

A Practical Guide to Conducting Systematic Literature Review Using R Software

Priyalatha Govindasamy^{a*}, Amelia Mohd Noor^a, Nor Junainah Mohd Isa^a, Antonio Olmos^b, Kathy Green^c

^a Faculty of Human Development, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, MALAYSIA

^b Aurora Research Institute, Colorado, USA

^c University of Denver, Colorado, USA

Editor:

Zahari Suppian, PhD

Abstract

A systematic literature review (SLR) is a highly rigorous inquiry of existing literature that addresses a formulated question. The review systematically searches, identifies, selects, appraises, and synthesizes research evidence relevant to the question that makes an SLR rigorous. Practicing this rigorous scientific inquiry can be overwhelming for first-time users. In the effort to assist the researchers, tools are introduced to ease their SLR inquiries. Therefore, this paper aims to provide the reader with practical steps to conduct a systematic review using R software. In this demonstration, we describe a systematic review and its challenges briefly. Then, readers are guided thoroughly with the steps from managing until sharing the bibliographic information with other co-researchers to conduct a systematic literature review. Every step of the procedure is explained in detail supported by the codes for R and screenshots of the output.

Keywords: *Systematic literature review; review guide; R software*

1. Introduction to Systematic Review

Systematic literature review (SLR) is an organized approach to collecting, evaluating, integrating, and presenting findings from multiple research studies to address research questions on a specific topic of interest (Snyder, 2019). It provides an overall understanding of the investigated topic. SLR adheres to a standardized guideline from searching articles till reporting the findings of the study. Among the most followed guide in the systematic literature review is the Preferred Reporting Items of Systematic Review (PRISMA). These standardized guidelines are used to ensure the credibility and transparency of the review have been met to generate a reliable outcome. Besides guidelines, organizations such as Cochrane and the Campbell Collaboration support systematic reviews by registering and storing reviews to be accessible to the people of interest on the topics. The repositories for systematic literature reviews are intended to provide a reliable platform to gather evidence on the topic studied for decision making or further actions.

The increase in primary literature requires a systematic review to accurately and comprehensively present the accumulated knowledge on the topic studied (Borah, Brown, Capers, & Kaiser, 2017). Although systematic literature review is on the rise, it is considered as time - consuming with a high

requirement of human resources. Unfortunately, not many have the luxury of time nor human power to successfully conduct a SLR. To overcome these challenges, the authors found that the free, open-source statistical software R can assist researchers in conducting a systematic literature review. The packages in R software can save time and reduce the need for extensive human resources in conducting a systematic review efficiently. Therefore, this paper is aimed to share best practices in conducting SLR efficiently with tools in R. This includes steps in extracting and reading bibliographic files saved under various file formats, and guiding ways to manage and utilize bibliographic files for a systematic review.

In this article, we use R (R Core Team, 2014) to demonstrate the implementation and use of R software in conducting systematic literature reviews efficiently. The use of R is increasing gradually in all fields because it is very powerful, it is continuously updated and maintained by the top statisticians in the world, and it is open-source software so it is free. Several interfaces can be used to run the software. The authors use R-studio (RStudio, 2015). The R software requires different packages, just like other statistical programs require specific routines for specific purposes (i.e., modules in SPSS). The user can download those packages from the Internet and install them, and it is recommended to run updates

*✉ : gpriyalatha@fpm.upsi.edu.my

from time to time. There are a very large number of packages intended to run almost every imaginable analysis. Readers are encouraged to search for resources in the R project for statistical computing website (<http://www.r-project.org/>), for notes on how to download and for a list of packages supported under R.

This paper is aimed to provide a comprehensive guide in illustrating steps from managing until screening for suitable articles for a systematic literature review. Searching is a process of finding relevant articles for the topic of interest from research-based search engines. As a result of the search, a list of potential articles are populated by the repositories. The repositories provide an overview of the articles in the form of bibliographic data. Generally, the data consist of basic descriptive information of an article which includes title, journal name, author's name, abstract and more. The depth of bibliographic data varies across the repositories. Moreover, the repositories are also found to have different file formats for storing the data. Among the common bibliographic file formats are rich information system (ris), BibTeX bibliography (bib), and comma separated values (csv). It is crucial to combine files from different article repository because researchers need to filter the search list to identify the relevant and suitable articles for their topic of study. Combining the files helps researchers to sort out duplicates and irrelevant articles from the searched list. However, repositories are not consistent in reporting their` data despite being in the same file format.

