Table of Contents

S1. Documenting the Retrieval Process

Table S1.1

Template for Documenting Records

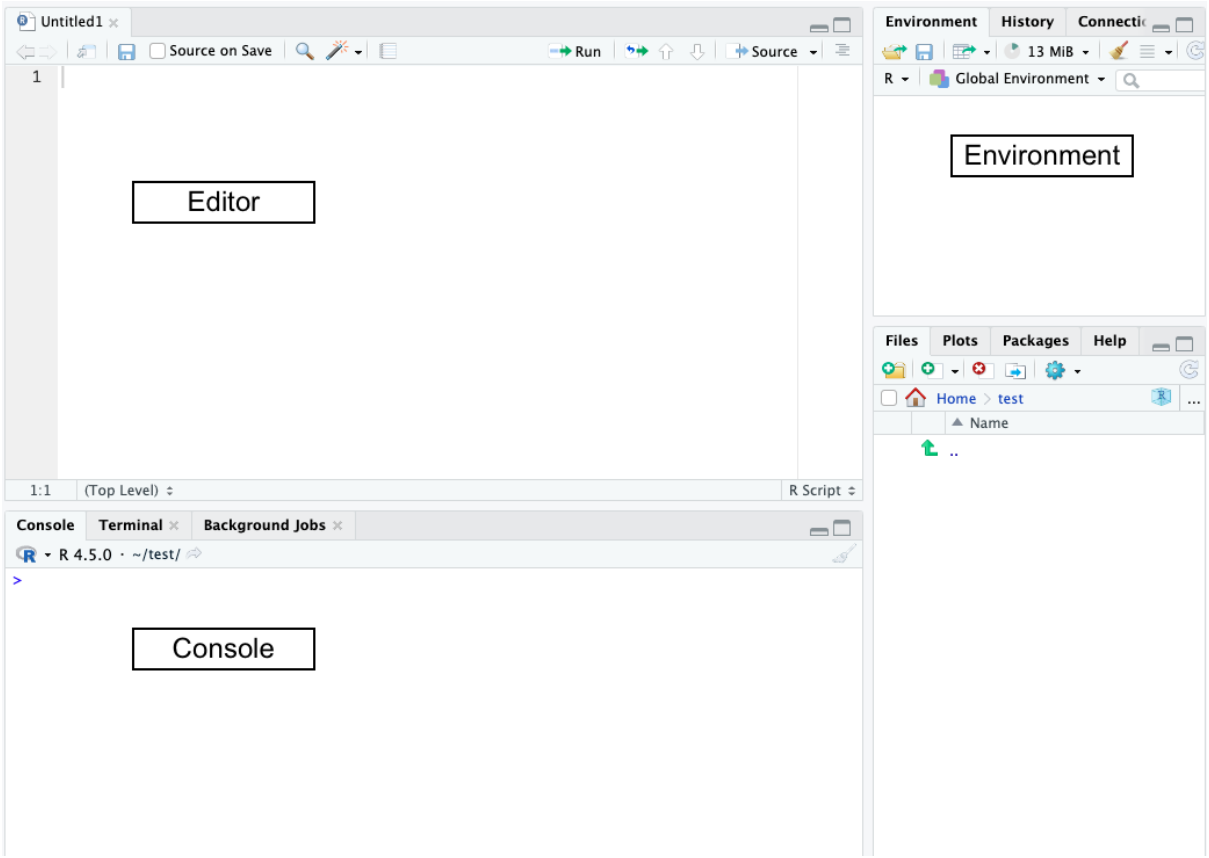
Database Search *Keywords: [Insert the Search String]*			
Database	**Number of Studies Retrieved**	**Date of Retrieval**	**Retrieval Conducted by Who**
ERIC			
PsycInfo			
PubMed			
Scopus			
Web of Science			
TOTAL **With duplicates**			
TOTAL **Without duplicates**			

S2. Data Import and Preparation in RStudio

Before analysing the data, the dataset must first be imported into RStudio. There are a few terminologies that researchers need to be familiar with when navigating the interface of RStudio, which would be: the console, the environment, and the editor. First, the console is where the results and output of the code will appear. Next, the environment is where loaded datasets, variables, and objects are listed. Lastly, the editor is the space for researchers to write their code. Figure S2.1 illustrates the interface of RStudio.

