2026 ■ Vol. 22 ■ no. 2

Conducting Meta-Analysis of Single Proportions Using R: A Practical Tutorial for Behavioral Researchers

Vikas Yadav [a], Tanwi Trushna [a], Akhil Dhanesh Goel [b], Uday Kumar Mandal [a] & Yogesh Damodar Sabde [a]

[a] ICMR National Institute for Research in Environmental Health (ICMR-NIREH), Bhopal, India

[b] All India Institute of Medical Sciences (AIIMS), Jodhpur, India

Abstract ■ Meta-analysis is a powerful statistical tool that combines results from multiple studies to produce a pooled estimate, providing a clearer, more accurate picture of the research question. In behavioral research, meta-analysis of single proportions, such as the prevalence of specific behaviors or psychological conditions, is particularly useful for quantifying behavioral patterns across diverse populations. This article offers a practical guide to conducting meta-analyses of single proportions in R. Using the R package 'meta', this guide explains how to conduct a meta-analysis using the Generalized Linear Mixed Model (GLMM) approach, perform subgroup, sensitivity and meta-regression analyses, and detect publication bias. The step-by-step R scripts provided are designed to make meta-analysis accessible to behavioral researchers with varying levels of statistical expertise, particularly in low-resource settings. By synthesizing data from multiple smaller studies, meta-analysis helps bridge critical knowledge gaps and supports the development of evidence-based behavioral research and intervention strategies. To support the reproducibility of the analyses, the R script and a sample dataset are available in a GitHub repository.

Keywords ■ Prevalence; Meta-regression; Publication bias; Sensitivity analysis; Subgroup analysis; Baujat Plot; Radial plot; funnel plot; forest plot; Doi plot. **Tools** ■ R.

drvikasyadav@gmail.com

10.20982/tqmp.22.2.p102

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ No reviewer

Introduction

Meta-analysis is a widely used statistical tool for pooling data from multiple studies to generate a single, more accurate estimate of an outcome compared to individual studies (Higgins et al., 2022; Patole, 2021). Estimating single proportions using meta-analysis is particularly valuable in the fields of behavioral research, where the prevalence of specific behavioral traits, symptoms, or conditions is essential for understanding the scope and distribution of psychological phenomena (Barendregt et al., 2013; Barker et al., 2021). Prevalence, which refers to the proportion of a population affected by a disease or trait at a given time, is a critical measure for guiding interventions, resource allocation, and policy-making (Evans, 2009; Noordzij et al., 2010).

In low- and middle-income countries, where conducting large-scale national surveys may be financially or logistically unfeasible, meta-analysis offers an effective alternative (Asogwa et al., 2022). By pooling data from smaller studies conducted in different regions or populations, researchers can generate more reliable estimates of disease prevalence across diverse settings (Ntais & Talias, 2024). These methods also allow for the combination of data from studies that vary in sample sizes and populations (Barker et al., 2021).

Conducting a meta-analysis of single proportions, however, requires the appropriate statistical approach. Among the available methods, the Generalized Linear Mixed Model (GLMM) is often the preferred approach for estimating the pooled effect, especially when dealing with binary outcomes such as disease prevalence (Schwarzer et al., 2019; Schwarzer & Rucker, 2022). In their 2019 study,

Schwarzer et al. pointed out that some commonly used transformations, such as the inverse Freeman-Tukey double arcsine transformation, can yield seriously misleading results when applied to meta-analyses of single proportions (Schwarzer et al., 2019). GLMM, on the other hand, provides a more robust approach by appropriately accounting for between-study variability and generating more accurate pooled estimates (Lin & Chu, 2020; Schwarzer et al., 2019). This makes it an excellent choice for behavioral researchers dealing with data on the prevalence of a specific disease, behavioral traits, or conditions.

