diet is related to a less incidence of cardiovascular disease, we will take the response as the time to cardiovascular event. To do so, we will select 'Survival' response type, setting the 'toevent' as time variable and 'event' as indicator variable (taking 'yes' as case).

- **Step 4. Take a look at the plots** to see which continuous variables should be treated as normal and which not. By default, `compareGroups` performs a Shapiro-Wilks normality test to decide which variables are normal. But we may want to check the normality assumption graphically.
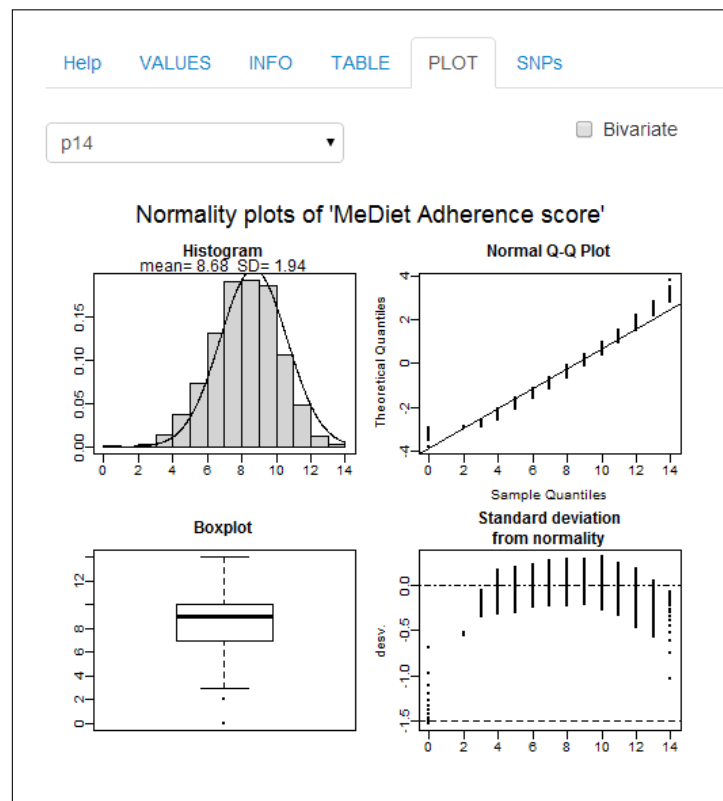
- **Step 5. Set the options.**

  - **Type:** Specify the variables to be treated as normal and the ones as non normal. To specify that all of them will be treated as normal, except 'p14' for which median and quartiles are displayed instead of mean and SD:
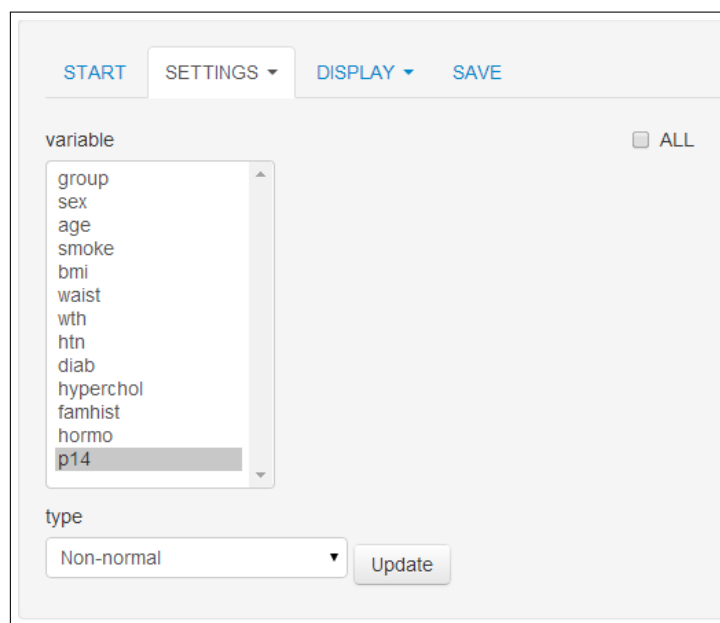
- **Hide:** To hide the 'No' category for the ('yes'/'no') variables such as hypertension, diabetes, etc., we will type 'No' in the 'hide no' input text window. Also, to hide the 'female' category for we will select the 'sex' variable and 'female' category and press 'Update' button afterwards.