Figure S2.1

Interface of RStudio



To begin, researchers should open a new R Script to write their code. This can be done using either one of the two proposed methods. First, in the bottom-right panel of the RStudio interface, click on “New Blank File” and select “R Script” from the dropdown menu (Figure S2.2). Alternatively, researchers may navigate to the top menu bar, click on “File” and select “New File” from the dropdown menu, and click on “R Script” (Figure S2.3).

Figure S2.2

Opening R Script

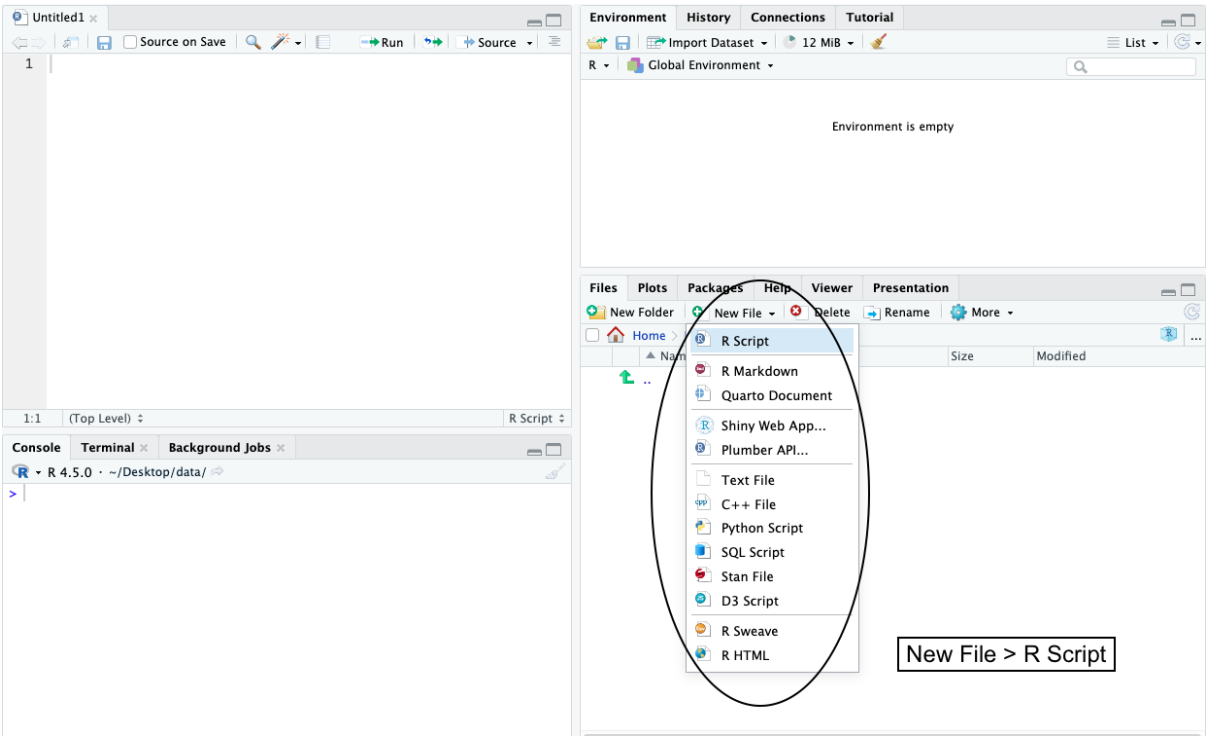
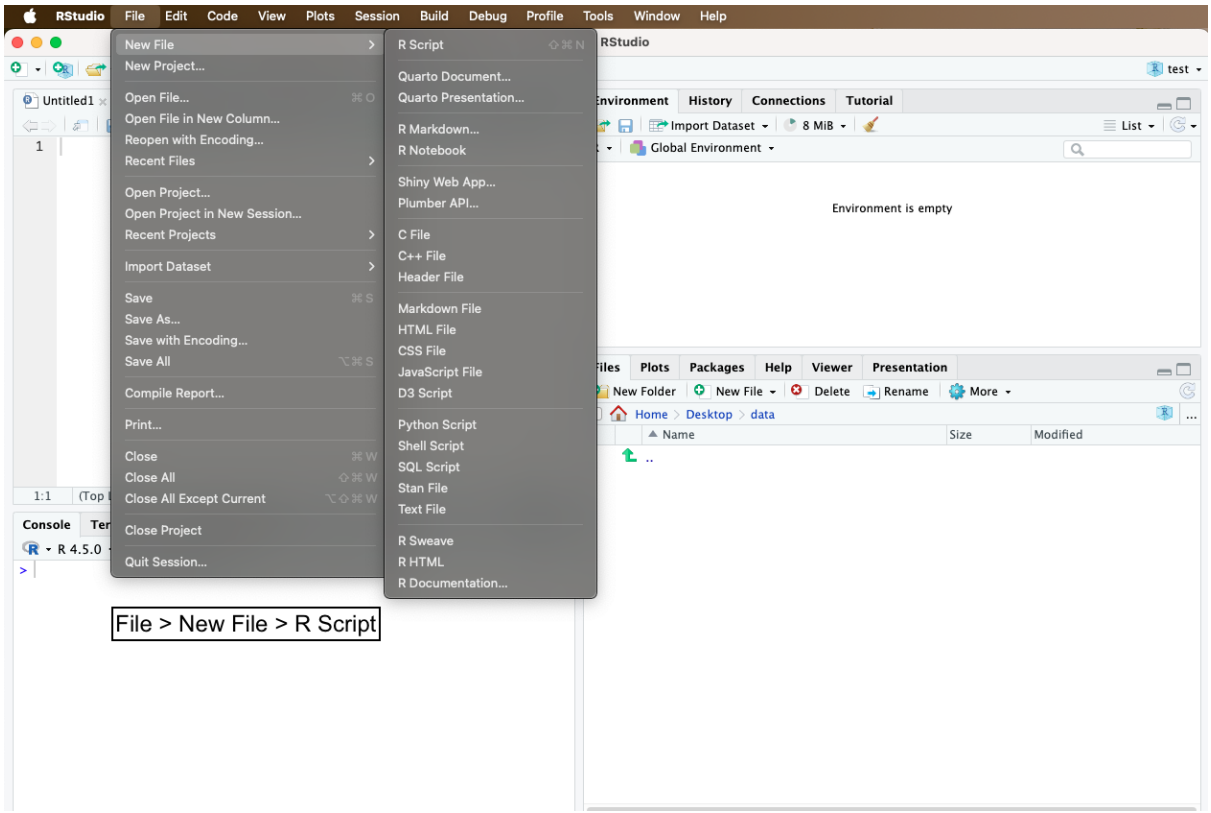