GLMM analyses can be performed using various statistical software, but the most prominent are R and SAS (Daly & Soobiah, 2022; R Core Team, 2025; SAS Institute Inc, 2025). R, a powerful statistical tool, offers several packages that simplify the process of performing meta-analyses (Balduzzi et al., 2019; Viechtbauer, 2010). The `meta` package in R is widely used for conducting meta-analyses, providing comprehensive functionality for pooling effect sizes, testing for publication bias, and conducting subgroup analyses (Balduzzi et al., 2019; Schwarzer, 2006; Schwarzer & Rucker, 2022). Since R is an open-source software, it can be easily accessed by researchers in resource-constrained settings (Shimizu & Ferreira, 2023). Despite the simplicity of performing GLMM-based meta-analysis in R, health researchers seldom utilize it in practice (Lin & Chu, 2020). A 2020 review of the methodology of recent prevalence meta-analyses revealed that, despite statistical limitations, researchers often opt for other transformations instead of using the GLMM (Borges Migliavaca et al., 2020). This may be due to a lack of availability of clear guidance presented in simple, non-mathematical language.

Previous literature has described methods for conducting meta-analysis of proportions in R; however, resources such as Wang (2023), while comprehensive, are primarily concept-oriented and not specifically designed for step-by-step implementation and interpretation by applied researchers (Wang, 2023). The present manuscript addresses this gap by providing a simplified, fully reproducible, and user-friendly GLMM-based workflow.

A comprehensive, step-by-step guide could encourage behavioral researchers to adopt this statistically robust method. This manuscript, therefore, aims to guide researchers through the entire data analysis process, including preparing data for analysis, setting up the R environment, conducting a meta-analysis using the Generalized Linear Mixed Model (GLMM), visualizing the results, and performing advanced statistical techniques such as meta-regression and sensitivity analyses.

The practical focus of this article ensures that even researchers with limited statistical experience can perform their own meta-analyses of single proportions using R. However, readers should note that all analyses were conducted in R (version 4.5.3) using RStudio (version 2026.01.2+418 “Apple Blossom”) on a Windows 64-bit computer, with the `meta` package (version 8.3-0) and the `metasens` package (version 1.5-3) (Balduzzi et al., 2019; Posit Team, 2025; R Core Team, 2025; Schwarzer, 2006; Schwarzer et al., 2025). This script can be run on macOS or other operating systems with minor modifications. Users on different platforms may also notice slight variations in functionality or output.

Methodology:

Preparing data for analysis

Systematic review and meta-analysis involve several key steps, including: (a) formulating a clear research question; (b) developing and registering a protocol; (c) defining eligibility criteria; (d) conducting a comprehensive literature search; (e) screening and selecting studies; (f) extracting relevant data; (g) assessing study quality; (h) performing statistical analysis and evaluating heterogeneity; (i) assessing publication bias; and (j) interpreting and reporting the findings. In the present study, we focus specifically on steps (f), (h), (i), and (j) (Bashir & Conlon, 2018).

Once the relevant studies have been selected, the analysis phase commences, with the initial step being the meticulous extraction of data from the eligible studies. In a meta-analysis of single proportions, data are typically extracted from studies that report disease prevalence. Essential variables include the study name, year, setting (e.g., community vs. hospital), participants’ age, total sample size, number of positive cases, and quality scores. Gender-specific disease prevalence data (e.g., for males and females) may also be included if available. These data should be stored in a CSV file format to facilitate easy import into R.

We will use an example of ‘Disease X’ to demonstrate the format in which data should be prepared for this meta-analysis (see Table 1). Each included study will be identified by the `study` variable, which lists the study name (Author et al., year), along with the `year` of publication. The `setting` variable will indicate whether the study was conducted in a community-based or hospital-based setting, while the `age` variable will report the mean age of participants. The `total` variable will represent the overall sample size, and the `positive` variable will denote the number of confirmed cases of Disease X within the study population. For studies that provide gender-specific data, the `males` and `females` variables will indicate whether separate prevalence data are available for males and females, with `malespos` and `femalespos` recording the number of positive cases in each group. The `maletotal` and `femaletotal` variables will capture the total sample size