For example, some apply capital letters for the labels, names the labels differently or do not provide the information at all. Due to the inconsistency, researchers need to organize and clean the data before processing any bibliographic file.

Upon preparing the files, the researcher can begin reviewing the title and abstract to filter articles for the final study. The abstract and title review can be shared among the team members. The review decision from different team members can be merged back into a single file. With the identified articles, researchers could find the full paper to review and extract information for the systematic review. In the following section, the authors will illustrate how R software can be used to facilitate the bibliographic data management for an efficient systematic literature review. In section A, examples of merging bibliographic files from different file formats are presented. Section B presents data preparation that includes cleaning and organization. Then, cleaned data will be distributed among the team members for review (Section C). The R software is used to illustrate steps within sections A-C.

2. Section A: Merging files of different format in R

This section cover steps in loading and merging files from various formats into a single file. The R code and output are presented below (Figure 1-4). The result for setting up, importing and merging files can be viewed under the Environment tab and console window in RStudio. Figure 4 presents the example of a merged file saved in csv format.

2.1 Setting up R

First, the working directory has to be set (Figure 1). It is followed by loading the required packages for this review process. Users are required to install these packages if they do not have revtools (Westgate, 2019), tidyverse (Wickham et al., 2019), and textclean (Rinker, 2018) packages installed in their Rstudio. The statements followed by single # represent comments while double ## refereed to outcome from the functions (Figure 1).

```
# Set working directory
setwd("C:/Users/S406U/Desktop/BSI-18")
# Loading required library
library("revtools") # contains function for review
library("tidyverse") # contains function for data manipulation

## -- Attaching packages ----- tidyverse 1.3.0
--

## v ggplot2 3.2.1   v purrr  0.3.3
## v tidyr  1.0.0   v stringr 1.4.0
## v readr  1.3.1   v forcats 0.4.0

## Warning: package 'stringr' was built under R version 3.6.2

## -- Conflicts -----
- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

Figure 1: Setting up working directory

2.2 Import bibliographic data into R

Bibliographic data from different format files are uploaded in RStudio. It is advised that all downloaded bibliographic files should be saved under the working directory specified earlier. In this example, the

bibliographic files from ScienceDirect and Eric search engines are retrieved. The file from ScienceDirect are saved in ris format while files from the Eric search engine are saved in nbib format (Figure 2).

```
# Read data files saved from Sciencedirect and Eric repository
Sciencedirect <-read_bibliography("Test.ris") # ris format
Eric <-read_bibliography("ERIC2019-07-30_23.25.26.nbib") #nbib format
```

Figure 2: Loading files into RStudio

2.3 Extract only relevant information

The bibliographic data provides multiple types of information relating to the articles. Therefore, it is the user's responsibility to carefully select the relevant information for their review. Not all the bibliographic files are consistent in the provided information. Consequently, users need to check the assigned names to the column in each bibliographic file. Next, the users need to extract the information from specific columns. In the following example, there are similar

and differently labelled columns between Ericdata and Sciendata (Figure 3). Upon extracting the required columns, the labels or headings of the columns are standardized across the two datasets. New variables or columns are created and filled with "NA". These variables are added specifically to the bibliographic dataset that does not provide the information. By adding variables, the number of columns and headings will be standardized across the dataset thus making it easier to merge them into a single data file.

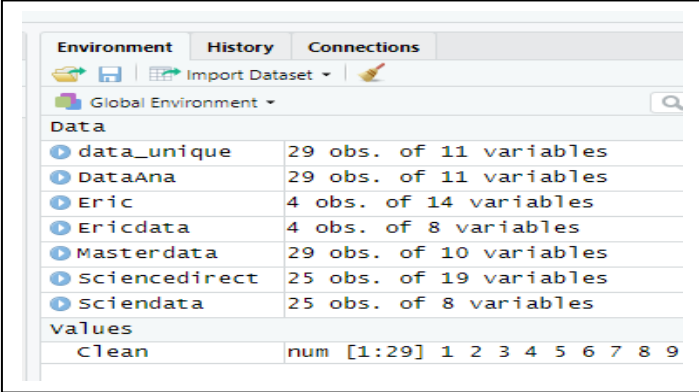
```
## Select the required column from the dataset
Ericdata <-Eric %>% select("label", "author", "DP", "journal","title", "doi")
Sciendata <-Sciencedirect %>% select("label", "author", "year", "journal", "title", "doi",
"keywords", "abstract")

