# Receiver Operator Characteristic Tools Graphic User Interface Extension for R Commander

**Daniel-Corneliu LEUCUŢA[1,\*], Mihaela HEDEŞIU[2], Oana Cristina ALMĂŞAN[3], Andrei ACHIMAŞ[1]**

[1]Iuliu Haţieganu University of Medicine and Pharmacy Cluj-Napoca, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania
[2]Iuliu Haţieganu University of Medicine and Pharmacy Cluj-Napoca, Department of Oral and Maxillofacial Radiology, 32 Clinicilor, 400006, Cluj-Napoca, Romania
[3]Iuliu Haţieganu University of Medicine and Pharmacy Cluj-Napoca, Department of Prosthetic Dentistry, 32 Clinicilor, 400006, Cluj-Napoca, Romania
E-mail: dleucuta@umfcluj.ro

* Author to whom correspondence should be addressed; Tel.: +40-264-597256 int 2502; Fax: +4-0264-593847.

## Abstract

*Background:* Receiver Operator Characteristic (ROC) curve, is a graphical plot which presents the performance of a binary classifier when the discrimination cutoff is varied. The aim of this work was to create an extension for R Commander that offers a graphical user interface for Receiver Operator Characteristic tools provided by several existing command line accessible packages like pROC and ROCR. *Material and Methods:* The extension was built and tested with R version 3.2.0 and R Commander 2.1-7. *Results:* We built an extension called RcmdrPlugin.ROC that we uploaded on the CRAN servers. The extension adds a new menu called ROC, along with two submenus pROC and ROCR that broadly corresponds to commands available to access the functions of these packages. The pROC menu offers several commands: to plot a ROC curve for a dataset or for a logistic regression model, to compare paired and unpaired ROC curves, each providing the following tabs: General (to select the variables for the analysis, and options for switching cases with controls); Smoothing (allowing the user to select different types of smoothing – binominal, density, distributions like normal, lognormal, ...); AUC (to specify the partial area under the curve (AUC) options), CI (to select the options of confidence intervals (CI) – the level, computing method: DeLong, bootstrap, ...); Plot (for the plotting options). The ROCR dialogue window offers more options in choosing the performance measures for the plot. *Conclusion:* The RcmdrPlugin.ROC extension helps less advanced users of R accessing ROC tools in a friendly graphical user interface.

**Keywords:** Receiver Operator Characteristic (ROC); User-Computer Interface; R; Software

## Introduction

Receiver Operator Characteristic (ROC) curve, is a graphical plot which presents the performance of a binary classifier when the discrimination cutoff is varied. The ROC curves and associated analysis tools like area under ROC, comparisons of ROC curves, finding an optimal diagnostic threshold, are important analysis tools used in medical research, particularly in diagnostic field. R [1] is a free software environment for statistical computing and graphics that offers

packages providing these tools, accessible by command line only, and such being difficult to use by non-specialists. R Commander [2] graphical user interface (GUI) to R is one of the most comprehensive in respect of functions offered and extensibility. There are at least thirty-five (35) extensions to RCommander and their number is growing (e.g. design of experiments [3], meta-analysis [4], and evidence based medicine [5]). At the present moment there is no GUI to access the existing powerful ROC tools in R.

The aim of this work was to create an extension for R Commander that offers a GUI for Receiver Operator Characteristic tools provided by several existing command line accessible packages like pROC [6] and ROCR [7].

## Material and Methods

The extension (Plug-in) for the R Commander graphical user interphace was written in the S programming language in Rstudio [8] Integrated Development Environment, built and tested with R version 3.2.0 and R Commander 2.1-7 onto an Ubuntu Linux platform and checked on a Windows winbuilder server [9]. For building the interface, the extension uses abstracted commands in R of Tcl/Tk. These interfaces offer graphical access to the functions provided by the pROC and ROCR packages. The developed software was published online on the Comprehensive R Archive Network (CRAN) network of servers [10], as free software under the license terms: GNU General Public License 2.

## Results and Discussion

An extension called RcmdrPlugin.ROC that we uploaded on the CRAN servers (https://cran.r-project.org/web/packages/RcmdrPlugin.ROC/index.html) was successfully developed.