Table 1 ■ Extracted data for meta-analysis of 35 studies of 'Disease X' prevalence (available as supplementary file 'meta_analysis_data.csv' at https://github.com/ResearchCore/PrevalenceMetaR)

study	setting	age	total	positive	males	malepos	maletotal	females	femalespos	femaletotal	quality
Author 01 et al., 2018	Hospital	27	705	46	yes	23	380	yes	23	325	5
Author 02 et al., 2018	Hospital	21	100	8	yes	2	53	yes	6	47	5
Author 03 et al., 2014	Hospital	53	322	89	yes	13	150	yes	76	172	5
Author 04 et al., 2013	Hospital	45	1000	142	yes	7	482	yes	135	518	6
Author 05 et al., 2017	Hospital	44	333	46	yes	1	165	yes	45	168	6
Author 06 et al., 2014	Hospital	42	400	48	yes	12	170	yes	36	230	5
Author 07 et al., 2014	Hospital	47	334	42	yes	12	145	yes	30	189	8
Author 08 et al., 2016	Hospital	58	1062	215	yes	66	541	yes	149	521	4
Author 09 et al., 2016	Hospital	54	1062	215	yes	66	491	yes	149	571	4
Author 10 et al., 2013	Hospital	48	139	18	yes	4	72	yes	14	67	7
Author 11 et al., 2017	Hospital	36	200	20	yes	4	87	yes	16	113	6
Author 12 et al., 2018	Community	31	350	41	yes	14	176	yes	27	174	6
Author 13 et al., 2017	Community	37	800	81	yes	22	334	yes	59	466	6
Author 14 et al., 2016	Hospital	46	1072	186	yes	44	571	yes	142	501	5
Author 15 et al., 2018	Hospital	35	633	70	yes	29	302	yes	41	331	5
Author 16 et al., 2010	Hospital	49	541	85	yes	7	261	yes	78	280	6
Author 17 et al., 2008	Community	47	300	39	yes	9	145	yes	30	155	5
Author 18 et al., 2017	Hospital	37	1000	92	yes	7	543	yes	85	457	8
Author 19 et al., 2018	Hospital	57	461	105	yes	6	251	yes	99	210	6
Author 20 et al., 2015	Hospital	51	720	135	yes	20	401	yes	115	319	5
Author 21 et al., 2018	Community	34	1000	92	yes	28	367	yes	64	633	5
Author 22 et al., 2016	Hospital	32	1000	103	yes	34	554	yes	69	446	5
Author 23 et al., 2018	Community	59	400	136	yes	16	190	yes	120	210	4
Author 24 et al., 2017	Community	21	166	2	no			no			9
Author 25 et al., 2015	Hospital	28	400	32	no			no			6
Author 26 et al., 2013	Hospital	39	810	78	no			no			5
Author 27 et al., 2018	Hospital	27	483	24	no			no			7
Author 28 et al., 2011	Hospital	34	305	31	yes	30	145	yes	1	160	7
Author 29 et al., 2014	Community	25	250	15	yes	2	145	yes	13	105	5
Author 30 et al., 2019	Hospital	32	263	25	yes	9	125	yes	16	138	5
Author 31 et al., 2015	Hospital	24	500	28	yes	8	276	yes	20	224	6
Author 32 et al., 2016	Hospital	43	200	30	yes	9	109	yes	21	91	5
Author 33 et al., 2019	Community	29	1000	31	yes	6	400	yes	25	600	6
Author 34 et al., 2016	Hospital	24	1000	31	yes	6	448	yes	25	552	6
Author 35 et al., 2017	Hospital	52	100	67	yes	32	54	yes	35	46	4

for males and females, respectively. The `quality` variable represents the score assigned to each included study using a critical appraisal checklist specifically designed for prevalence studies, with scores ranging from 1 to 9, where 1 indicates the lowest methodological quality and 9 the highest.