## rename the column to have standardized label
names(Ericdata)[names(Ericdata) == "DP"] <- "year"

## added a missing column and filled in with "NA"
Ericdata$keywords <-NA
Ericdata$abstract <-NA

## Merge datasets into a single file
###Since, both datasets have equal column, rowbind is used to bind them.
Masterdata <-rbind(Sciendata, Ericdata)
## rename the column name to all uppercase to standardize the lable. This is optional
colnames(Masterdata) <- toupper(colnames(Masterdata))
## saving the combined data into a csv file. This data will be stored under the pre-defined
working directory
write.csv(Masterdata, "Masterdata.csv")
```

Figure 3: Modifying the data file for standardization



The screenshot shows the RStudio Environment pane with the following data objects:

Data	Observations	Variables
data_unique	29 obs.	11 variables
DataAna	29 obs.	11 variables
Eric	4 obs.	14 variables
Ericdata	4 obs.	8 variables
Masterdata	29 obs.	10 variables
Sciencedirect	25 obs.	19 variables
Sciendata	25 obs.	8 variables

Below the data objects, the 'values' section shows a 'Clean' variable with a numeric vector of length 29, containing values 1 through 9.

Figure 4: Example of a merged data file

3. Section B: Data preparation

3.1 Data preparation

This section covers steps in cleaning and managing data for review. The variable "title" is recoded into a new variable called "TITLEnew" where the contents are changed to small letters. Next, the modification was performed on the content of doi column. Some

cells in DOI column contains "doi" character followed by numbers. To standardize, the specific string (doi) was trimmed to retain only the values. Figure 5 presents the R code for data preparation. The outcome of the data preparations presented in Figures 6 and 7.

```
## Here, the information are organized and cleaned prior to the analysis
### the title variable is recoded into a new variable with lower capital letters
Masterdata$TITLEnew<-tolower(Masterdata$TITLE)

