

compareGroups Web User Interface (WUI) Tutorial

Isaac Subirana^{1,2,3}

January 22, 2015

¹CIBER Epidemiology and Public Health (CIBERESP), Spain

²IMIM - Parc de Salut Mar, Barcelona

³Statistics Department, University of Barcelona, Spain

`isubirana@imim.es`

Contents

1	Introduction	1
2	Menus and options	1
2.1	Control menu (left side)	2
2.2	Results panels (right side)	3
3	Example	4

1 Introduction

The `compareGropus` package (version ≥ 2.1) incorporates a Web User Interface (WUI) based on `shiny` R package [?] (<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 existing 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 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). Also, it is possible to select example data already present in the `compareGroups` package (REGICOR and PREDIMED) and SNPS from `SNPassoc` package.
- **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	&
not	!

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)

- **INFO** A table showing the available data for each variable and possibly each group, as well as some other information such as the selection criteria, etc.
- **VALUES** Using this tab, the user can take a look at the data contained in the data set.
 - **Summary:** A table with the the name, label and a basic summary of all the variables contained in the data set.
 - **Extended:** A table with the raw values in the data set. By default only the first ten rows are displayed, but the user can set the number of rows to be shown and navigate through the data set.
- **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, although this is the less nice format, it is the fastest option to be displayed and, in consequence, may be the most useful way to see how the bivariate table looks like when some options are modified (e.g. grouping variable) ‘on line’.
- **PLOT** Univariate or bivariate (taking into account the groups if proceeds) plots of selected variable.
- **SNPs** Descriptives of SNPs (Single Nucleotide Polymorphisms) with appropriate statistics and tests for these genetic variants. 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.
- **HELP** Use this tab to know more about `compareGroups` and its Web User Interface.
 - **About:** Explanation about both the `compareGroups` R-package and its Web User Interface (WUI) version. Also, it contains the security rules of the data sent by the user when using WUI remotely (from www.comparegroups.eu).
 - **Panel:** Help navigation panel that mimics both the control and result panels.

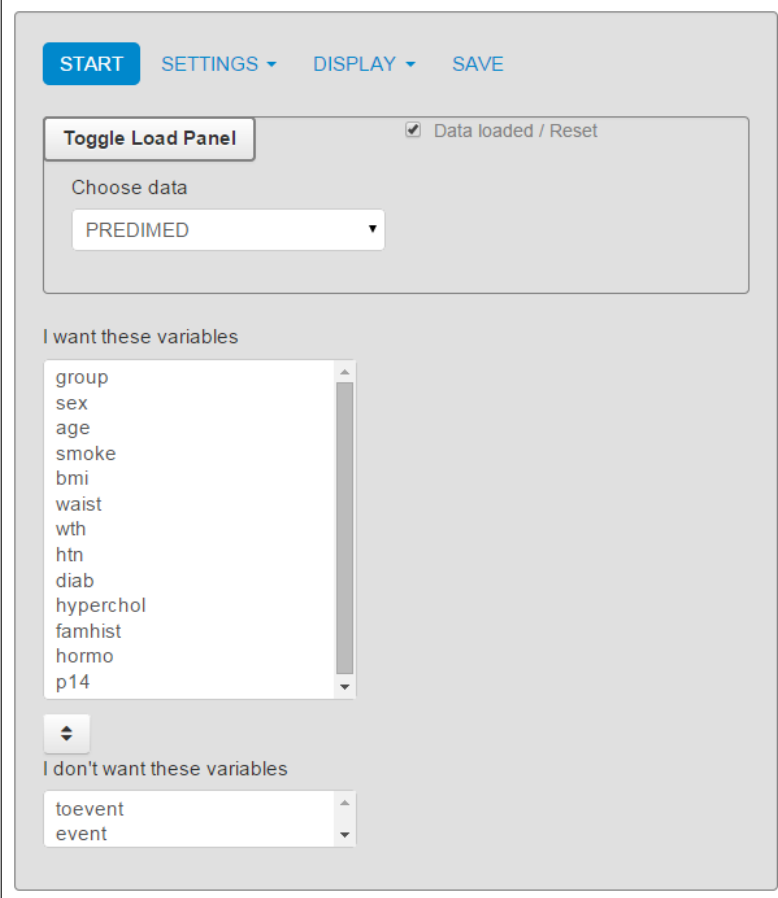
3 Example

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 and select the variables to be analysed** In this example we will load the PREDIMED data set from the PREDIMED study [?]. This data is already available in the `compareGroups` package. 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.



