VERSION 1

AUG 16, 2018

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.spmedk6

Protocol Citation: Lloyd Balbuena 2018. UK REF 2014 Analysis Data and R Script.
protocols.io
<https://dx.doi.org/10.17504/protocols.io.spmedk6>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
 We use this protocol and it's working

Created: Aug 16, 2018

Last Modified: Aug 16, 2018

PROTOCOL integer ID:
 14797

UK REF 2014 Analysis Data and R Script V.1

Lloyd Balbuena¹

¹University of Saskatchewan

Lloyd Balbuena

ABSTRACT

This collection of files contains the raw data and R script for the paper entitled 'The UK Research Excellence Framework and the Matthew Effect: Insights from machine learning.'

Data:

The file **balbuena_REF_2014.dta** is a Stata 12 file that contains the data for the paper. It is also saved in CSV format as **Balbuena_REF_2014.csv**

The file **balbuena_ref_syntax.R** contains the syntax used to run the analysis.

The file **balbuena_REF.RData** file contains the data objects in RData format. See the instructions below if you wish to replicate the analysis. To simply view the data in Excel without re-running the analysis, the universities in the training set (n = 79) and testing set (n=30) are provided in the excel file "REF_2014_schools.xls"

ATTACHMENTS

[balbuena_REF_2014.csv](#) [balbuena_REF_2014.dta](#) [balbuena_REF.RData](#)

- 1 Open the syntax file in R / RStudio.

☐ balbuena_ref_syntax.r

- 2 Load the RData file using the command Session | Load Workspace. If you wish to import the Stata or Excel files (#1), use the Part I section of the syntax file. Otherwise, go straight to Part II.

☐ balbuena_ref.rdata

☐ balbuena_ref_2014.dta

☐ balbuena_ref_2014.csv

- 3 Run the commands line by line or only the parts you wish to replicate.

- 4 If you wish to replicate the BART analysis only, go straight to Part III after having loaded the data.