Figure S2.3

Opening R Script



Once the R Script is open, the next step is to set the working directory. Setting the working directory ensures that RStudio knows where to locate the dataset that is to be analysed. This can be achieved through one of the two proposed ways. Researchers may navigate to the bottom-right panel of the RStudio and look for the folder where the dataset was saved. Click on the cogwheel labelled “More” and select “Set as Working Directory” from the dropdown menu (Figure S2.4). Alternatively, researchers may navigate from their menu bar, click on “Session”, followed by “Set as Working Directory”, and select “Choose Directory” from the dropdown menu (Figure S2.5). Afterwards, select the folder where the data file was saved and click open. For both methods, a command starting with `setwd()` should appear in the console (Figure S2.6). Researchers are encouraged to copy that

command and paste it into the R Script editor for future reference (Figure S2.6). Once the working directory is set, researchers may use the following code to read in the data:

```
data = read.csv("SPC.csv")
```

In this command, data serves as the object where the dataset is saved to. Researchers may replace “data” with a name of their choice. Researchers can also replace “SPC.csv” with the actual name of their dataset file.

Figure S2.4

Setting the Working Directory

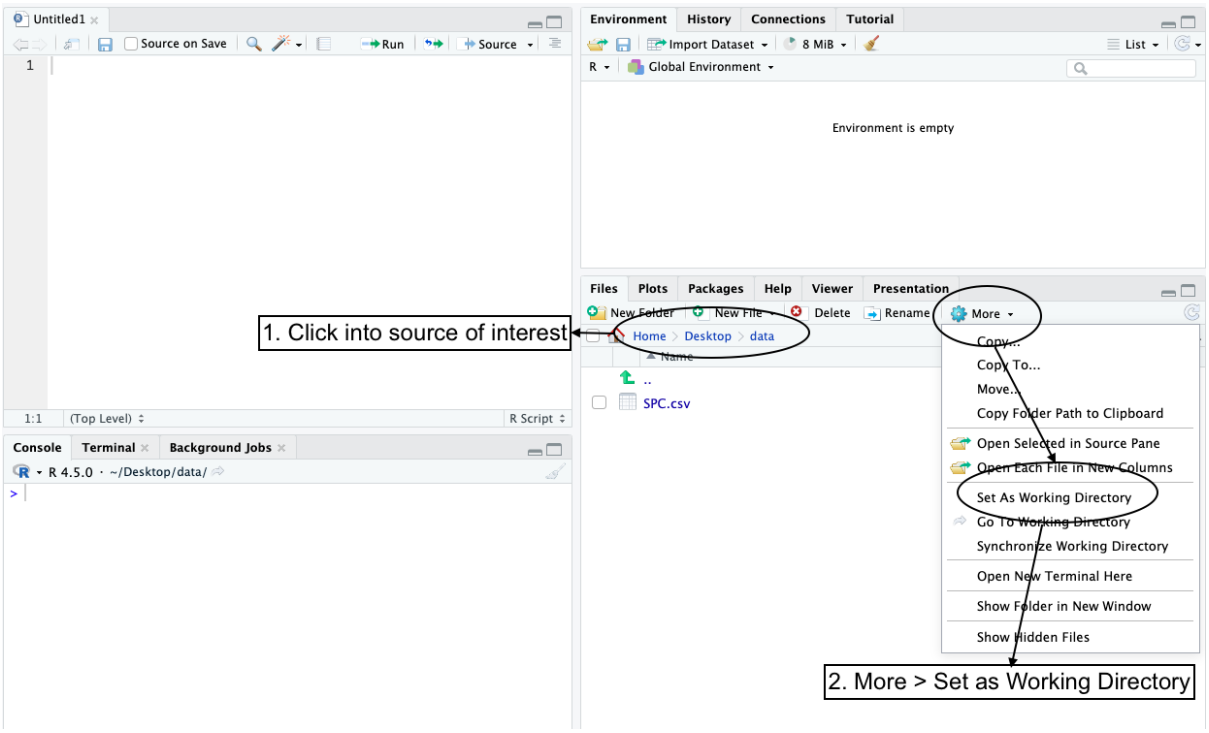


Figure S2.5

Setting the Working Directory

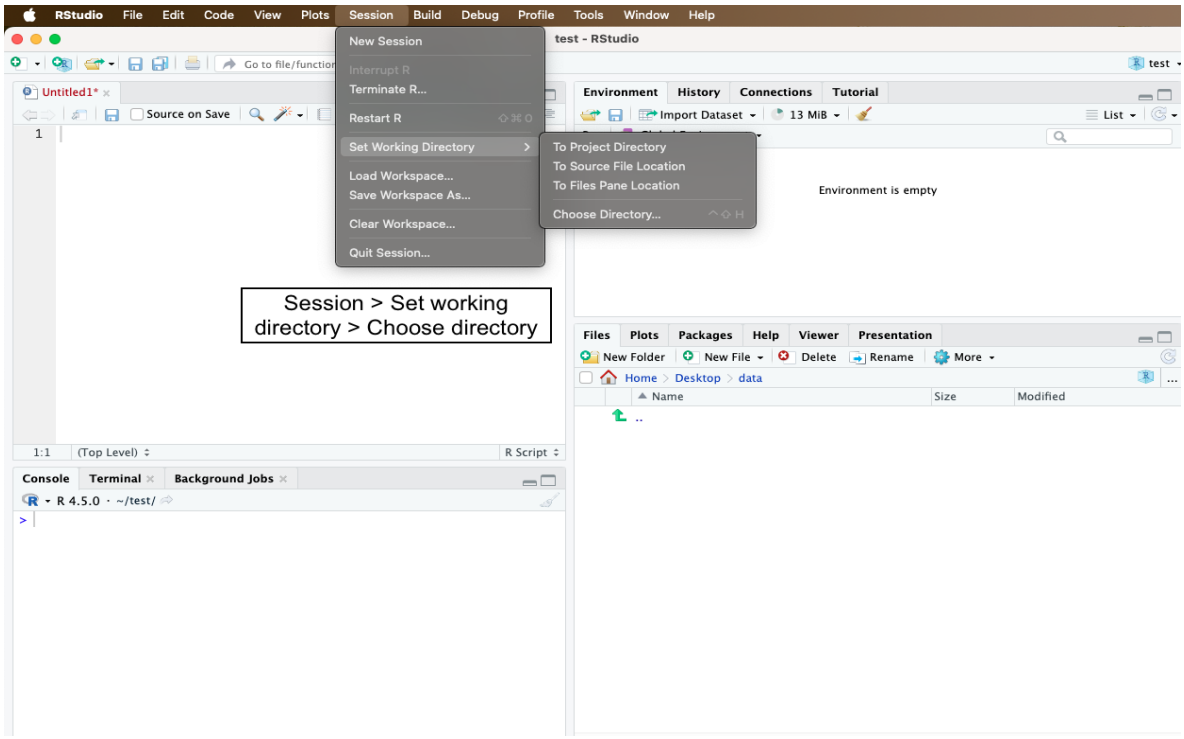
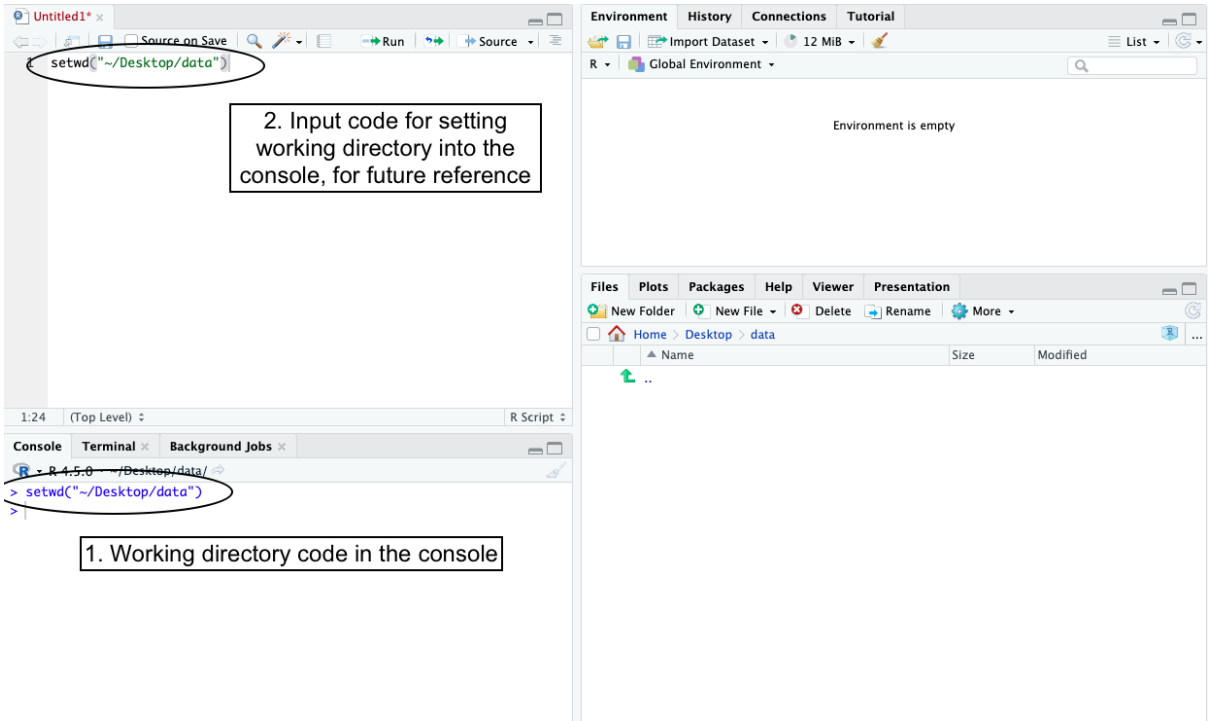