The screenshot shows the cGroupsWUI web application interface. At the top, there are four buttons: 'START' (highlighted in blue), 'SETTINGS', 'DISPLAY', and 'SAVE'. Below these buttons is a 'Toggle Load Panel' section. It contains a 'Choose data' dropdown menu with 'PREDIMED' selected. To the right of this dropdown is a checkbox labeled 'Data loaded / Reset' which is checked. Below the 'Toggle Load Panel' is a section titled 'I want these variables'. It contains a list of variables: group, sex, age, smoke, bmi, waist, wth, htn, diab, hyperchol, famhist, hormo, and p14. Below this list is a double-headed arrow button. Below the arrow button is a section titled 'I don't want these variables'. It contains a list of variables: toevent and event.

- **Step 2. 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).

START **SETTINGS** ▾ DISPLAY ▾ SAVE

Response type

☐ None

☐ Group

☒ Survival

Select time-to-event variable

toevent ▾

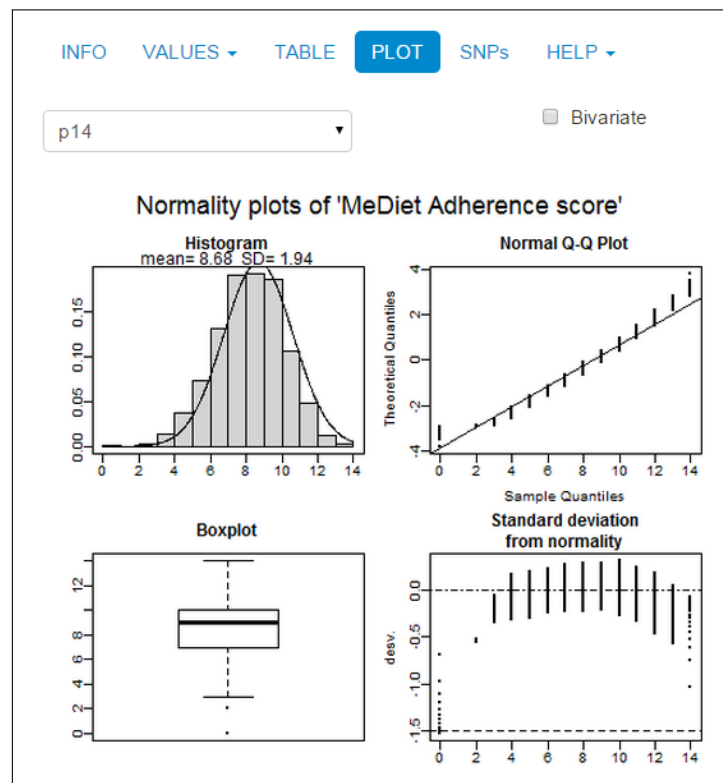
Select status variable

event ▾

Select event category

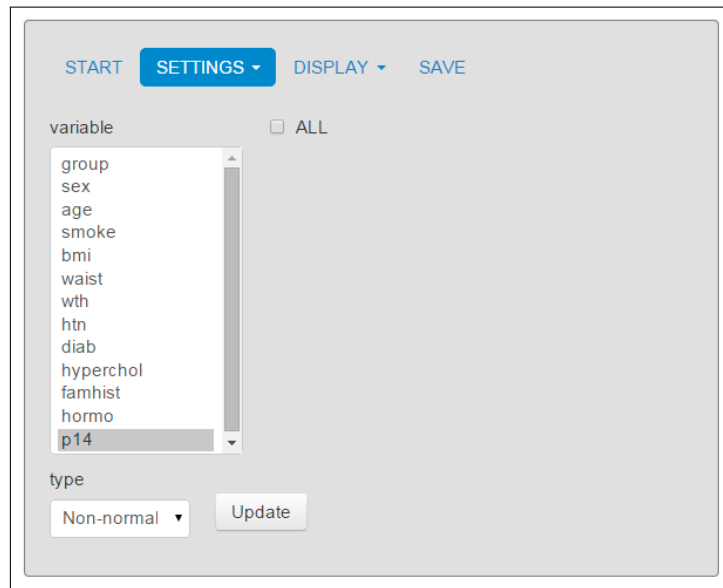
2:Yes ▾

- **Step 3. 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 4. 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:



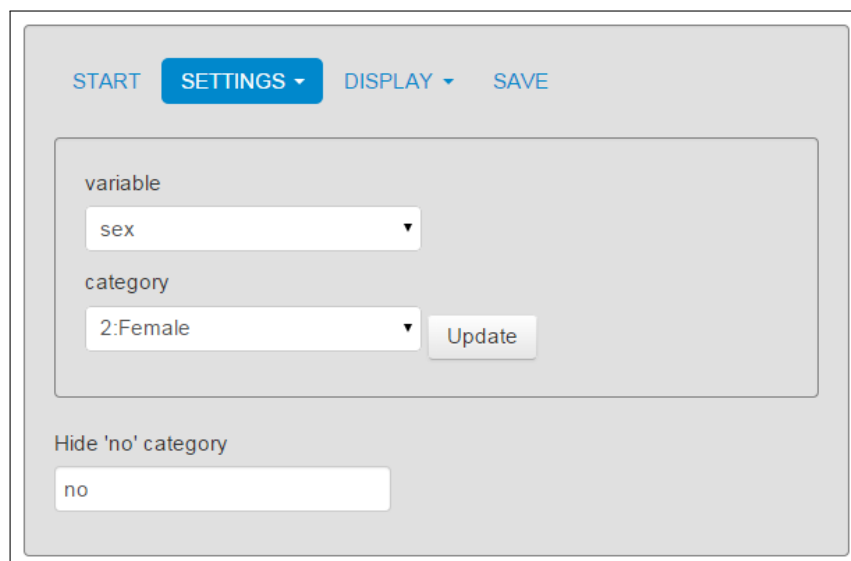
START SETTINGS DISPLAY SAVE

variable ☐ ALL

group
sex
age
smoke
bmi
waist
wth
htn
diab
