# StatAid: An R package with a graphical user interface for data analysis

Vincent Alcazer[1, 2]

1 Cancer Research Center of Lyon, INSERM U1052, Lyon, FRANCE **2** Hospices Civils de Lyon, Lyon, FRANCE

Summary

Data analysis is a crucial step in every research project in life science. Every clinician or researcher is one day faced with the need to perform statistical analysis. However, few free accessible solutions exist to date and most of the relevant software programs need a paid license. R is a free language which allows one to perform statistical analysis (R Core Team, 2017). While the R environment is very powerful, its learning curve can be very steep in the beginning, especially for people with no previous coding skills or those with less time to learn them. A graphical user interface has already been provided as an independent package, but its features are limited for medical and applied life science studies and its usage remains difficult and unintuitive for new users (Fox, 2005). Other free software programs exist, such as iNZight or Jamovi. However, while providing solutions with multiple features such as variable recoding, they do not guide the user through the analysis and can lack some key features such as time-dependent outcome analysis.

`StatAid` is a free open-source software provided as an R package which allows clinicians and researchers to perform statistical analysis through an intuitive graphical interface. It has been developed using the Shiny package (Chang, Cheng, Allaire, Xie, & McPherson, 2020), while Golem was used for package compilation and deployment (Guyader, Fay, Rochette, & Girard, 2020).

This software guides the user through all the steps of data analysis and includes multiple features such as:

- Exploratory data analysis: distribution, count, missing-values and outliers check
- Descriptive analysis, simple comparative analysis and publication ready 'table 1' output
- Publication-ready graph customization
- Paired data analysis (e.g. for repeated measures and matched case-control studies)
- Univariate analysis and models for continuous and categorical outcome: correlation, linear and logistic regression
- Univariate analysis and models for time-dependent outcome: Kaplan-Meier curves and Cox regression
- Multivariate analysis and models for continuous, categorical and time-dependent outcomes

Its user-friendly interface can guide clinicians or researchers, even those without previous software experience, through statistical analysis. In addition to a local version, a ready-to-use online version (https://vincentalcazer.shinyapps.io/StatAid/) is also available, providing access to `StatAid` everywhere, even on hospital/research center computers where external software installation is not allowed.

Statement of need

`StatAid` is an R package designed to fit the needs of every-day statistical analysis in science. This package provides all the tools necessary to perform data analysis in an intuitive, ready-to-use graphical interface. Users without coding skills or the availability of softwares with paid-licenses can easily perform all the steps of a good statistical analysis, from data-exploration/quality check to multivariate modeling.

Compared with other similar free software, `StatAid` has been designed to quickly produce publication-ready graphs and tables by guiding the user through their data analysis and providing multiple graph customization options. By limiting the number of choices and integrating different checks and variable controls, `StatAid` helps the user prevent bad test use or bad graph choice. Besides, as an evolving software, `StatAid` also provides the possibility for users to request the implementation of additional features or to contribute to software development. Its open-source aspect can also be seen as a security for people working with sensitive data (e.g. data from clinical trials/patients). The online version of `StatAid` enables it to be accessable everywhere, even on computers with restrictive policies for software installation.

`StatAid` was designed to be used by clinicians, researchers, students, and any person wanting to perform statistical analysis with no prior coding skills. Primarily designed for medical/life science data analysis, `StatAid` can also easily be extended to other fields.

References

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2020). Shiny: Web application framework for R. Retrieved from https://CRAN.R-project.org/package=shiny

Fox, J. (2005). The R commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software*, *14*(9). doi:10.18637/jss.v014.i09

Guyader, V., Fay, C., Rochette, S., & Girard, C. (2020). Golem: A framework for robust shiny applications. Retrieved from https://CRAN.R-project.org/package=golem

R Core Team. (2017). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/