The extension adds a new menu called ROC, along with two submenus (see Figure 1) pROC and ROCR that broadly corresponds to commands available to access the functions of the packages pROC and ROCR.
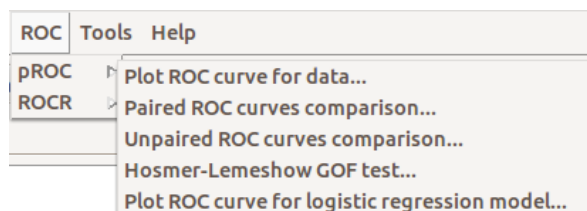


**Figure 1.** The ROC menu and pROC submenu added by the extension

The pROC menu offers several commands: to plot a ROC curve for a dataset or for a logistic regression model, to compare paired and unpaired ROC curves, to compute a Hosmer-Lemeshow goodness of fit test (see Figure 1). All these commands, except the Hosmer-Lemeshow test, open a window with the same tabs and options, with a few differences specific to each command purpose. The tabs are: General (to select the variables for the analysis, and options for switching cases with controls, or changing the output from 0-1 range to percentages); Smoothing (allowing the user to select different types of smoothing); AUC (to specify the area under the curve (AUC) options), CI (to select the options of confidence intervals (CI) computation); Plot (for the plotting options) (see Figure 2).
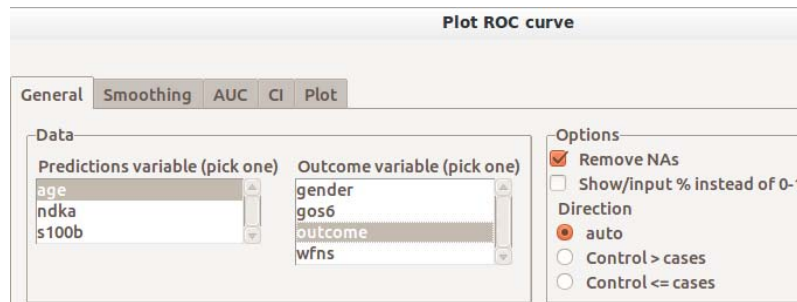
**Figure 2.** The General tab, with variable selection, and data options

The Smoothing tab offers different smoothing methods options: binomial, density, distribution (normal, lognormal, exponential, …), logcondens, and some finer adjustments.

The AUC tab allows the option to compute a partial AUC, by specifying the limits of it, and to choose the focus on sensitivity or specificity.

The CI tab provides the interface to specify the confidence level, the method for computing the confidence interval (DeLong, bootstrap – with the number of replicates), and the type of CI: for sensitivity, specificity, AUC, or for specific thresholds – all data points, local maximas, or custom defined ones.

The Plot tab offers many options for displaying different elements: the smoothing curves, AUC polygon, confidence interval, grid, identity line, text (for AUC, thresholds, p-values), labels, colors, line types, and confidence interval type (Figure 3). For paired and unpaired ROC curves the tabs have duplicate options for each of the curves that are compared (Figure 3).
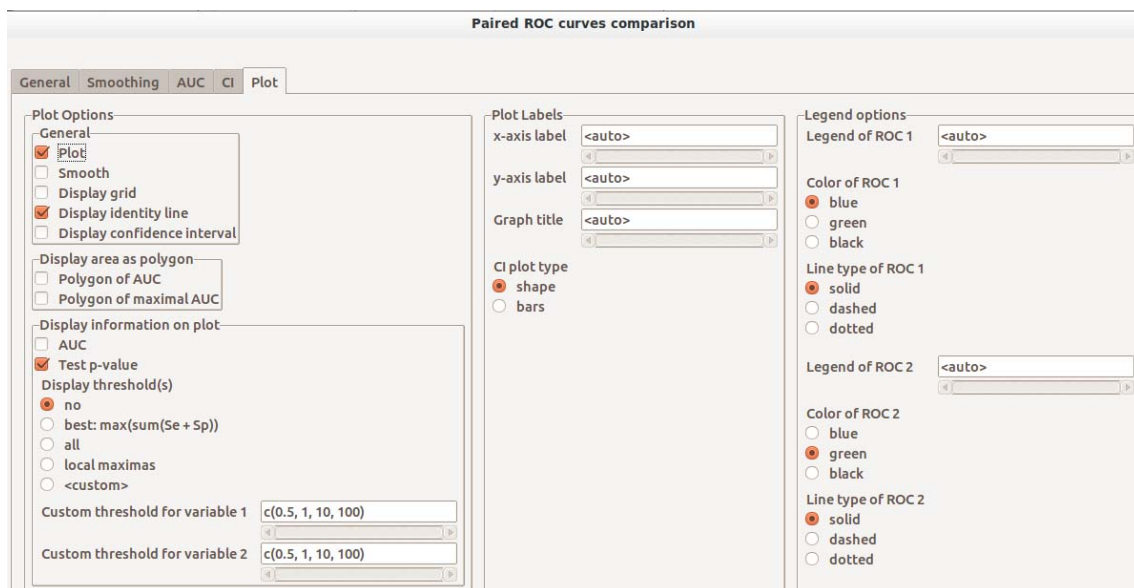


**Figure 3.** The Plot tab with numerous options for displaying two ROC curves

The dialogues by default output the confidence intervals for the AUC.