hyperchol
famhist
hormo
p14

type
Non-normal Update

- **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.



START SETTINGS DISPLAY SAVE

variable
sex

category
2:Female Update

Hide 'no' category
no

- **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.

The screenshot shows the 'SETTINGS' window with the following configuration:

- Reference category:**
 - variable: sex
 - category: 2:Female
 - Update button
- Multiplying factor:**
 - variable: A list box containing 'age', 'bmi', 'waist' (selected), and 'wth'.
 - factor: 5
 - Update button
- ☐ ALL

- **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.

The screenshot shows the 'SETTINGS' window with the following configuration:

- Global subset**
 - Write subset expression: sex==1
 - Apply button
- Variable subset**
 - variable: A list box containing 'group', 'sex', 'age', and 'smoke'.
 - Write subset expression: (empty text box)
 - Apply button
- ☐ ALL

- **Step 5. 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:

START SETTINGS **DISPLAY** SAVE

☐ ALL
 ☐ p-overall
☒ Descriptives
 ☐ p-trend
☒ OR/HR
 ☒ OR/HR p-value
☐ Available
 ☐ NA category
☐ Pairwise p-value
☐ Simplify

- **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:

START SETTINGS **DISPLAY** SAVE

Frequencies:
☒ %
 ☐ n (%)
 ☐ n

Means:
☒ Mean (SD)
 ☐ Mean+-SD

Quantiles: Median [a, b]:

[a,]:

[, b]:

brackets
 ☐ Squared
 ☒ Rounded

separator
 ☒ Semicolon
 ☐ Comma
 ☐ Slash

Update

- **Decimals:** Moreover, 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:

The screenshot shows the 'SETTINGS' tab of a software interface. At the top, there are buttons for 'START', 'SETTINGS' (with a dropdown arrow), 'DISPLAY' (with a dropdown arrow), and 'SAVE'. The main content area is divided into three sections:

- p-values:** A label 'Number of decimals' is followed by a text input field containing the value '3'.
- Descriptives:** This section has a 'variable' dropdown menu showing 'group', 'sex', 'age', and 'smoke'. To the right of this menu is a checkbox labeled 'ALL' which is currently unchecked. Below the variable menu is a label 'Number of decimals (-1: default)' followed by a text input field containing '-1' and an 'Update' button.
- OR/HR:** This section also has a 'variable' dropdown menu showing 'group', 'sex', 'age', and 'smoke'. To the right is a checkbox labeled 'ALL' which is currently checked. Below the variable menu is a label 'Number of decimals (-1: default)' followed by a text input field containing '3' and an 'Update' button.

A vertical scrollbar is visible on the right side of the settings panel.

- **Headers:** Finally, some ‘key’ words in the header of the descriptive tables, (such as ‘p.ratio’, ...) may be changed:

The screenshot shows the 'DISPLAY' tab of the same software interface. At the top, there are buttons for 'START', 'SETTINGS' (with a dropdown arrow), 'DISPLAY' (with a dropdown arrow), and 'SAVE'. The main content area contains several text input fields and one button:

- All:** A label followed by a text input field containing '[ALL]' and an 'Apply' button.
- overall p-value:** A text input field containing 'p.overall'.
- p-value for trend:** A text input field containing 'p.trend'.
- OR/HR p-value:** A text input field containing 'p-value'.
- Available data:** A text input field containing 'N'.
- Caption (only for PDF):** A text input field containing 'NULL'.

- **Step 6. 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:**

INFO VALUES TABLE PLOT SNPs HELP

HTML PDF R console

Resize: 1

Var	No event N=2528	Event N=151	HR	p-value
Intervention group:				
Control	29.8%	38.4%	Ref.	Ref.
MedDiet + Nuts	36.7%	27.2%	0.538 [0.361;0.803]	0.002
MedDiet + VOO	33.5%	34.4%	0.708 [0.487;1.030]	0.071
Sex: Male	100%	100%	1.000 [1.000;1.000]	.
Age	66.0 (61.0;71.0)	68.0 (63.0;75.0)	1.056 [1.031;1.082]	<0.001
Smoking:				
Never	27.0%	16.6%	Ref.	Ref.
Current	24.8%	26.5%	1.629 [0.988;2.686]	0.056
Former	48.2%	57.0%	1.924 [1.233;3.004]	0.004
Body mass index	29.1 (26.9;31.5)	29.2 (26.9;31.6)	1.014 [0.966;1.065]	0.565
Waist circumference	103 (97.0;109)	104 (97.0;110)	1.037 [0.954;1.127]	0.391
Waist-to-height ratio	0.61 (0.58;0.65)	0.62 (0.58;0.66)	13.171 [0.848;204.538]	0.065
Hypertension	77.7%	79.5%	1.204 [0.811;1.787]	0.358
Type-2 diabetes	52.0%	64.2%	1.591 [1.140;2.219]	0.006
Dyslipidemia	66.6%	58.9%	0.805 [0.582;1.114]	0.191
Family history of premature CHD	17.4%	14.6%	0.894 [0.569;1.405]	0.627
Hormone-replacement therapy	0.00%	0.00%	1.000 [1.000;1.000]	.
MeDiet Adherence score	9.00 (8.00;10.0)	8.00 (7.00;10.0)	0.873 [0.808;0.942]	0.001