### remove apostrophe on titlenew
library("textclean")

## Warning: package 'textclean' was built under R version 3.6.2

Masterdata$TITLEnew<-strip(Masterdata$TITLEnew, apostrophe.remove = TRUE)
### discard part of a string character
library("stringr")
Masterdata$DOInew <-str_remove(Masterdata$DOI, "doi.")
```

Figure 5: Code for data cleaning through recoding and trimming

	F	G	H	I	J	D
	TITLE	DOI	KEYWORDS	ABSTRACT	TITLEnew	D
nomics	The scale of Work-Related Affective Fe	doi:10.1016/j.aperg	Work-related affect	Interest in the field	the scale of workrel	1
ffective D	The influence of the working alliance o	10.1016/j.jad.2019.	Bipolar disorder anc	Background The wo	the influence of the 1	1
	Chapter 36 - Sleep Disorders in Chronic do	10.1016/B978-0-12	Actigraphy and Chr	Sleep is an essential	chapter sleep disorc	1
	Chapter 4 - Intervention research on th	1016/B978-0-12	Behavioral intervent	The efficacy of psyc	chapter interventio	1
ffective D	The NEO-FFI domain of openness to ex	1016/j.jad.2019.	Ketamine and Temp	Background There	the neoffi domain o	1
ffective D	Suicide prevention Using the number o	1016/j.jad.2019.	Suicidal ideation an	Background Suicid	suicide prevention 1	1
	Chapter 78 - Pain and Chronic Kidney I	10.1016/B978-0-12	Chronic Kidney Dise	Several studies eval	chapter pain and ch	1
	56 - Bone Metastases	10.1016/B978-0-32	bone metastases an	Bone metastasis is a	bone metastases 1	1
ffective Di	Psychometric properties of a Haitian Ki	10.1016/j.jad.2019.	Haiti and Non-partn	Background In this	psychometric prope	1
	Chapter 109 - Postherpetic Neuralgia	10.1016/B978-0-32	Herpes zoster and N	Postherpetic neural	chapter postherpeti	1
ffective Di	Effectiveness and adherence of telepho	10.1016/j.jad.2019.	Telephone and Psyc	Background The aim	effectiveness and ac	1
ffective Di	Is prenatal maternal distress context-d	10.1016/j.jad.2019.	Pregnancy and Anxi	Background Matern	is prenatal materna	1
azardous	Investigation of polycyclic aromatic hy	10.1016/j.jhazmat.2	Polycyclic aromatic	The concentrations	investigation of pol	1
ierontolo	NA	10.1016/j.archger.2	Burden and Caregivi	Objective This study	NA	1
ffective Di	Focusing on insomnia symptoms to be	10.1016/j.jad.2019.	Depression and Insc	Background Disturb	focusing on insomn	1
nd Individ	Grateful people are happier because th	10.1016/j.paid.2019.	Time perspectives a	Past studies demon	grateful people are	1

Changed the original title to all lower cases.

Figure 6: Result of recoding variable

Initially, we combined bibliographic data from various files format. Consequently, there is a possibility for the same article to appear in more than one dataset. With the use of R code, we can identify and eliminate duplicate entries. In this example, the title of the articles is used to find the duplicates. Any entries that have an exact match on their title will be counted and the frequency will be presented. The selection of matching variables are dependent on the users. Additionally, the entries with missing title information are also excluded. Only the single and complete entries will be saved into a new dataset.

Figure 8 presents the R code for identifying and removing duplicates and missing entries and saving them into a new dataset. This is followed by the results of finding duplicates in the data files. The "n_duplicates" column in the data file indicates the frequency of the duplicate entries. The red boxed examples shows that the title appears twice in the combined dataset retrieved from various search engines (Figure 6). The elimination of entries is shown through the reduction of observations from the data file. (Figure 10).

G	H	I	J	K
DOI	KEYWORDS	ABSTRACT	TITL Enew	DOI new
doi.10.1016/j.apergo.2019.10.016	Work-related affect	Interest in the field	the scale of workre	10.1016/j.apergo.2019.10.016
doi.10.1016/j.jad.2019.10.016	Bipolar disorder and	Background The wo	the influence of the	10.1016/j.jad.2019.10.016
doi.10.1016/B978-0-12-819122-4	Actigraphy and Chr	Sleep is an essential	chapter sleep disord	10.1016/B978-0-12-819122-4
10.1016/B978-0-12-819122-4	Behavioral intervent	The efficacy of psyc	chapter intervention	10.1016/B978-0-12-819122-4
10.1016/j.jad.2019.10.016	Ketamine and Temp	Background There a	the neoffi domain o	10.1016/j.jad.2019.10.016
10.1016/j.jad.2019.10.016	Suicidal ideation an	Background Suicide	suicide prevention u	10.1016/j.jad.2019.10.016
10.1016/B978-0-12-819122-4	Chronic Kidney Dise	Several studies eval	chapter pain and ch	10.1016/B978-0-12-819122-4
10.1016/B978-0-323-85443-3	bone metastases an	Bone metastasis is a	bone metastases	10.1016/B978-0-323-85443-3

First part of the content (doi.) is trimmed

Figure 7: Results of trimming a particular character from the content

3.2 Finding and eliminating duplicates