The ROCR dialogue window offers more options in choosing the axes for the plot, and some different customizations compared to pROC package (see Figure 4). For axes the following options are provided: accuracy, sensitivity, specificity, error rate, false positive rate, recall, positive predictive value, negative predictive value, precision, prediction-conditioned fallout, prediction-conditioned miss, phi correlation coefficient, mutual information, chi square statistic, odds ratio, lift value, precision/recall F measure, ROC convex hull, area under the ROC curve, precision/recall break-even point, calibration error, mean cross-entropy, root mean squared error, SAR measure, expected cost, explicit cost.
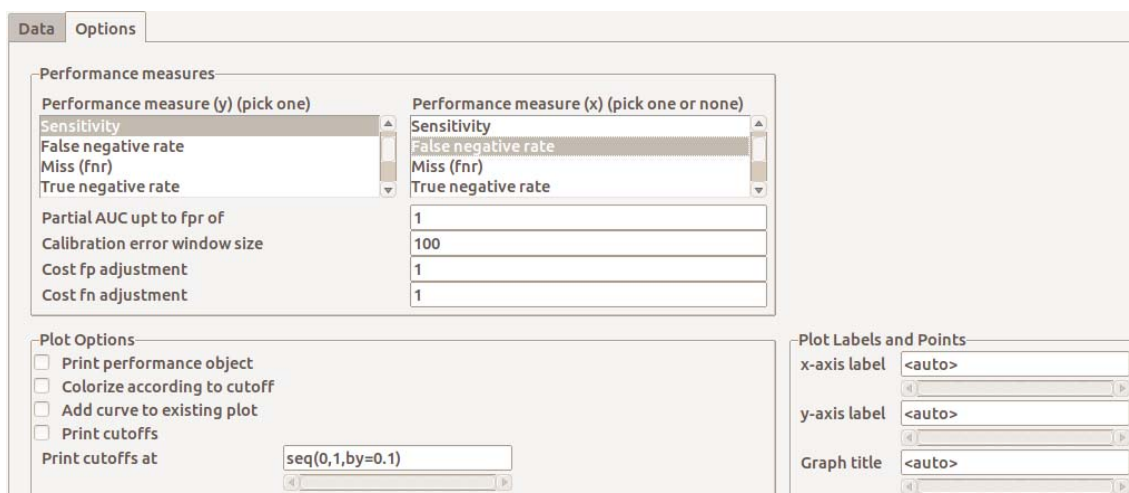
**Figure 4.** The ROCR, Options tab allowing more options to choose the performance measures for the plot

Both the pROC and ROCR options can be used for logistic regression models, being accessible from their corresponding submenus. For these models, the Hosmer-Lemeshow test is also available.

The extension has been built successfully, and it offers simplified graphical access to functions from pROC and ROCR packages.

The RCommander GUI offers functionality to import, display, analyze data, write scripts, output results, and extend this functionality by the help of extensions (there are many other features that RCommander offers). This way there is no need to recreate an interface for basic functions like importing, modifying data. New functionality can be added using this approach. This extension allows adding new features that can operate on loaded data sets, by means of the RCommander interface.

The confidence intervals for the AUC are displayed by default to help the improvement of reporting statistics in the published literature.

A comparison of features offered by different graphical user interfaces in our plugin, IBM SPSS 23 [11], Medcalc 15 [12], and S-Plus pROC plugin [6] is depicted Table 1. The comparison is not exhaustive, but it points the main capabilities of each interface.

Our extension compares very well with other graphical user interface software packages, having many features for ROC analyses. Except plotting false positive and negative rate, cost, calibration error, odds ratio, phi correlation coefficient, accuracy, it offers the same features as the pROC SPlus package, after which it was developed interface-wise. SPSS has fewer features, while MedCalc offers some features that can't be found in our extension (Outputting positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, cost values, and computing the sample size).

