TableOne: an online web application and R package for summarising and visualising data

Agapios Panos,¹ Dimitris Mavridis ^{1,2}

¹Department of Primary Education, University of Ioannina, Ioannina, Greece
²Université Paris Descartes, Faculté de Médecine, Paris, France

Correspondence to
 Professor Dimitris Mavridis,
 Department of Primary Education, University of Ioannina, Ioannina, Greece;
 dmavridi@uoi.gr

Received 18 May 2020
 Revised 4 June 2020
 Accepted 9 June 2020
 Published Online First
 14 July 2020

ABSTRACT

Objective To develop an easy-to-use R package and web application that summarises baseline characteristics across different arms of a clinical trial or different exposures.

Methods Tables and figures are the efficient means of visualising, communicating and summarising data. It is common in comparative effectiveness research to provide a synopsis of characteristics and outcomes across the various treatment groups. The popularity of such a table has earned it a name and we simply call it the 'TableOne', as it is usually the first TableOne encounters looking at a published clinical trial. Such a table includes not only descriptive statistics for each group but also appropriate tests (p values and 95% CIs) for checking for differences across groups. We have developed an R package (called TableOne) (accessible through <https://github.com/agapiospanos/TableOne>) that quickly summarises and compares results across different groups. We have also extended it to an online web application that is easily handled by the researcher. All computations are done in R and plots are produced using the plotly library. We provide a detailed description on how to use the web application.

Results The application guides the user in a step by step format (wizard) and it is accessible through any browser in the following link (<https://esm.uoi.gr/shiny/tableone/>). Finally, appropriate interactive plots are provided for each variable.

Conclusions This easy-to-use web application will help researchers quickly and easily to visualise differences across treatment groups or different exposures.

INTRODUCTION

Tables and figures are the efficient means of visualising, communicating and summarising data. Most research designs summarise important attributes of the participants enrolled at the start of the study (baseline characteristics) via a table. For example, in meta-analyses, we may have study characteristics across treatment and control groups. Randomised clinical trials (RCTs) typically report a table with baseline characteristics and their comparison across groups. When confronted with observational evidence, we explore the various groups to check if treatment selection depends on individuals' characteristics, which, in turn, may confound results. Such a table has established itself as the first table presented in a scientific paper and this is why it is also known as TableOne. Typically, TableOne includes demographic characteristics (eg, age, sex, race), medical history (eg, stroke, diabetes, family history), examination findings (eg, systolic/diastolic blood pressure (SBP/DBP),

glycated haemoglobin (HbA1c), test measurements), lifestyle characteristics (eg, exercise, dietary intake).

METHODS

We would also like to explore for differences in these characteristics across groups (eg, control and treatment group, exposure or non-exposure to a risk factor) by applying appropriate tests and presenting 95% CIs. Additionally, graphical displays should complement data presentation. The variable type plays an important role when presenting data as each type of variable has appropriate descriptive statistics, figures and tests. Variables are of different types, there are dichotomous variables (eg, sex, previous stroke), continuous variables (eg, SBP/DBP, HbA1c, weight), ordinal variables (eg, Likert items, modified Rankin Scale (MRS)). For continuous data, one may choose between mean and SD or, if distributions are far from normal, median and IQR, whereas for variable on the nominal and ordinal scale we use absolute and relative frequencies. We developed an R package (called TableOne) that quickly summarises and compares results across different groups. The R package is now available through Github in <https://github.com/agapiospanos/TableOne> and we are in the process of uploading it in the Comprehensive R Archive Network (CRAN). We have also extended it to an online web application that is easily handled by the researcher. We will place more emphasis on the web application in this manuscript. The application guides the user in a step by step format (wizard) and it is accessible through any browser in the following link (<https://esm.uoi.gr/shiny/tableone/>). Finally, appropriate interactive plots are provided for each variable. We offer a variety of graphs such as barplots, stacked barplots, histograms, pie charts, boxplots and scatterplots. All computations are done in R¹ and plots are produced using the plotly library.²

We used data from the Evros Stroke Registry (ESR), a prospective population-based study evaluating the incidence of first-ever stroke in the Evros prefecture in north-eastern Greece. The ESR contains data from consecutive patients with first-ever ischaemic stroke recorded during a 24-month period. We focus mainly on two different groups, the Embolic Stroke of Undetermined Source (ESUS) and non-ESUS strokes. Data are presented in detail in Tsivgoulis *et al*³ and the dataset is available in <https://github.com/agapiospanos/TableOne>

RESULTS

We used the TableOne web application in <https://esm.uoi.gr/shiny/tableone/> to present baseline characteristics of the ESR across ESUS and non-ESUS

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Panos A, Mavridis D. *Evid Based Ment Health* 2020;**23**:127–130.

TableONE

Instructions

In the file that you will use, please make sure that all the columns have names and there are not empty column names in the 1st row.

If you experience any difficulties do not hesitate to contact us via the contact form in our website!