```
### Identify the duplicate entries in the data
Clean <-find_duplicates(Masterdata, match_variable = "TITLnew")
### Extract the unique articles and discard the duplicates
data_unique <-extract_unique_references(Masterdata, Clean)

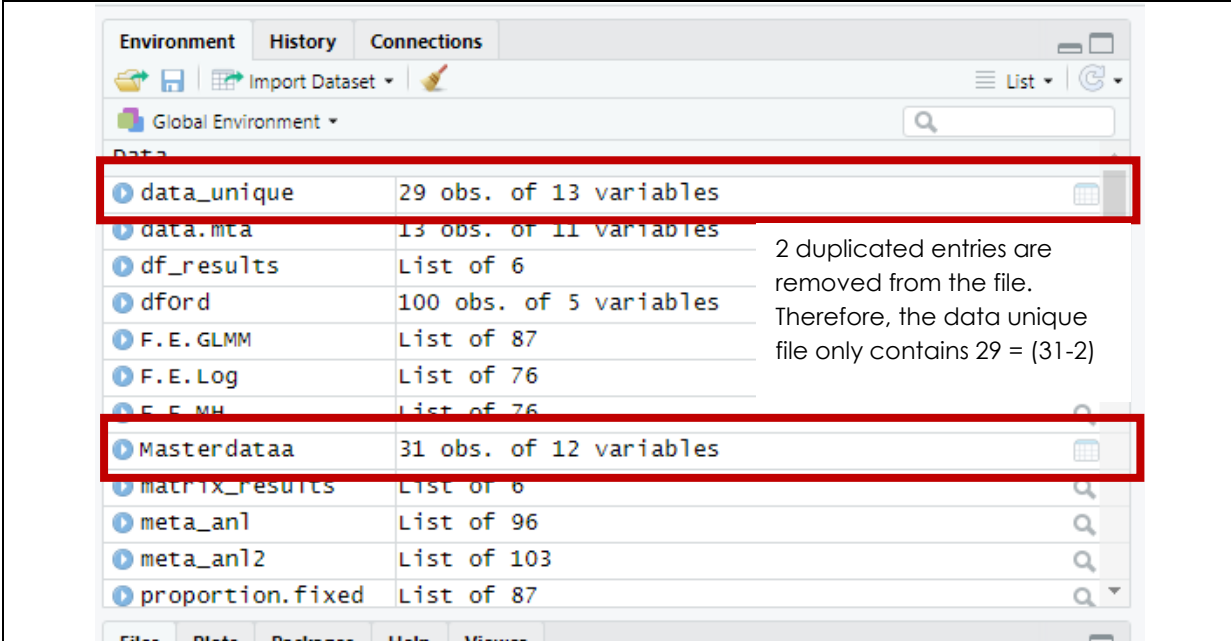
##remove rows of TITLnew column that contains NA
DataAna <-data_unique[!is.na(data_unique$TITLnew),]