- **OR:** Since we have hidden the females, we have to do the same for the HR. Also, for the waist we will change the scale to 5. Doing this, the HR will represent the change in 5 units.

– **Subset:** If we want to do the analysis only for male type 'sex==1' or 'sex==2' in only females must be analysed in the 'global subset' text input window.

- **Step 6. What to be displayed and how.**

  – **Show:** Then, we specify which columns (information) to be displayed in the bivariate table. For instance, if we want to display the Odds Ratio, its p-value, but not the descriptives for the entire cohort p.overall:

– **Format:** Also, we may want to display only the percentages for categorical variables, mean and standard deviation with plus/minus symbol and rounded brackets for first and third quantiles for continuous normal and non-normal variables, respectively:

– **Decimals:** Also, it may be important to report more decimals for the odds ratio. By default, two decimals are reported. If we want to report 3 decimals for all variables:

- **Step 7. Visualize the table:** To see the output, click on TABLE tab on the right panel. The table can be displayed in different formats:

  i) **HTML:**

Var	No event N=6072	Event N=252	HR	p.ratio
Intervention group:				
Control	32.0%	38.5%	Ref.	Ref.
MedDiet + Nuts	33.4%	27.8%	0.66 [0.48;0.89]	0.008
MedDiet + VOO	34.5%	33.7%	0.70 [0.53;0.94]	0.018
Sex: Male	41.6%	59.9%	2.05 [1.59;2.64]	<0.001
Age	66.9±6.14	69.4±6.65	1.06 [1.04;1.09]	<0.001
Smoking:				
Never	62.2%	45.2%	Ref.	Ref.
Current	13.3%	19.4%	1.96 [1.40;2.74]	<0.001
Former	24.5%	35.3%	2.02 [1.53;2.67]	<0.001
Body mass index	30.0±3.81	29.8±3.92	0.99 [0.96;1.02]	0.455
Waist circumference	100±10.6	102±10.6	1.09 [1.03;1.15]	0.003
Waist-to-height ratio	0.63±0.07	0.63±0.07	5.27 [0.83;33.6]	0.079
Hypertension	82.8%	83.3%	1.10 [0.79;1.53]	0.578
Type-2 diabetes	46.8%	63.9%	1.88 [1.46;2.44]	<0.001
Dyslipidemia	72.9%	59.9%	0.62 [0.49;0.80]	<0.001
Family history of premature CHD	22.7%	20.2%	0.93 [0.68;1.26]	0.640
Hormone-replacement therapy	1.77%	0.45%	0.29 [0.04;2.10]	0.223
MeDiet Adherence score	9.00 (7.00;10.0)	8.00 (7.00;10.0)	0.88 [0.83;0.94]	<0.001

  ii) **PDF:**

Table 1: Summary descriptives table by groups of '[to-event ; event = Yes]'

	No event N=6072	Event N=252	HR	p.ratio
Intervention group:				
Control	32.0%	38.5%	Ref.	Ref.
MedDiet + Nuts	33.4%	27.8%	0.66 [0.48;0.89]	0.008
MedDiet + VOO	34.5%	33.7%	0.70 [0.53;0.94]	0.018
Sex: Male	41.6%	59.9%	2.05 [1.59;2.64]	<0.001
Age	66.9±6.14	69.4±6.65	1.06 [1.04;1.09]	<0.001
Smoking:				
Never	62.2%	45.2%	Ref.	Ref.
Current	13.3%	19.4%	1.96 [1.40;2.74]	<0.001
Former	24.5%	35.3%	2.02 [1.53;2.67]	<0.001
Body mass index	30.0±3.81	29.8±3.92	0.99 [0.96;1.02]	0.455
Waist circumference	100±10.6	102±10.6	1.09 [1.03;1.15]	0.003
Waist-to-height ratio	0.63±0.07	0.63±0.07	5.27 [0.83;33.6]	0.079
Hypertension	82.8%	83.3%	1.10 [0.79;1.53]	0.578
Type-2 diabetes	46.8%	63.9%	1.88 [1.46;2.44]	<0.001
Dyslipidemia	72.9%	59.9%	0.62 [0.49;0.80]	<0.001
Family history of premature CHD	22.7%	20.2%	0.93 [0.68;1.26]	0.640
Hormone-replacement therapy	1.77%	0.45%	0.29 [0.04;2.10]	0.223
MeDiet Adherence score	9.00 (7.00;10.0)	8.00 (7.00;10.0)	0.88 [0.83;0.94]	<0.001