Setting up the R environment

It is important to ensure that the subsequent steps are run on the latest versions of R, RStudio, and all necessary packages; failure to install these components correctly may lead to analysis errors. First-time users can download R from the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/bin/windows/base/, and the desktop version of RStudio from the Posit website at https://posit.co/download/rstudio-desktop/. Before starting the analysis, users will need to install the required packages, `meta` and `metasens` (Balduzzi et al., 2019; Schwarzer et al., 2025), if they have not been installed previously. These can be installed using the `install.packages` command:

```
# Installing Required Packages
install.packages("meta")
install.packages("metasens")
```

The `library()` function is used to activate installed packages in R, so their functions can be used. The `library(meta)` loads the `meta` package for conducting meta-analysis, `library(grid)` loads the `grid` package for graphical functions, and `library(metasens)` loads the `metasens` package for sensitivity analyses. After installing and activating the required packages, choose a folder to store all data files, code files, and exported graphs. Download the supplementary dataset (i.e., meta_analysis_data.csv) and R script files (i.e., meta_analysis_script.R) from the provided link (https://github.com/ResearchCore/PrevalenceMetaR) into this folder. The `getwd()` function can be used to check the

Listing 1 ■ Performing Meta-Analysis Using Random-Effects GLMM

```
prev.meta.glmm <- metaprop(
  event    = positive,
  n        = total,
  data     = data,
  studlab  = study,
  pscale   = 100,
  common   = FALSE,
  method   = "GLMM",
  prediction = TRUE
)
```

current working directory (i.e., folder where the files are stored) and ensure it is set correctly. The `setwd()` function helps to change it and update the folder path to match the user's system structure. This can be done by simply copying the folder path, pasting it inside the quotes, and replacing any backslashes (\) with forward slashes (/). Alternatively, in RStudio, the user can navigate to the desired folder using the 'Files' pane (located in the bottom right corner) and set the working directory through the 'More' menu. This ensures that all output files, such as graphs exported as PDFs, are stored in the correct location for analysis. These codes are detailed here.

```
# Loading Required Packages and
# Setting the Working Directory
library(meta)
library(grid)
library(metasens)
getwd()
setwd("folder_path")
```

To import data from a CSV file in R, the `read.csv()` function can be used. A convenient approach is to use `choose.files()` within this function. This opens a pop-up window that allows the user to browse their system and select the desired CSV file. Once selected, the file is imported into R, with the `header = TRUE` argument ensuring that the first row of the file is treated as column headers for the dataset. Use the `View(data)` command to display the imported dataset in a spreadsheet-like tabular format (see code below). This allows users to visually inspect the data, verify that variables have been imported correctly, and quickly identify potential issues such as missing values, incorrect variable types, or data entry errors.

```
# Importing and Viewing Data
data <- read.csv(choose.files(),
          header = TRUE)
View(data)
```

Meta-Analysis Using GLMM

The Generalized Linear Mixed Model (GLMM) approach is particularly well-suited for handling binary outcomes such as disease prevalence (Schwarzer et al., 2019; Schwarzer & Rucker, 2022). To perform the meta-analysis of single proportions, the `metaprop()` function is used (see Listing 1).

This command calculates the pooled proportion of positive cases across multiple studies, accounting for variability between them. In meta-analyses of prevalence, random-effects models are commonly used due to the inherent heterogeneity across studies, which is why the `common = FALSE` argument is used to exclude fixed-effects model results (Borges Migliavaca et al., 2020). The `pscale = 100` parameter ensures that the pooled prevalence is expressed as a percentage. Specifying `prediction = TRUE` in the metaprop function yields an additional prediction interval that reflects the expected range of prevalence in a future study, incorporating between-study heterogeneity.

Running `summary(prev.meta.glmm)` will provide a detailed overview of the meta-analysis. This meta-analysis output typically includes several key components (see Listing 2). It first reports the number of included studies (k), total number of observations (o), and total number of events (e), which describe the overall dataset used in the analysis. This is followed by the pooled estimate from the random-effects model along with its 95% confidence interval, representing the overall proportion (or prevalence) across studies while accounting for between-study variability. The output also includes the prediction interval and the measures of heterogeneity such as τ^2 (tau squared, indicating between-study variance), τ (its standard deviation), I^2 (percentage of total variability due to heterogeneity rather than chance), and H (relative excess variability; Choi & Kang, 2025; Cordero & Dans, 2021; Higgins et al., 2003).