##save the data to distribute to different readers.
write.csv(DataAna, "DataAna.csv")
```

Figure 8: Code for finding and eliminating duplicate entries

	DOInew	n_duplicates
related affective feelings woraf	10.1016/j.apergo.2019.102945	1
he working alliance on the treatment and o...	10.1016/j.jjad.2019.09.014	1
orders in chronic kidney disease	10.1016/B978-0-12-815876-0.00036-X	1
ion research on therapies that aim to treat ...	10.1016/B978-0-12-815015-3.00004-0	2
1 of openness to experience moderates ket...	10.1016/j.jjad.2019.09.010	1
n using the number of health complaints a...	10.1016/j.jjad.2019.08.025	1
chronic kidney disease	10.1016/B978-0-12-815876-0.00078-4	1
	10.1016/B978-0-323-47674-4.00056-6	1
properties of a haitian kreyl version of the tra...	10.1016/j.jjad.2019.09.004	1
etic neuralgia	10.1016/B978-0-323-54947-9.00109-7	1
adherence of telephoneadministered psyc...	10.1016/j.jjad.2019.09.023	2
tal distress contextdependent comparing u...	10.1016/j.jjad.2019.09.048	1
olycyclic aromatic hydrocarbons in soils fr...	10.1016/j.jhazmat.2019.121158	1

Figure 9: Indication of duplicate entries in a combined dataset



Environment History Connections

Global Environment

Data

data_unique	29 obs. of 13 variables	
data.mta	13 obs. of 11 variables	
df_results	List of 6	2 duplicated entries are removed from the file.
dford	100 obs. of 5 variables	Therefore, the data unique file only contains 29 = (31-2)
F.E. GLMM	List of 87	
F.E. Log	List of 76	
F.E. MH	List of 76	
Masterdataa	31 obs. of 12 variables	
matrix_results	List of 6	
meta_an1	List of 96	
meta_an12	List of 103	
proportion.fixed	List of 87	

Files Plots Packages Help Viewer

Figure 10: Changes in the number of observations after eliminating duplicate entries

4. Section C: Distributing data for review

Once the bibliographic data is cleaned, the content is ready to be reviewed. Here, users can divide and distribute the data to multiple reviewers. In this example, we list three names of the reviewers. Then, the task to review the articles for the study is equally divided among the three reviewers. Once the tasks are

divided equally, the function automatically splits the master data file into three separate csv format files. These files are saved under the names of the reviewers in the pre-defined working directory. Figure 11 presents the R code for saving, distributing and screening bibliographic data files to reviewers. Figure 12 presents an example of the reviewer's files.


