Replication Report Brookhart et al. (2006)

Juan Claramunt[1], Felix Clouth[2]

[1] Leiden University

[2] Tilburg University

April 18, 2023

Replication Report, Brookhart et al. (2006)

Abstract

In this document, we describe a replication study of "Variable Selection for Propensity Score Models" by Brookhart et al. (Brookhart et al. 2006). This report includes an analysis of the data-generating mechanism, simulation study set-up and methods described in the original paper, the feasibility of its implementation based on the information available in the original document and the results we were able to replicate. Furthermore, we analyze whether the presentation of the methods and results on the original paper can be improved. We were able to replicate the results of the first experiment of the original paper by making a few small assumptions. However, we were not able to replicate the results of experiment two. This was probably caused by a lack of explanation of certain concepts used on the article, the lack of experience of the authors of this report and the extra difficulty that requires replicating a figure instead of a table.

Correspondence concerning this replication report should be addressed to:
f.j.clouth@tilburguniversity.edu

1 Introduction

This replication report documents the replication attempt of the simulation study Brookhart et al. (2006). Following the definition of Rougier et al. (2017) we understand the replication of a published study as writing and running new code based on the description provided in the original publication with the aim of obtaining the same results. This replication study was independently done by another pair of researchers. In the discussion section we will include an overview of the common findings an differences.

2 Method

2.1 Information basis

This replication was conducted mainly on information provided in Brookhart et al. (2006). Software code or additional information (such as supplementary materials) were not available. The addition of the open source code (according to the original paper, it exists but it is not published) will make the replication of the study much easier. Additionally, we used information provided in the splines2 and ROCR R-package documentation. However, this information was not referred to in the main article.

2.2 Data Generating Mechanism

Information provided in the above mentioned sources indicated that the following simulation factors were systematically varied in generating the artificial data.

2.2.1 Experiment 11

Simulation factor	No. levels	Levels
Varied		
Sample size	2	500, 2500
Fixed		
α_0		0.5
α_1		4
α_2		1
α_3		0
α_4		0.5
β_0		0

Simulation factor	No. levels	Levels
β_1		0.5
β_2		0
β_3		0.75
Dichotomous exposure A		sampled from rbern(N, prob=pnorm($q = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3$))
Outcome Y		sampled from rpois(N, lamda=$exp(\alpha_0+\alpha_1*(((1+exp(-3*X_1))^{-1})-0.5)+\alpha_2*X_2+\alpha_3*X_3+\alpha_4*A))$
Three covariates		independently sampled from rnorm(N, 0, 1)

2.2.2 Experiment 1 (sensitivity analysis)

Simulation factor	No. levels	Levels
Varied		
Standard deviation of covariates	2	0.5, 1.5
α_4	3	0.25, 0.5, 1
β_0	2	-1, 0
Fixed		
Sample size		500
α_0		0.5
α_1		4
α_2		1
α_3		0
β_1		0.5
β_2		0
β_3		0.75
Dichotomous exposure A		sampled from rbern(N, prob=pnorm($q = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3$))
Outcome Y		sampled from rpois(N, lamda=$exp(\alpha_0+*\alpha_1*(((1+exp(-3*X_1))^{-1})-0.5)+\alpha_2*X_2+\alpha_3*X_3+\alpha_4*A))$

2.2.3 Experiment 2

Simulation factor	No. levels	Levels
Varied		
Sample size	2	500, 2500
α_1	21	[0, .01, ..., .2]
β_1	26	[0, .05, ..., 1.25]
Fixed		
α_0		0.5
α_2		1
α_3		0
α_4		0.5
β_0		0
β_2		0
β_3		0.75
Three covariates		independently sampled from rnorm(N, 0, 1)
Outcome Y		sampled from rpois(exp($\alpha_0 + \alpha_1 * X_1 + \alpha_2 * X_2 + \alpha_3 * X_3 + \alpha_4 * A$))
Dichotomous exposure A		sampled from rbern(N, prob=pnorm($q = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3$))

The design was full-factorial in the main analyses of Experiment 1 and Experiment 2. For the sensitivity analyses of Experiment 1, all parameters were held at their default values while a single parameter was altered.