τ^2 represents the absolute magnitude of between-study variance on the scale of the effect size; larger values indi-

Listing 2 ■ Summarizing Meta-Analysis Results and Heterogeneity Statistics

```
> summary(prev.meta.glmm)
# Individual study details omitted from this listing.
Number of studies: k = 35
Number of observations: o = 19411
Number of events: e = 2448

                        events             95%-CI
Random effects model 11.5168 [9.0439; 14.5577]
Prediction interval          [2.4788; 39.9936]

Quantifying heterogeneity (with 95%-CIs):
 tau^2 = 0.6270; tau = 0.7918; I^2 = 95.8% [94.9%; 96.5%]; H = 4.86 [4.41; 5.36]

Test of heterogeneity:
           Q d.f.  p-value
 Wald 803.67   34 < 0.0001
 LRT  931.39   34 < 0.0001

Details of meta-analysis methods:
- Random intercept logistic regression model
- Maximum-likelihood estimator for tau^2
- Calculation of I^2 based on Q
- Prediction interval based on t-distribution (df = 34)
- Logit transformation
- Clopper-Pearson confidence interval for individual studies
- Events per 100 observations
```

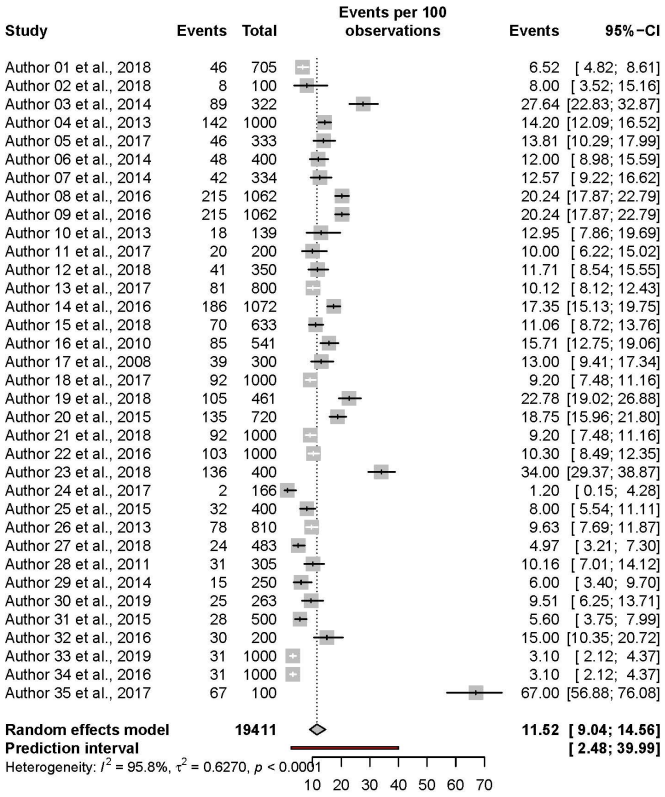
cate greater heterogeneity, though interpretation depends on the metric used (e.g., the logit scale). τ (the square root of τ^2) reflects the standard deviation of true effect sizes across studies and can be interpreted as the typical deviation of individual study estimates from the overall pooled effect (Borenstein, 2023; Choi & Kang, 2025).

I^2 represents the proportion of total variability in effect estimates that is due to between-study heterogeneity rather than chance. Although I^2 values of around 25%, 50%, and 75% have been commonly used as cut-offs to indicate low, moderate, and high heterogeneity, respectively, these thresholds are arbitrary and should be interpreted with caution in the context of meta-analysis of prevalence data (Borenstein et al., 2021; Higgins et al., 2003; Migliavaca et al., 2022). Importantly, I^2 does not convey the absolute magnitude of heterogeneity or the range of true effects across studies; therefore, it should not be used in isolation. Complementary measures such as τ, τ^2, and especially prediction intervals are more informative for understanding the dispersion and clinical relevance of effects across studies (Borenstein, 2023; Choi & Kang, 2025; Migliavaca et al., 2022).

