



## A Computer-Aided SAS Macro for the Evaluation of the Simulation Performances in Missingness Settings

Urko Aguirre<sup>1,\*</sup>, Inmaculada Arostegui<sup>2,4</sup>, Jose M. Quintana<sup>3</sup>

<sup>1</sup> Research Unit, REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas, Hospital Galdakao-Usansolo, Galdakao, Spain; [urko.aguirrelarracoechea@osakidetza.eus](mailto:urko.aguirrelarracoechea@osakidetza.eus).

<sup>2</sup> Department of Applied Mathematics, Statistics and Operational Research; REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas. Faculty of Science and Technology, Leioa. Spain; [inmaculada.arostegui@ehu.es](mailto:inmaculada.arostegui@ehu.es)

<sup>3</sup> Research Unit, REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas, Hospital Galdakao-Usansolo, Galdakao, Spain; [josemaria.quintanalopez@osakidetza.eus](mailto:josemaria.quintanalopez@osakidetza.eus).

<sup>4</sup> BCAM-Basque Center for Applied Mathematics, Bilbao, Spain.

\* Author to whom correspondence should be addressed;

E-Mail: [urko.aguirrelarracoechea@osakidetza.eus](mailto:urko.aguirrelarracoechea@osakidetza.eus); Tel.: +34 94 400 71 05; Fax: +34 94 400 71 32.

*Published: 7 December 2015*

---

**Abstract:** Model validation has become a topic of great interest to many fields such as industry, medicine or even to government. Its main challenge is to provide stable and credible tools so that the decision-maker with the information necessary can make high-consequence judgments. This process requires simulation modelling and consequently, some guidelines or evaluation criteria are essential in order to draw meaningful conclusions. A computer-aided SAS® macro is developed using the SAS/IML programming language. Researchers should provide the dataset to be analyzed and the true values to be compared. As a result, the statistical program shows measures (i.e., number of simulations to be performed, bias, accuracy, coverage, etc...) which help investigators to make decisions with a minimal effort of programming. Numerical results of the aforementioned statistical parameters, plots and a report are returned by the statistical tool. Although this macro is focused on the missingness setting, it is applicable to any other discipline. We encourage researchers to use it to make better statistical assessments of the used methods.

---

**Keywords:** simulation; missing data; SAS®; macro

**Mol2Net YouTube channel:** <http://bit.do/mol2net-tube>

### 1. Introduction

Statistical modelling is a powerful tool which is used to describe mathematically real-life world issues.

One of the most important roles of the statistical modelling is to create stable and robust mathematical expressions: models should be

internally consistent and provide the best accurate estimates. Model validation is one of the approaches to address this problem.

Model validation is a straightforward and powerful process. It is concerned with building the right model. As noted, it is utilized to determine that a model is an accurate representation of the real system. Validation is usually achieved through the calibration of the model, an iterative process of comparing the model to actual system behavior and using the discrepancies between the two, and the insights gained, to improve the model. This process is repeated until model accuracy is judged to be acceptable. To this end, it requires simulation modeling.

Simulation is a computational technique that relies on repeating a computation on many different random samples in order to estimate a statistical quantity. These techniques provide empirical estimation of the sampling distribution of the parameters of interest that could not be achieved from a single study and enable the estimation of accuracy measures, such as the bias in the estimates of interest, as the truth is known [1]. Based on several parameters applied to the simulation performances, researchers would be able to determine whether the developed statistical model is acceptable.

### 3. Materials and Methods

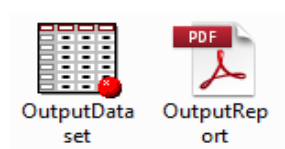
The main routine is composed of explicit-passed several parameters needed for the assessment of the performed simulations. First of all, as input parameter, researchers should provide a dataset including the obtained simulation results. Moreover, the true values of the parameters to be assessed are also required. Finally, the type of missing data mechanisms (missing completely at random, missing at random and missing not at random) and missingness rate should be provided.

This statistical procedure evaluates the simulation performances according to the following mathematical expressions [2]:

The main objective of this manuscript is to provide a statistical tool that assess the performed simulation models.

## 2. Results and Discussion

Once executed the SAS® macro, results are given for the overall dataset or stratified by a specified group. These numerical outputs are displayed in a Word /PDF file or if desired, stored in a dataset (see Figure 1).



**Figure 1.** Output Delivery System of the SAS® macro.

The advances in computer technology have allowed simulation studies to be more accessible. This statistical tool allows to compute the most used parameters to make a proper simulation assessment. However, performing simulations is not simple; to perform a proper simulation assessment several parameters and designs should be taken into account.

- Raw Bias: It is defined as the difference between the true value and the mean average of the simulated coefficients.
- Relative bias: Relative bias was calculated by dividing the raw bias (difference between the mean value over simulation results and the true parameter) by the true value.
- Standardized bias: We compared the true value of the beta regression coefficient of the considered interaction factor in the model with the corresponding value obtained with each of the analyzed methods, relative to the standard error of the simulated value.
- Coverage: The coverage of a confidence interval is the proportion of times that the obtained confidence interval (using the simulated values) contains the true specified

parameter value. If the coverage value is below the 90%, the performance of the interval procedure will be troublesome.

- Relative width: If one procedure has a similar or higher rate of coverage than another but yields intervals that are substantially shorter, then it should be preferred. Shorter intervals translate into greater accuracy and higher power.

The computer-aided SAS® macro is internally programmed using the SAS/IML language under

SAS 9.4 release. Different SAS statements such as PROC MEANS or DATA steps are used to set up the aforementioned statistical parameters.

#### 4. Conclusions

The SAS® macro is highly accessible and we encourage researchers to use it to make better statistical assessments of the used methods.

#### Author Contributions

All authors have equally contributed.

#### Conflicts of Interest

The authors declare no conflict of interest.

#### References and Notes

1. Burton A.; Altman, D.G.; Royston, P.; Holder, R.L. The design of simulation studies in medical statistics. *Statistics in Medicine* **2006**, *25*, 4279-4292.
2. Aguirre U.; Arostegui, I; Esteban, C; Quintana, J.M. Assessment of the performance of imputation techniques in observational studies with two measurements. *International Journal of Statistics in Medical Research* **2015**, *4*(3), 240-251.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions defined by MDPI AG, the publisher of the Sciforum.net platform. Sciforum papers authors the copyright to their scholarly works. Hence, by submitting a paper to this conference, you retain the copyright, but you grant MDPI AG the non-exclusive and unrevocable license right to publish this paper online on the Sciforum.net platform. This means you can easily submit your paper to any scientific journal at a later stage and transfer the copyright to its publisher (if required by that publisher). (<http://sciforum.net/about> ).