2.3 Research Comparison

The study compares different specifications of a propensity scores (PS) model using a Monte-Carlo simulation. The model examines the effects of the inclusion of three types of covariates: X1 (a true confounder), X2 (a predictor of the outcome) or X3 (a variable related to the exposure but not the outcome). The exposure effects are estimated using two methods: (1) by adjusting for the PS in a multivariate outcome model in which the effect of the estimated PS was modeled through a regression spline. In this case, the propensity score model fit is given by the following formula:

$$E[Y|\hat{PS}, A] = exp\{\lambda + \sum_k \psi_k B_k(\hat{PS}) + \gamma A\}$$

where λ is the baseline rate, B_k are B-spline basis functions and γ is the treatment effect. (2) Using subclassification in which strata are defined by quintiles of the estimated propensity scores and taking the average treatment effect across strata.

2.4 Performance measures

To measure the performance of the different specifications of the PS models of the original simulation study, the authors measured the bias, the variance and the mean square error of the estimates of the parameter α_4 and the average C-stat.

2.5 Technical implementation

While the original simulation study was carried out in R version 1.9.1 running on a Windows XP platform, our replication was implemented using the R programming environment version 4.0.3 on a Windows 10 environment (details regarding software versions can be obtained from the section Reproducibility Information). The corresponding R code can be obtained from https://github.com/replisims/your_repo.
The following table provides an overview of replicator degrees of freedom, i.e. decisions that had to be made by the replicators because of insufficient or contradicting information. Issues were resolved by discussion among the replicators. Decisions were based on what the replicators perceived to be the most likely implementation with likeliness estimated by common practice and/or guideline recommendations. Wherever feasible multiple interpretations where implemented.

Issue	Replicator decision	Justification
It was not clear whether to estimate the PS or to calculate them from the population model	Estimate the PS	It should result in a better fit for each simulated dataset.
The implementation of the cubic spline estimation was not clearly explained. Moreover, the reference included does not provide enough insight	We used the function bs() in splines package, with specifying three knots splines::bs(PS, knots = quantile(PS, probs = c(0.25, 0.5, 0.75)))	Only available function in R. We used the default options.
Subclassification approach not fully defined. Not clear how to obatin the estimates of the exposure rate within strata.	We used the mean as estimator	It is the most reasonable choice.

Issue	Replicator decision	Justification
No description of c-statistic	Hmisc::somers2(PS, data[,"exposure"])["C"]	Most likely meaning of the c-statistic. Most common R package to compute the area under the ROC curve.
Some values required for experiment 2 were missing	Used the same values we used in experiment 1	If the values are omitted, the most likely reason is that the values are the same for both experiments.
Definition and scales of the variables measured in the axes of the figures of experiment 2 were not clearly described.	No solution	
Estimators of experiment 2 were not clearly explained.	Use the estimators of experiment 1 that fitted better the brief explanation	It is the most reasonable choice.
According to the Methods section, the PS was adjusted using the spline function in experiment 2. However, in the results section we can find the quintile approach as well.	No solution	

2.6 C-statistic issue

In a first version of the original paper we were trying to replicate, it was not specified what the *c* statistic mean. Therefore, it was not clear what measure we were replicating nor how we were supposed to do so. In latter versions, it is stated that the *c*-statistics they use in this research is the area under the receiver operating characteristic (ROC) curve.