Additionally, other tests of heterogeneity (e.g., the Wald test and the likelihood ratio test, LRT) provide the Q statistic with its corresponding degrees of freedom ('d.f') and p-value. A statistically significant p-value (typically < 0.05) indicates the presence of heterogeneity beyond chance (Cordero & Dans, 2021; Higgins et al., 2003). The output further specifies the methodological details used in the analysis, including the statistical model, the estimation method for τ^2, the transformation applied, and the approach for calculating confidence intervals for individual studies; all of which can be used in reporting results.

A prediction interval, as described previously, represents the expected range within which the true prevalence in a future study may lie, accounting for between-study heterogeneity. A wide prediction interval indicates substantial variability and limited generalizability of the pooled estimate. Unlike I^2, which only quantifies the proportion of variability due to heterogeneity, the prediction interval reflects its magnitude in real units and is therefore more informative for interpreting heterogeneity in prevalence meta-analysis (Choi & Kang, 2025; Migliavaca et al., 2022).

Figure 1 ■ Forest Plot of the Pooled Prevalence of Disease X



Visualizing Results:

A clear visual representation is key in meta-analysis. Saving outputs as PDF files is recommended because it preserves high-quality vector graphics that can be scaled without losing resolution. Additionally, PDF files can easily be converted to other journal-recommended formats, such as TIFF or PNG, using free online tools. Therefore, we have provided all the code required to generate the output as a PDF file. All codes listed for graphical outputs create files using the base R graphics device.

Forest Plot

A forest plot displays the results of each study and the pooled estimate. It provides a visual representation of the data and is essential for understanding the heterogeneity between studies (Patole, 2021). The commands for creating forest plots in R is detailed in the next instructions, and the output is shown in Figure 1.

```
# Generating a Forest Plot
# for Pooled Prevalence
pdf("forest.prev.meta.glmm.pdf",
    width = 10, height = 15)
forest(prev.meta.glmm, sortvar = study,
    common = FALSE)
grid.text(
  "Pooled Prevalence of Disease X",
  y = unit(0.98, "npc"),  # near top
  gp = gpar(fontsize = 16,
            fontface = "bold")
)
dev.off()
```

Running this code generates a PDF file named "forest.prev.meta.glmm.pdf" in the working directory, with dimensions specified as 10 × 15 inches. In R, the `pdf()` function defines figure dimensions in inches by default. To keep units clear, figure sizes may be planned in centimeters and then converted to inches (1 inch = 2.54 cm) for use in the code. This is because the `pdf()` function in R defines figure dimensions in inches by default and does not support alternative units such as centimeters. The `forest()` function draws a forest plot from the meta-analysis object `prev.meta.glmm`. The argument `sortvar = study` arranges studies alphabetically by study name in the out-

put. The option `common = FALSE` suppresses the fixed-effect (common-effect) summary, so that only the random-effects pooled estimate is displayed. If a user wishes to include the fixed-effect model results in the plot, this can be achieved by removing the `common = FALSE` argument (or by explicitly setting `common = TRUE`), which will add the fixed-effect pooled estimate and its confidence interval to the figure.

Notably, the `meta` package does not provide a direct option within `forest()` to add a custom title. When multiple forest plots are generated, this can make figure identification difficult; therefore, adding a descriptive title programmatically during plot creation is advisable to ensure figures remain easily recognizable, especially when multiple images are merged or shared. These titles can also be removed in the final publication if required. The `grid.text()` command (from the grid graphics system) is used to add such custom annotations to the plot after it has been drawn. In this example, it places a bold title, "Pooled Prevalence of Disease X", at the top of the figure. The argument `y = unit(0.98, "npc")` specifies the vertical position in normalized parent coordinates (npc), where 1 corresponds to the top of the plotting region; thus, a value of 0.98 positions the text just below the upper margin. The graphical parameters defined by `gp = gpar(fontsize = 16, fontface = "bold")` control the text appearance, allowing precise customization of font size and style without modifying the underlying forest plot.