iii) `R`:


```
Help    VALUES    INFO    TABLE    PLOT    SNPs


○ HTML   ○ PDF   ● R console


--------Summary descriptives table by '[  toevent ;  event = Yes ]'---------


                                 No event          Event              HR       p.ratio
                                 N=6072            N=252
---------------------------------------------------------------------------------------
Intervention group:
    Control                      32.0%             38.5%            Ref.         Ref.
    MedDiet + Nuts               33.4%             27.8%      0.66 [0.48;0.89]   0.008
    MedDiet + VOO                34.5%             33.7%      0.70 [0.53;0.94]   0.018
Sex: Male                        41.6%             59.9%      2.05 [1.59;2.64]  <0.001
Age                              66.9±6.14         69.4±6.65  1.06 [1.04;1.09]  <0.001
Smoking:
    Never                        62.2%             45.2%            Ref.         Ref.
    Current                      13.3%             19.4%      1.96 [1.40;2.74]  <0.001
    Former                       24.5%             35.3%      2.02 [1.53;2.67]  <0.001
Body mass index                  30.0±3.81         29.8±3.92  0.99 [0.96;1.02]   0.455
Waist circumference              100±10.6          102±10.6   1.09 [1.03;1.15]   0.003
Waist-to-height ratio            0.63±0.07         0.63±0.07  5.27 [0.83;33.6]   0.079
Hypertension                     82.8%             83.3%      1.10 [0.79;1.53]   0.578
Type-2 diabetes                  46.8%             63.9%      1.88 [1.46;2.44]  <0.001
Dyslipidemia                     72.9%             59.9%      0.62 [0.49;0.80]  <0.001
Family history of premature CHD  22.7%             20.2%      0.93 [0.68;1.26]   0.640
Hormone-replacement therapy      1.77%             0.45%      0.29 [0.04;2.10]   0.223
MeDiet Adherence score     9.00 (7.00;10.0) 8.00 (7.00;10.0) 0.88 [0.83;0.94]  <0.001
---------------------------------------------------------------------------------------
```

Note that HTML and PDF formats incorporate an horizontal bar to customize the table size.

- **Step 8. Saving the table.** Finally, once the table contains all the desired figures and in the appropriate format, it can be downloaded and stored in different formats:

  a) PDF: a LaTeX compiler such as MikTex must be installed to build the PDF document with the descriptive table. The user can select the font size and whether table must be placed vertically or horizontally (landscape option).

  b) CSV: a plain text file with columns separated by commas or semicolons. For Windows users, this format is useful since it can be opened by Excel.

  c) HTML: a web browser is opened and the table can be easily copied and pasted to Word, for instance.

  d) TXT: a plain text file which can be opened by any text editor program and which contains the table as in `R` console with a nice format. Once the 'download' button is pressed, the file is automatically stored in your PC/Mac.

References

[1] R. Estruch, E. Ros, J. Salas-Salvadó, MI. Covas, D. Corella, F. Arós, E. Gómez-Gracia, V. Ruiz-Gutiérrez, M. Fiol, J. Lapetra, RM. Lamuela-Raventos, L. Serra-Majem, X. Pintó, J. Basora, MA. Muñoz, JV. Sorlí, JA. Martínez, MA. Martínez-González, and PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a mediterranean diet. *N Engl J Med*, 368(14):1279–90, Apr 2013.

[2] RStudio and Inc. *shiny: Web Application Framework for R*, 2014. R package version 0.9.1.