2.7 Experiment 2 issue

Whilw trying to replicate the results of the second experiment we found multiple issues. First, they use multiple transformation of the results that are not defined in the paper. Second, it is far more difficult to compare the results obtained in a graph than in a table. While we could easily check that the results we obtained for experiment 1 were exactly the same as the results obtained in the original paper, it was incredibly difficult to do the same with experiment 2. Furthermore, there is no open code available to

Table 5: Table 1. Simulation experiment 1, with results based on an analysis in which the propensity score is entered into an outcome model as a parametric spline term.

	Variable(s) in propensity score model							
	X_1	X_2	X_3	X_1, X_2	X_1, X_3	X_2, X_3	X_1, X_2, X_3	None
n = 500								
Bias	0.008	0.601	0.744	0.006	0.007	0.744	0.006	0.599
Variance	0.030	0.025	0.047	0.020	0.040	0.038	0.030	0.040
MSE	0.030	0.386	0.600	0.020	0.040	0.591	0.030	0.398
avg c-stat	0.67	0.52	0.76	0.67	0.82	0.76	0.82	
n = 2500								
Bias	0.008	0.601	0.744	0.006	0.007	0.744	0.006	0.592
Variance	0.030	0.025	0.047	0.020	0.040	0.038	0.030	0.009
MSE	0.030	0.386	0.600	0.020	0.040	0.591	0.030	0.359
avg c-stat	0.67	0.51	0.76	0.67	0.81	0.76	0.81	

replicate these figures. These difficulties did not allow us to replicate the figures of experiment 2.

3 Results

3.1 Experiment 1

In this section we show the tables we obtain during the replication of experiment 1.

3.2 Experiment 2

We were not able to replicate the figures of experiment 2.

Table 6: Table 2. Simulation experiment 1, with results based on an analysis using subclassification in which strata are defined by quintiles of the estimated propensity score.

	Variable(s) in propensity score model							
	X_1	X_2	X_3	X_1, X_2	X_1, X_3	X_2, X_3	X_1, X_2, X_3	None
n = 500								
Bias	0.027	0.607	0.802	0.030	0.037	0.801	0.033	0.599
Variance	0.022	0.018	0.050	0.016	0.048	0.039	0.038	0.040
MSE	0.023	0.386	0.693	0.016	0.050	0.680	0.039	0.398
n = 2500								
Bias	0.029	0.597	0.763	0.029	0.063	0.762	0.061	0.592
Variance	0.004	0.003	0.011	0.003	0.012	0.009	0.010	0.009
MSE	0.005	0.360	0.594	0.004	0.016	0.589	0.014	0.359

Table 7: Table 3. Sensitivity analysis of simulation study 1. We consider nine different perturbations of the simulation parameters. Results are from 1,000 simulations of data (n = 500), using a parametric spline to adjust for the estimated propensity score.