Finally, `dev.off()` closes the PDF graphics device and saves the completed plot to the file. If not all studies are fully visible, the height argument can be increased to accommodate additional rows. To sort studies by effect size rather than by name, the sortvar argument may be changed to `sortvar = TE`. Minor overlap between test statistics and heterogeneity measures in the forest plot was avoided by enabling automatic width adjustment (`calcwidth.tests = TRUE` and `calcwidth.hetstat = TRUE`), which allocates sufficient horizontal space for overall effect tests and heterogeneity statistics (τ^2, I^2, Q). Alternatively, overlap can be reduced by removing selected x-axis tick marks using the at argument in the `forest()` function. By default, the forest plot automatically generates tick marks based on the scale and range of the data. If these default ticks (e.g., 0, 20, 40, 60, and 80) overlap with the heterogeneity statistics, they can be adjusted by specifying alternative positions, for example, `at = c(40, 60, 80)`. This removes the overlapping ticks (0 and 20) and retains only those that improve the clarity of the plot.

The forest plot visually summarizes the results of the meta-analysis, showing individual study estimates as squares with horizontal lines representing their 95% confidence intervals, where the size of each square reflects the study weight. The pooled estimate from the random-effects model is displayed as a diamond, with its width indicating the overall confidence interval (Brush et al., 2024; Chang et al., 2022; Sarkar & Baidya, 2025). By default, a forest plot displays the prediction interval as a bar beneath the pooled estimate. If the prediction interval is not required, it can be omitted by setting the argument `prediction = FALSE` in the `forest()` function, which removes it from the graphical output (see previous in-text code). The degree of overlap among study confidence intervals provides a visual impression of heterogeneity, while the vertical reference line allows comparison of individual studies with the overall effect (Chang et al., 2022; Favorito, 2023; Sarkar & Baidya, 2025). Importantly, in meta-analyses of prevalence, forest plots should be interpreted with caution, as high heterogeneity and wide variability across studies are common and expected; therefore, visual overlap of confidence intervals alone may not reliably indicate consistency, and emphasis should be placed on the range and distribution of estimates rather than solely on the pooled summary effect (Bigna, 2026; Migliavaca et al., 2022).

Funnel plot

Funnel plots help identify potential publication bias by showing the relationship between study size and effect size (Spineli & Pandis, 2021). Symmetry suggests no bias, while asymmetry may indicate bias (Patole, 2021; Sterne et al., 2011). The command for creating funnel plots in R is detailed in the following code, and the output is shown in Figure 2.

```
# Generating a Funnel Plot for
# Assessing Publication Bias
pdf("funnel.prev.meta.glmm.pdf",
    width = 15, height = 10)
par(mar = c(7, 7, 2.1, 2.1),
    cex.lab = 2.5,
    cex.axis = 2.5)
funnel(
  prev.meta.glmm,
  xlab = "Logit Transformed Proportion",
  ylab = "Standard Error",
  cex  = 2.5,
  lwd  = 7 )
grid.text(
  "Funnel Plot",
  y  = unit(0.98, "npc"),
  gp = gpar(fontsize = 16,
            fontface = "bold") )
dev.off()
```

Figure 2 ■ Funnel Plot for Assessment of Publication Bias

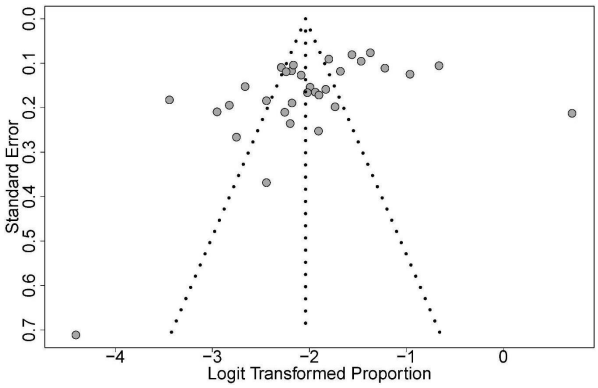
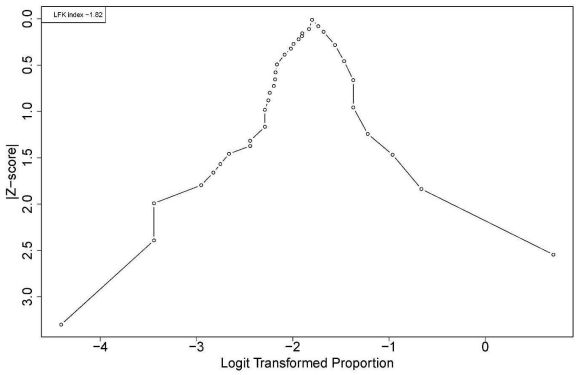