ii) **PDF:**

INFO VALUES TABLE PLOT SNPs HELP

HTML PDF R console

Resize: 1 5 10

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

	No event N=2528	Event N=151	HR	p-value
Intervention group:				
Control	29.8%	38.4%	Ref.	Ref.
MedDiet + Nuts	36.7%	27.2%	0.538 [0.361;0.803]	0.002
MedDiet + VOO	33.5%	34.4%	0.708 [0.487;1.030]	0.071
Sex: Male	100%	100%	1.000 [1.000;1.000]	.
Age	66.0 (61.0;71.0)	68.0 (63.0;75.0)	1.056 [1.031;1.082]	<0.001
Smoking:				
Never	27.0%	16.6%	Ref.	Ref.
Current	24.8%	26.5%	1.629 [0.988;2.686]	0.056
Former	48.2%	57.0%	1.924 [1.233;3.004]	0.004
Body mass index	29.1 (26.9;31.5)	29.2 (26.9;31.6)	1.014 [0.966;1.065]	0.565
Waist circumference	103 (97.0;109)	104 (97.0;110)	1.037 [0.954;1.127]	0.391
Waist-to-height ratio	0.61 (0.58;0.65)	0.62 (0.58;0.66)	13.171 [0.848;204.538]	0.065
Hypertension	77.7%	79.5%	1.204 [0.811;1.787]	0.358
Type-2 diabetes	52.0%	64.2%	1.591 [1.140;2.219]	0.006
Dyslipidemia	66.6%	58.9%	0.805 [0.582;1.114]	0.191
Family history of premature CHD	17.4%	14.6%	0.894 [0.569;1.405]	0.627
Hormone-replacement therapy	0.00%	0.00%	1.000 [1.000;1.000]	.
MeDiet Adherence score	9.00 (8.00;10.0)	8.00 (7.00;10.0)	0.873 [0.808;0.942]	0.001

iii) R:

INFO
VALUES
TABLE
PLOT
SNPs
HELP

HTML
PDF
R console

```

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

```

	No event N=2528	Event N=151	HR	p-value
Intervention group:				
Control	29.8%	38.4%	Ref.	Ref.
MedDiet + Nuts	36.7%	27.2%	0.538 [0.361;0.803]	0.002
MedDiet + VOO	33.5%	34.4%	0.708 [0.487;1.030]	0.071
Sex: Male	100%	100%	1.000 [1.000;1.000]	.
Age	66.0 (61.0;71.0)	68.0 (63.0;75.0)	1.056 [1.031;1.082]	<0.001
Smoking:				
Never	27.0%	16.6%	Ref.	Ref.
Current	24.8%	26.5%	1.629 [0.988;2.686]	0.056
Former	48.2%	57.0%	1.924 [1.233;3.004]	0.004
Body mass index	29.1 (26.9;31.5)	29.2 (26.9;31.6)	1.014 [0.966;1.065]	0.565
Waist circumference	103 (97.0;109)	104 (97.0;110)	1.037 [0.954;1.127]	0.391
Waist-to-height ratio	0.61 (0.58;0.65)	0.62 (0.58;0.66)	13.171 [0.848;204.538]	0.065
Hypertension	77.7%	79.5%	1.204 [0.811;1.787]	0.358
Type-2 diabetes	52.0%	64.2%	1.591 [1.140;2.219]	0.006
Dyslipidemia	66.6%	58.9%	0.805 [0.582;1.114]	0.191
Family history of premature CHD	17.4%	14.6%	0.894 [0.569;1.405]	0.627
Hormone-replacement therapy	0.00%	0.00%	1.000 [1.000;1.000]	.
MedDiet Adherence score	9.00 (8.00;10.0)	8.00 (7.00;10.0)	0.873 [0.808;0.942]	0.001

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

- **Step 7. 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.

START SETTINGS ▾ DISPLAY ▾ SAVE

Select format

PDF ▾ Download

Resize

normalsize ▾

☐ Landscape