	Variable(s) in propensity score model							
	X_1	X_2	X_3	X_1, X_2	X_1, X_3	X_2, X_3	X_1, X_2, X_3	None
std1 = 0.5								
Bias	-0.003	0.281	0.364	-0.006	-0.001	0.360	-0.005	0.278
Variance	0.031	0.021	0.047	0.021	0.041	0.037	0.029	0.037
MSE	0.031	0.101	0.180	0.021	0.041	0.166	0.029	0.115
std1 = 1.5								
Bias	0.008	0.855	1.009	0.007	0.004	1.013	0.008	0.847
Variance	0.033	0.028	0.048	0.024	0.046	0.040	0.035	0.041
MSE	0.033	0.759	1.066	0.024	0.046	1.066	0.035	0.759
std2 = 0.5								
Bias	0.003	0.595	0.738	-0.001	0.001	0.734	-0.004	0.599
Variance	0.007	0.015	0.020	0.005	0.011	0.018	0.009	0.017
MSE	0.007	0.369	0.564	0.005	0.011	0.557	0.009	0.376
std2 = 1.5								
Bias	0.015	0.602	0.751	0.014	0.019	0.748	0.014	0.586
Variance	0.106	0.055	0.146	0.090	0.147	0.122	0.121	0.137
MSE	0.106	0.418	0.710	0.090	0.147	0.680	0.121	0.480
std3 = 0.5								
Bias	-0.003	0.698	0.732	-0.004	-0.002	0.732	-0.003	0.691
Variance	0.032	0.025	0.045	0.022	0.038	0.035	0.027	0.041
MSE	0.032	0.512	0.581	0.022	0.038	0.571	0.027	0.518
std3 = 1.5								
Bias	-0.001	0.495	0.734	-0.004	-0.006	0.734	-0.008	5.07
Variance	0.027	0.022	0.055	0.021	0.049	0.044	0.037	0.041
MSE	0.028	0.267	0.593	0.021	0.049	0.582	0.038	0.518
a4 = 0.25								
Bias	-0.004	0.593	0.731	0.000	-0.007	0.737	0.000	0.600
Variance	0.032	0.025	0.049	0.023	0.044	0.039	0.033	0.040
MSE	0.032	0.377	0.583	0.023	0.044	0.582	0.033	0.399
a4 = 1								
Bias	0.004	0.601	0.735	0.005	0.004	0.735	0.005	0.597
Variance	0.033	0.027	0.052	0.023	0.047	0.042	0.035	0.040
MSE	0.033	0.387	0.592	0.023	0.047	0.582	0.035	0.397
b0 = 1								
Bias	-0.012	0.562	0.687	-0.010	-0.015	0.690	-0.012	0.575
Variance	0.034	0.021	0.052	0.022	0.047	0.039	0.032	0.041

4 Discussion

4.1 Replicability

In general, the data-generating mechanism is easy to replicate. It is clearly described in the original paper and it can be easily coded. Also, experiment 1 can be replicated with only a few replication degrees of freedom. However, availability of open source code will improve its replicability. The initial lack of explanation for the c-statistic or the issue with the model with splines could have been solved easily if the code was available. Experiment two is more difficult to replicate, or, at least, the figures on the paper. The model estimation and the computation of the measures are similar to experiment one. However, once these results are obtained, it is necessary to use some transformations that are not explained at all in the paper. Also, as stated previously, replicating a complex figure is not an easy task.In fact, we were not able to do so. The fact that we were not able to replicate experiment 2 does not mean that it is not possible. It is likely that more experienced researchers in the field would be able to replicate experiment 2.

4.2 Replicator degrees of freedom

In experiment 1, we needed to guess some concepts due to the lack of information on the article. However, the choices were straightforward and it we were able to replicate the experiment. In experiment 2, the lack of information about some concepts and scales did not allow us to replicate the experiment. Probably, when writing the paper the authors addressed an audience with a higher experience in the field than the authors of this report. However, it is always a good practice to write a paper as self-explanatory as possible. This will help reaching a wider audience and reduce the replicator degrees of freedom. At the same time, these technical details are not relevant for understanding the main concept or the message the authors want to present. They are mainly necessary for the purpose of replication.

4.3 Equivalence of results

The equivalence of results in experiment 1 is extremely good. Despite using a different R version and code, we were able to obtain very similar same tables. We cannot judge the results of experiment 2 because we were not able to obtain the same figures.

5 Comparison with other pair performing the same replication study

The other pair was able to perform an even more successful replication. Both groups were able to replicate experiment 1 with similar issues that we solved similarly yet independently. On the other hand, the other pair was able to replicate figure 2 and figure 3 from the original paper and almost able to replicate figure 4.
A discussion between both groups leaded to the following conclusions:

The order in which the method, measures and results are presented in the paper can be improved in order to facilitate the understanding and replicability of the research.
The original paper omit some basic information assuming it is common knowledge. However, we do not consider it common knowledge. This fact could explain why one of the groups was able to replicate more than the other. Actually, in latter versions of the paper they included explanations of some concepts (e.g. meaning of c-statistic).
The first experiment is explained with more detail, which makes it easier to replicate. On the other hand, both groups consider that experiment two contains some concepts that are not fully explained, which hurts its replicability.
Both groups consider that the original code of the research could be available publicly since it is coded in an open source programming laguage (R), and it does not require new functions that could be under copyrigth. Publishing the code will improve the replicability and transparency of the research.