```
## save and distribute to article list to different readers.
```{r}
read cleaned data
write.csv(DataAna, "DataAna.csv")
List the reviewer
reviewers <- c("Amy", "June", "Priya")
TaskDis <- distribute_tasks(DataAna, reviewers = reviewers)
Screen abstracts
Priya <- read.csv("reviewer_Priya.csv")
screen_abstracts(Priya)
```
```

Figure 11: Codes to save and distribute data files for multiple reviewers

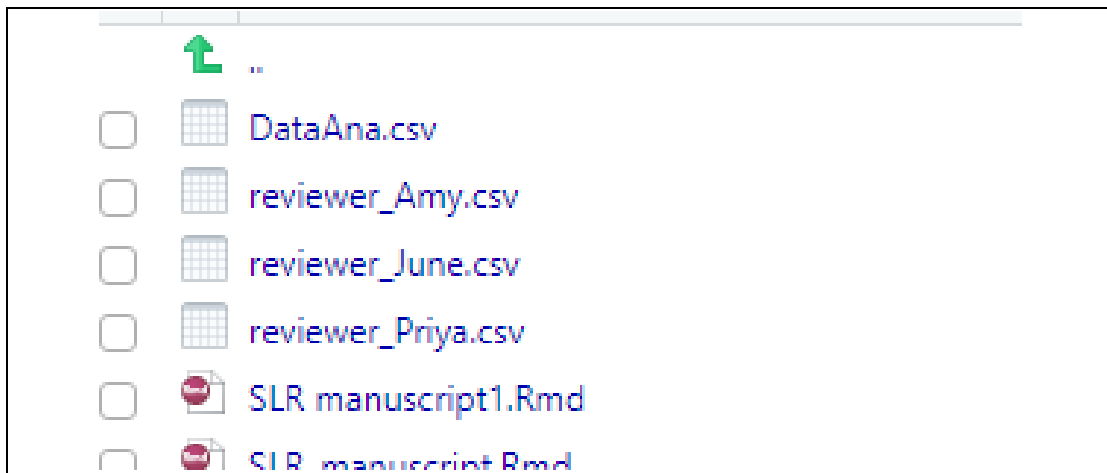


Figure 12: Example of reviewers file

4.1 Screen title and abstracts

In this step, reviewers start screening the assigned files to them. Reviewers screen through the title and abstracts of the articles to determine their suitability of the study for the review. The shiny app that be obtained from R can be generated to help users to review the bibliographic list. It allows researchers to screen the title and abstracts to finalize articles for the review. Using the interface, researchers can decide on

the articles by clicking “select” or “exclude” options. Upon selection, the abstract and title will change its color to blue and red otherwise. Additionally, researchers can include comments about the abstract in the show notes windows section (Figure 13). Save the decision “Vetted reviewer_Priya.csv. The results from the review can be saved and opened in a csv format (Figure 14)

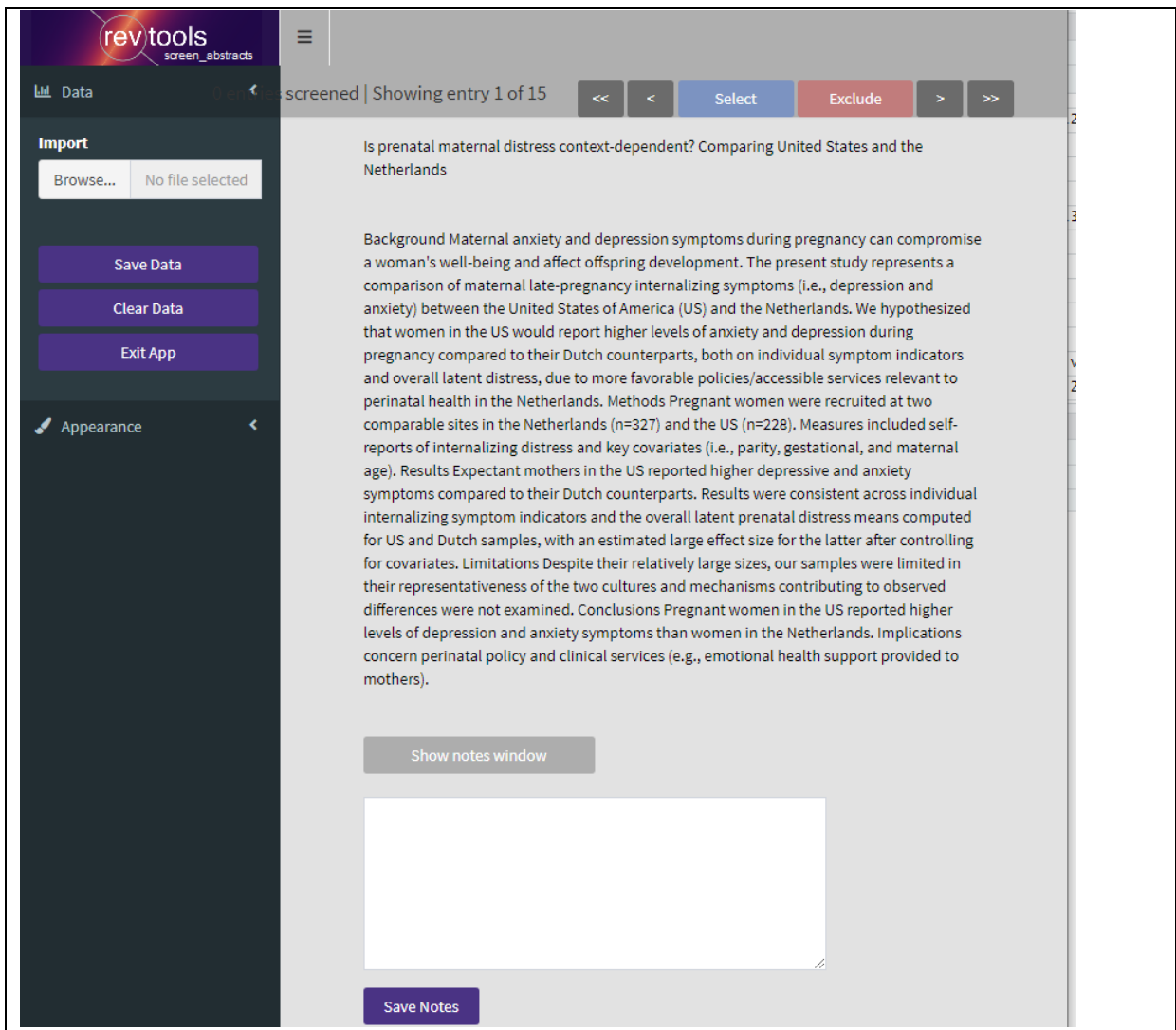


Figure 13: Shiny app interface for reviewing purposes

| doi | keywords | abstract | titlenew | cpinew | n_duplicate | order_initi | order_e | order_ran | order_sele | color | selected | notes |
|-----------|-------------|-------------|-------------|-----------|-------------|-------------|---------|-----------|------------|---------|----------|-------------------------|
| 10.1016/j | Bipolar dis | Backgrou | the influer | 10.1016/j | 1 | 1 | 1 | 2 | 2 | #993f3f | excluded | |
| 10.1016/E | Behavioral | The efficac | chapter in | 10.1016/E | 2 | 2 | 2 | 11 | 11 | #000000 | NA | |
| 10.1016/j | Ketamine | Backgrou | the neoffi | 10.1016/j | 1 | 3 | 3 | 10 | 10 | #000000 | NA | |
| 10.1016/j | Haiti and N | Backgrou | psychome | 10.1016/j | 1 | 4 | 4 | 3 | 3 | #405d99 | selected | Psychometric properties |
| 10.1016/j | Pregnancy | Backgrou | is prenatal | 10.1016/j | 1 | 5 | 5 | 1 | 1 | #405d99 | selected | Relevant to the study |
| 10.1016/j | Polycyclic | The conce | investigati | 10.1016/j | 1 | 6 | 6 | 7 | 7 | #000000 | NA | |
| 10.1016/j | Time pers | Past studi | grateful pe | 10.1016/j | 1 | 7 | 7 | 15 | 15 | #000000 | NA | |