<https://esm.uoi.gr/en/contact-page/>

Choose an excel file

Please choose the excel file that contains the data (.csv, .xlsx, .xls)

Or use the example data

Example data reference:

Tsivgoulis G, Kargiotis O, Katsanos AH, et al. Incidence, characteristics and outcomes in patients with embolic stroke of undetermined source: a population-based study. *J Neurol Sci* 2019;401:5-11.

version 0.2.6
Designed & Developed by Agapios Panos and Dimitris Mavridis

TableONE

Select which excel columns you want to include in TableOne export and click next

ESUS, Death, MRS1, Age, SBP, [

ESUS	☒
Days_of_RS	☒
Recurrent_stroke	☒
Death	☒
Days_of_Death	☒
MRS1	☒
MRS2	☒
MRS3	☒
MRS4	☒
Age	☒

Designed & Developed by Agapios Panos and Dimitris Mavridis

TableONE

Specify the grouping information

Select the excel column that contains the grouping information for the treatment and control group

ESUS
Death
MRS1
Age
SBP
DBP
NIHSS1
female_sex
smoking

Select treatment group value

Select control group value

version 0.2.6
Designed & Developed by Agapios Panos and Dimitris Mavridis

TableONE

Specify the dichotomous variables

Select the excel columns that contains dichotomous values

Death, female_sex, smoking

version 0.2.6
Designed & Developed by Agapios Panos and Dimitris Mavridis

Figure 1 On the top left corner, it is the initial page of the application where one chooses the dataset and the excel sheet (ESUS.Evros.xls). On the top right corner, we choose the variables we want to work with. On the bottom left corner, we choose the grouping variable (here ESUS) and on the bottom right corner we specify the dichotomous variables. DBP, diastolic blood pressure; ESUS, Embolic Stroke of Undetermined Source; MRS, modified Rankin Scale; NIHSS, National Institute of Health Stroke Scale; SBP, systolic blood pressure.

Figure 2 TableOne main output, the table that summarises the evidence on the selected variables across the grouping variable.

cryptogenic stroke patients. Initially, the application asks us to upload an excel file with the data. You can see the initial page on the top left corner of [figure 1](#). The excel file can have the labels of each column in the first row and the application will recognise those as the names of the variables. Then, we chose the variables we want to work with (top right corner of [figure 1](#)). Here, we chose ESUS (whether a cryptogenic stroke is classified as an ESUS or non-ESUS), death, MRS1, age, DBP, SBP, National Institute of Health Stroke Scale, female_sex (whether a patient is a woman or not) and smoking. We have to define the grouping variable ESUS (bottom left corner of [figure 1](#)) and then the application asks to define which variables are dichotomous (in our

case death, female_sex and smoking —bottom right corner of [figure 1](#)), which are ordinal (MRS1) and the rest are taken to be continuous. It also asks us for which continuous variables we would rather have median and IQR reported instead of mean and SD. Then the application asks us to put labels, if desired, to the included variables and finally proceeds to produce the table ([figure 2](#)). For dichotomous and ordinal variables, we have percentages and ORs and χ^2 tests, respectively. For continuous variables, we have mean, SD and t-test or median, IQR and Mann-Whitney tests if distributions are not normal. In [figure 3](#), we present some of the graphical options of the application. The left-side plot gives histograms of the National Institutes of

Figure 3 Graphs produced by the TableOne application. The left-hand side plot depicts histograms of NIH stroke scale and the right-hand side plot depicts stacked barplots of the modified Ranking Scale across ESUS and non-ESUS cryptogenic strokes. ESUS, Embolic Stroke of Undetermined Source; NIHSS, National Institute of Health Stroke Scale.

Health (NIH) Stroke scale for the ESUS and non-ESUS groups whereas the right-side plot gives a stacked barplot for the MRS, which is the typical way of presenting this type of data.

DISCUSSION

We developed an easy-to-use web application and R library for summarising and communicating evidence and exploring differences across competing groups. It can be used for tabular and graphical representations. This is an ongoing project and it aims to help researchers visualise their data base. Just like with other shiny tools or easy-to-use methods, there is the risk that some researchers will use the application to produce descriptives, tests and plots that are not suitable for the data at hand. For example, one may a wrong test or produce an incomprehensible graph. There is also the risk that one without a proper statistical education interprets results from tests too naively. For example, we have noticed that some clinicians tend to over-emphasise p value results, focusing on statistical significance and ignoring the actual effect value. It is likely that we may have an important but not statistically significant effect or vice versa. Also, with multiple tests, we expect to find false positive results by chance. Even in RCTs, we may found imbalance of baseline characteristics across groups and this does not invalidate the randomisation process. Hence, we warn readers to avoid checking for baseline imbalances in RCTs or at least be very cautious when doing so. Generally, we suggest advising a

statistician for the proper analysis and interpretation of results. We aim to continuously improve the TableOne web application. Immediate improvements include expanding to cases where we have more than two groups and include more options about what to present in the table.

Contributors AP programmed the web application and R code under the guidance of DM. DM wrote the manuscript and AP made comments.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests No, there are no competing interests.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository (<https://esm.uoi.gr/shiny/tableone/>)

ORCID iD

Dimitris Mavridis <https://orcid.org/0000-0003-1041-4592>

REFERENCES

- 1 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018URL. <https://www.R-project.org/>
- 2 Plotly Technologies Inc. *Collaborative data science publisher: Plotly technologies Inc. place of publication: Montréal, Qc date of publication: 2015 URL*, 2015. <https://plot.ly>
- 3 Tsigvoulis G, Kargiotis O, Katsanos AH, et al. Incidence, characteristics and outcomes in patients with embolic stroke of undetermined source: a population-based study. *J Neurol Sci* 2019;401:5–11.