6 Contributions

Authors made the following contributions according to the CRediT framework https://casrai.org/credit/
Primary Replicator:

- Data Curation
- Formal Analysis (supporting)
- Investigation
- Software
- Visualization (lead)
- Writing - Original Draft Preparation
- Writing - Review & Editing

Co-Pilot:

- Formal Analysis (lead)
- Investigation
- Software (supporting)
- Visualization (supporting)
- Validation
- Writing - Review & Editing

References

10 Brookhart, M Alan, Sebastian Schneeweiss, Kenneth J Rothman, Robert J Glynn, Jerry Avorn, and Til Stürmer. 2006. "Variable Selection for Propensity Score Models." *American Journal of Epidemiology* 163 (12): 1149–56. https://doi.org/10.1093/aje/kwj149.

Rougier, Nicolas P, Konrad Hinsen, Frédéric Alexandre, Thomas Arildsen, Lorena A Barba, Fabien C Y Benureau, C Titus Brown, et al. 2017. "Sustainable Computational Science: The {{ReScience}} Initiative." *PeerJ Computer Science* 3 (December): e142. https://doi.org/10.7717/peerj-cs.142.

Appendix

6.1 Code organization

The code and the files associated are organized in the form of a research compendium which can be found in the following git repository `https://github.com/replisims/Brookhart-2006-FJ`

```
## .
## +-- collectResults.R
## +-- computeResults.R
## +-- computeResultsExp2.R
## +-- defs.aux
## +-- defs.tex
## +-- flowchart.PNG
## +-- getResults.R
## +-- Lato-Black.ttf
## +-- Lato-BlackItalic.ttf
## +-- Lato-Bold.ttf
## +-- Lato-BoldItalic.ttf
## +-- Lato-Italic.ttf
## +-- Lato-Regular.ttf
## +-- MainScript.R
## +-- Method_exp1.R
## +-- Method_exp2.R
## +-- Method_new.R
## +-- Method_old.R
## +-- My Collection.bib
## +-- MyDataGeneration.R
## +-- MyEvaluationPC.R
## +-- plots.R
## +-- Preparation.R
## +-- README.md
## +-- references.bib
## +-- replicationFelix.Rproj
## +-- Replisims Juan and Felix.Rmd
## +-- Replisims-Juan-and-Felix.log
## +-- Replisims-Juan-and-Felix.pdf
## +-- Replisims-Juan-and-Felix.Rmd
## +-- Replisims-Juan-and-Felix.tex
## +-- replisims2.Rmd
## +-- ResultsEnvironment.RData
## +-- SimulationAllCells.R
## +-- UbuntuMono-Bold.ttf
## +-- UbuntuMono-BoldItalic.ttf
## +-- UbuntuMono-Italic.ttf
## +-- UbuntuMono-Regular.ttf
## \-- Untitled
##     +-- defs.aux
##     +-- defs.tex
##     +-- flowchart.PNG
##     +-- Lato-Black.ttf
##     +-- Lato-BlackItalic.ttf
##     +-- Lato-Bold.ttf
##     +-- Lato-BoldItalic.ttf
```

```
##     +-- Lato-Italic.ttf
##     +-- Lato-Regular.ttf
##     +-- references.bib
##     +-- UbuntuMono-Bold.ttf
##     +-- UbuntuMono-BoldItalic.ttf
##     +-- UbuntuMono-Italic.ttf
##     +-- UbuntuMono-Regular.ttf
##     +-- Untitled.log
##     +-- Untitled.pdf
##     +-- Untitled.Rmd
##     \-- Untitled.tex
```