We aim to develop the functions offered by this extension with other goodness of fit tests for logistic regression models, and classification indexes, and output of medical diagnostics indicators (like predictive values, likelihood ratios), to further help the potential users.

**Table 1.** Features comparison between RCommander ROC plugin, IBM SPSS 23 base, Medcalc, and S-Plus pROC plugin, graphical user interfaces. AUC – area under the curve, CI – confidence interval, Se – sensitivity, Sp – specificity, PPV – positive predictive value, NPV – negative predictive value, LR – likelihood ratio

| | ROC plugin | SPSS 23 [11] | Medcalc 15 [12] | S-Plus pROC [6] |
|---|---|---|---|---|
| AUC with 95% CI DeLong method | yes | no | yes | yes |
| AUC with 95% CI Hanley & McNeil method | no | no | yes | no |
| AUC with 95% CI using W statistic | no | yes | no | no |
| AUC with 95% CI bootstrap | yes | no | yes | yes |
| CI for Se, Sp, or thresholds (local maxims or specified) | yes | no | yes | yes |
| Partial AUC for Se or Sp | yes | no | no | yes |
| Correct partial AUC with McClish formula | yes | no | no | yes |
| Display ROC curve chart | yes | yes | yes | yes |
| Display ROC curve 95% CI | yes | no | yes | yes |
| Mark specific cutoff points | yes | no | yes | yes |
| Plot false positive and negative rate, cost, calibration error, odds ratio, phi correlation coefficient, accuracy | yes | no | no | no |
| Paired ROC curve comparison | yes | yes | yes | yes |
| Unpaired ROC curve comparison | yes | no | no | yes |
| Smoothing curve (binormal, different density distributions) | yes | no | no | yes |
| Output values for Se, Sp, thresholds | yes | yes | yes | yes |
| Output PPV, NPV, +LR, -LR, cost | no | no | yes | no |
| Direct plot of ROC curve for logistic regression models | yes | no | no | no |
| Cost for false positive and false negative adjustments | yes | no | yes | yes |
| Compute sample size | no | no | yes | no |

**List of abbreviations**

AUC = area under the curve
CI = confidence intervals
CRAN = Comprehensive R Archive Network
GUI = graphical user interface
ROC = Receiver Operator Characteristic

**Conflict of Interest**

The authors declare that they have no conflict of interest.

**Acknowledgements**

**References**

1. R Development Core Team. R: A Language and Environment for Statistical Computing. R

Foundation for Statistical Computing. Vienna, Austria; 2015. Available from: http://www.R-project.org/

2. Fox J. The R Commander: A Basic-Statistics GUI for R. J Stat Softw [Internet]. Aug 2005 [cited 2015 July 4];14(9):1-42. Available from: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/

3. Ulrike G. R Commander Plugin for (industrial) Design of Experiments [Internet]. 2014 [cited 2015 July 4]. Available from: http://CRAN.R-project.org/package=RcmdrPlugin.DoE

4. Re AC. Graphical User Interface for Conduting Meta-Analysis in R Graphical User Interface for Conduting Meta-Analysis in R [Internet]. 2013 [cited 2015 July 4]. Available from: http://CRAN.R-project.org/package=RcmdrPlugin.MA

5. Leucuța D, Călinici T, Drugan T, Istrate D, Achimaș A. Graphical User Interface Extension in R Commander for Evidence Based Medicine Indicators. Applied Medical Informatics 2014;35(1):11-6. Available from: http://ami.info.umfcluj.ro/index.php/AMI/article/view/498

6. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77. Available from: http://www.biomedcentral.com/1471-2105/12/77

7. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics 2005;21(20):3940-41. Available from: http://bioinformatics.oxfordjournals.org/content/21/20/3940.abstract

8. RStudio.com [Internet]. Boston: RStudio, Inc.; ©2015 [cited 2015 July 4]. Available from: http://www.rstudio.com/

9. Ligges U. Building and checking R source packages for Windows [Internet]. [updated 2015 April 10; cited 2015 July 4] Available from: http://win-builder.r-project.org/

10. The Comprehensive R Archive Network [Internet]. [cited 2014 August 23]. Available from: http://cran.r-project.org/

11. IBM Corp. IBM SPSS Statistics for Windows, version 23.0. Armonk, NY; 2014. Available from: http://www-03.ibm.com/software/products/ro/spss-stats-base

12. MedCalc Software. MedCalc, version 15.11. Ostend, Belgium; 2015. Available from: https://www.medcalc.org/download.php