Figure S2.6

Saving Working Directory Code into the Console



S3. Inter-Rater Agreement Methods

Cohen's κ

Cohen's κ, symbolised by the lowercase Greek letter, κ, ranges from 0 to +1. A value of 0 indicates the expected agreement between the screeners by random chance, while +1 represents perfect agreement between the screeners (McHugh, 2012). Calculating Cohen's κ is another method researchers may employ in calculating IRR/IRA in this stage. The following contingency table illustrates how researchers may compute Cohen's Kappa:

	Rater B: Yes	Rater B: No	Total
Rater A: Yes	a (e.g., 3)	c (e.g., 5)	e (e.g., 8)
Rater A: No	b (e.g., 1)	d (e.g., 7)	f (e.g., 8)
Total	g (e.g., 4)	h (e.g., 12)	N (e.g., 16)

1. To calculate Po, which is the proportion of cases where the screeners agree:

$$Po = \frac{a + d}{N}$$

Substituting values:

$$Po = \frac{3 + 7}{16} = 0.625$$

2. To calculate Pe, which is calculated based on the marginal totals of the contingency table:

$$Pe = [(\frac{e}{N}) \times (\frac{g}{N})] + [(\frac{f}{N}) \times (\frac{h}{N})]$$

Substituting values:

$$Pe = [(\frac{8}{16}) \times (\frac{4}{16})] + [(\frac{8}{16}) \times (\frac{12}{16})] = 0.5$$

3. To calculate Cohen's Kappa,

$$\kappa = \frac{Po - Pe}{1 - Pe}$$

Substituting values:

$$\kappa = \frac{0.625 - 0.5}{1 - 0.5} = 0.25$$

To interpret the final results of the IRR/IRA, Table 6 consolidates the standard interpretation of Cohen's κ value (Bajpai et al., 2015; Landis & Koch, 1977; McHugh, 2012). Higher values indicate greater agreement between the raters. Researchers should note that agreement levels below 'moderate' may require retraining the screeners or revising the screening criteria (McHugh, 2012).

Table S3.1

Interpretation of the Values of Cohen's κ

Value of Cohen's κ	Level of Agreement
.00-.20	None
.21-.39	Minimal
.40-.59	Weak
.60-.79	Moderate
.80-.90	Strong
Above .90	Almost perfect

Percent Agreement Method for Calculating IRR/IRA

Each screener will provide their evaluations for each criterion per paper by indicating their agreement or disagreement. To determine the overall inter-rater reliability, the percent agreement values for all the criteria are averaged, providing a raw agreement rate expressed as a percentage (%). The formula for this method is as follows:

$$\text{Raw Agreement} = \frac{Number\ of\ agreements}{Total\ number\ of\ records\ rated} \text{ x } 100$$

Each screener will provide their evaluations in a singular spreadsheet. For each criterion in each paper, screeners will indicate their agreement or disagreement as suggested:

- 'TRUE' or 'MATCH' or '1' to indicate agreement between the screeners
- 'FALSE' or 'RESOLVE' or '0' to indicate disagreement between the screeners

This process should be completed for each criterion across all papers. Afterwards, the inter-rater reliability for each criterion can be calculated using the aforementioned formula. This calculation should be repeated for each criterion individually. For example, if 30 out of 40 evaluations for a criterion are in agreement, the IRR/IRA for that criterion will be calculated as so: $\frac{30}{40} \times 100\% = 75\%$.

To determine the overall inter-rater reliability, the percent agreement values for all the criteria are averaged, providing a raw agreement rate expressed as a percentage (%). To facilitate the process, we recommend using the following formulas that can be used in either Google Sheets or Microsoft Excel, depending on which software the researcher has chosen to conduct the abstract and title screening with:

- =IF formula: To automate the marking of agreement or disagreement between the screeners
- =COUNTIF and =AVERAGE formula: To calculate the raw agreement rate between the screeners

Table S3.2 illustrates the process of calculating IRR or IRA with the percent agreement method.