| 10.1016/E | cancer anc | Cancer-rel | chapter ca | 10.1016/E | 1 | 8 | 8 | 13 | 13 | #000000 | NA | |
| 10.1016/j | NA | Backgrou | white mat | 10.1016/j | 1 | 9 | 9 | 5 | 5 | #000000 | NA | |
| 10.1016/E | Cardiovas | Epidemiol | chapter ep | 10.1016/E | 1 | 10 | 10 | 4 | 4 | #000000 | NA | |

Figure 14: Data file of selected decision and notes

5. Conclusion

A systematic literature review is a comprehensive approach that helps understand the overview of an investigated topic. For that, researchers are responsible to find all resources related to the topic and rigorously review them for evidence synthesis. The systematic process becomes tedious because of the need for researchers to search vast research repositories and organize identified titles for further synthesis. Therefore the application of tools will ease some of the tedious jobs in a review. Besides, the use of tools in the review can reduce human errors when conducting a systematic review. Thus provides transparency in the process and credibility to the results of the systematic review. Therefore, this practical guide is aimed to introduce the capability of an open-source tool called R in helping researchers conduct a systematic review. This guide is hoped to provide insights on ways to utilize free software in expediting the systematic review process. The introduction of free tools is anticipated to reduce human time in searching and organizing but provides an opportunity to focuses more on synthesizing the content.

Citation

Govindasamy, P., Mohd Noor, A., Mohd Isa, N. J., Olmos, A., & Green, K. (2020). A Practical Guide to Conducting Systematic Literature Review Using R Software. *Evaluation Studies in Social Sciences*, 1(1), 35-43. <https://doi.org/10.37134/esss.vol1.1.4.2020>

Reference

- Borah R, Brown, A.W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *British Medical Journal Open*, 7. DOI:10.1136/bmjopen-2016-012545.
- R Core Team (2014). R: A language and environment for statistical computing. (3.0.3) [Computer software]. Vienna, Austria: Foundation for Statistical Computing.
- Rinker, T. W. (2018). *textclean*: Text Cleaning Tools. R packages version 0.9.3.
- RStudio (2012). RStudio: Integrated development environment for R (Version 0.96.122) [Computer software]. Boston, MA. Retrieved May 20, 2012.
- Synder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. DOI: 10.1016/j.jbusres.2019.07.039
- Westgate, M. J. (2019). *revtools*: Tools to support evidence synthesis. R package version 0.4.1.
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.