Reproducibility Information

This report was last updated on 2022-03-06 16:00:08. The simulation replication was conducted using the following computational environment and dependencies:

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 4.0.3 (2020-10-10)
##  os       Windows 10 x64
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  Spanish_Spain.1252
##  ctype    Spanish_Spain.1252
##  tz       Europe/Berlin
##  date     2022-03-06
##
## - Packages -------------------------------------------------------------------
##  package        * version    date       lib
##  cachem           1.0.5      2021-05-15 [1]
##  callr            3.7.0      2021-04-20 [1]
##  cli              3.0.0      2021-06-30 [1]
##  crayon           1.4.1      2021-02-08 [1]
##  desc             1.3.0      2021-03-05 [1]
##  devtools         2.4.2      2021-06-07 [1]
##  digest           0.6.27     2020-10-24 [1]
##  dplyr          * 1.0.2      2020-08-18 [1]
##  ellipsis         0.3.2      2021-04-29 [1]
##  evaluate         0.14       2019-05-28 [1]
##  fansi            0.5.0      2021-05-25 [1]
##  fastmap          1.1.0      2021-01-25 [1]
##  fs               1.5.0      2020-07-31 [1]
##  generics         0.1.0      2020-10-31 [1]
##  glue             1.4.2      2020-08-27 [1]
##  htmltools        0.5.1.1    2021-01-22 [1]
##  knitr          * 1.36       2021-09-29 [1]
##  lifecycle        1.0.0      2021-02-15 [1]
##  magrittr         2.0.1      2020-11-17 [1]
##  memoise          2.0.0      2021-01-26 [1]
##  pillar           1.6.1      2021-05-16 [1]
##  pkgbuild         1.2.0      2020-12-15 [1]
##  pkgconfig        2.0.3      2019-09-22 [1]
##  pkgload          1.2.1      2021-04-06 [1]
```

```
##  prettyunits      1.1.1      2020-01-24 [1]
##  processx         3.5.2      2021-04-30 [1]
##  ps               1.6.0      2021-02-28 [1]
##  purrr            0.3.4      2020-04-17 [1]
##  R6               2.5.0      2020-10-28 [1]
##  remotes          2.4.0      2021-06-02 [1]
##  RepliSimReport   0.0.0.9000 2022-03-05 [1]
##  rlang            0.4.11     2021-04-30 [1]
##  rmarkdown        2.11       2021-09-14 [1]
##  rprojroot        2.0.2      2020-11-15 [1]
##  sessioninfo      1.1.1      2018-11-05 [1]
##  stringi          1.6.2      2021-05-17 [1]
##  stringr          1.4.0      2019-02-10 [1]
##  testthat         3.0.4      2021-07-01 [1]
##  tibble           3.1.2      2021-05-16 [1]
##  tidyselect       1.1.1      2021-04-30 [1]
##  usethis          2.0.1      2021-02-10 [1]
##  utf8             1.2.1      2021-03-12 [1]
##  vctrs            0.3.8      2021-04-29 [1]
##  withr            2.4.2      2021-04-18 [1]
##  xfun             0.26       2021-09-14 [1]
##  xtable         * 1.8-4      2019-04-21 [1]
##  yaml             2.2.1      2020-02-01 [1]
##  source
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.2)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  Github (replisims/RepliSimReport@5f14003)
```

```
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.3)
##  CRAN (R 4.0.0)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.5)
##  CRAN (R 4.0.2)
##  CRAN (R 4.0.0)
##
## [1] C:/Users/jclaramunt/Documents/R/win-library/4.0
## [2] C:/Program Files/R/R-4.0.3/library
```

The current Git commit details are:

```
## Local:    master C:/Users/jclaramunt/surfdrive/replicationFelix
## Remote:   master @ origin (https://github.com/jclaramunt/replicationFandJ)
## Head:     [32b0e4a] 2021-03-18: Last commit
```