Table S3.2

Percent Agreement Method Process of Calculating Inter-Rater Reliability with Microsoft Excel or Google Sheets

Screener 1 Abstract Screening Sheet Tab

	Column A	Column B	Column C	Column D
Row E	Paper	Criteria 1	Criteria 2	Criteria 3
Row F	ABC	Yes	Yes	No
Row G	EFG	No	Yes	No
Row H	HIJ	Yes	Yes	Yes

Screener 2 Abstract Screening Sheet Tab

	Column A	Column B	Column C	Column D
Row E	Paper	Criteria 1	Criteria 2	Criteria 3
Row F	ABC	Yes	No	Yes
Row G	EFG	No	Yes	Yes
Row H	HIJ	Yes	Yes	Yes

Inter-Rater Calculation (1)

	Column A	Column B	Column C	Column D
Row E	Paper	Criteria 1	Criteria 2	Criteria 3
Row F	ABC	=if(BF of Screener 1 = BF of Screener 2, “MATCH”, “RESOLVE”)	=if(CF of Screener 1 = CF of Screener 2, “MATCH”, “RESOLVE”)	=if(DF of Screener 1 = DF of Screener 2, “MATCH”, “RESOLVE”)
Row G	EFG	=if(BG of Screener 1 = BG of Screener 2, “MATCH”,	=if(CG of Screener 1 = CG of Screener 2, “MATCH”,	=if(DG of Screener 1 = DG of Screener 2, “MATCH”,

		“RESOLVE”)	“RESOLVE”)	“RESOLVE”)
Row H	HIJ	=if(BH of Screener 1 = BH of Screener 2, “MATCH”, “RESOLVE”)	=if(CH of Screener 1 = CH of Screener 2, “MATCH”, “RESOLVE”)	=if(BG of Screener 1 = BG of Screener 2, “MATCH”, “RESOLVE”)

Inter-Rater Calculation (2)

	Column A	Column B	Column C	Column D
Row E	Paper	Criteria 1	Criteria 2	Criteria 3
Row F	ABC	MATCH	RESOLVE	RESOLVE
Row G	EFG	MATCH	MATCH	RESOLVE
Row H	HIJ	MATCH	MATCH	MATCH
Row I		=COUNTIF(BE :BH, “MATCH”) /3	=COUNTIF(CE :CH, “MATCH”) /3	=COUNTIF(DE :DH, “MATCH”) /3
= AVERAGE (BI:DI) %				

The percent agreement method is straightforward to calculate (Gisev et al., 2013), with the results that are easy to interpret (Bajpai et al., 2015). However, the percent agreement method does account for chance agreement, potentially overestimating the agreement rate (Bajpai et al., 2015; McHugh, 2012). In contrast, Cohen’s κ accounts for the possibility of guessing, but yields results that are less intuitive to interpret (Bajpai et al., 2015). The choice of method depends on the likelihood of the screeners guessing their evaluation of each criterion. If guessing is of a significant concern, Cohen’s κ would be more appropriate for calculating IRR/IRA (McHugh, 2012). On the other hand, if the screeners are well-trained and guessing is unlikely, the percent agreement method may be sufficient (McHugh, 2012).

S4. Data Structure Formats

There are two main formats: wide and long. In a wide format, each study will occupy a single row, with separate columns representing a variable or outcome of interest, making it suitable for meta-analyses where the included studies report the same set of variables In a long format, each study may have multiple rows, with each row corresponding to a variable or outcome of interest, which is ideal for meta-analyses with multiple outcomes or for more complex methods such as multilevel meta-analyses. The format of the dataset also depends on the *R* packages that will be utilised. For example, visualisation packages such as *ggplot2* require a long format while both forms of datasets are suitable to be used with the meta package. Ultimately, the format of the dataset depends on two key factors: the type of meta-analysis being conducted and the requirements of the *R* packages. Figures S4.1 and S4.2 illustrate examples of wide and long formats respectively.

Table S4.1

Example of a Wide Format Dataset

Author	Effect Size Outcome 1	Effect Size Outcome 2
ABC	0.4	0.5
EFG	0.1	0.3
HIJ	0.5	0.5

Table S4.2

Example of a Long Format Dataset

Author	Variable of Interest	Effect Size
ABC	Outcome 1	0.1
ABC	Outcome 2	0.4
EFG	Outcome 1	0.3

References

Bajpai, S., Bajpai, R., & Chaturvedi, H. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, *41*, 20–27.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.