Figure 3 ■ Doi Plot for Assessing Publication Bias in the Prevalence of Disease X



Running the provided code generates a PDF file named `funnel.prev.meta.glmm.pdf` in the 'working directory' with dimensions of 15 inches in width and 10 inches in height. The `par()` function customizes the plot's appearance, with the `mar` argument defining the size of the margins (bottom, left, top, and right), where each unit corresponds to one line of text. The `cex.lab` and `cex.axis` parameters adjust the size of the axis labels and tick marks, respectively. The `funnel()` function creates the funnel plot based on the `prev.meta.glmm` model, with the X-axis labeled "Logit-Transformed Proportion" and the Y-axis labeled "Standard Error". The size of the symbols for each study is controlled by the `cex` argument, and the thickness of the plot lines is set by the `lwd` argument. After the plot is created, the `dev.off()` function closes the PDF device and saves the plot to the specified file. These parameters, such as symbol size and line thickness, can be easily adjusted to customize the appearance of the plot.

The funnel plot in meta-analysis of prevalence displays the relationship between the estimated proportion (usually on a transformed scale) and its precision (standard error; Afonso et al., 2024; Sterne et al., 2011). In the absence of publication bias or small-study effects, the plot typically appears as a symmetrical inverted funnel, with larger studies (more precise estimates) clustering near the top and smaller studies showing greater spread at the bottom (Afonso et al., 2024; Brush et al., 2024; Sterne et al., 2011). Deviations from symmetry may indicate potential publication bias, selective reporting, or true heterogeneity in prevalence across studies (Afonso et al., 2024; Sterne et al., 2011).

However, in prevalence meta-analyses, some asymmetry is common due to variability in study size, population characteristics, and underlying prevalence, and should therefore be interpreted cautiously in conjunction with statistical tests and other analyses (Afonso et al., 2024; Cheema et al., 2022; Sterne et al., 2011). Importantly, funnel plot asymmetry reflects small-study effects rather than publication bias alone, and statistical tests for asymmetry are often underpowered, especially when the number of studies is small; therefore, absence of asymmetry does not rule out bias, and its presence does not confirm it (Afonso et al., 2024; Cheema et al., 2022; Sterne et al., 2011).

Publication Bias Assessment

To statistically confirm the presence of publication bias, both Begg & Mazumdar's rank correlation test and Egger's linear regression test can be conducted (Begg & Mazumdar, 1994; Egger et al., 1997). The Begg's test is based on rank correlation between effect estimates and their variances, while Egger's test uses a linear regression approach to examine the association between effect sizes and their standard errors (Begg & Mazumdar, 1994; Egger et al., 1997). The `metabias()` function with `method.bias = "rank"` applies Begg's test, while `method.bias = "linreg"` performs Egger's test. The results, stored in `prev.meta.glmm.begs` and `prev.meta.glmm.egger`, respectively, are printed and provide test results and corresponding p-values (see Listing 3). For both tests, a p-value less than 0.05 is generally considered indicative of potential publication bias or small-study effects, whereas a non-significant result suggests no strong evidence of bias (Begg & Mazumdar, 1994; Egger et al., 1997). Egger's test is typically more sensitive but may also detect asymmetry due to heterogeneity rather than true publication bias, particularly in meta-analyses of prevalence (Furuya-Kanamori et al., 2018). Therefore, results from these tests should be interpreted cautiously and in conjunction with visual inspection of the funnel plot and other supporting analyses (Cheema et al., 2022).

Listing 3 ■ Performing Publication Bias Tests (Begg's and Egger's Tests)

```
> # Performing Publication Bias Tests
> # Begg and Mazumdar's rank correlation test
> prev.meta.glmm.begs <- metabias(prev.meta.glmm, method.bias = "rank")
> print(prev.meta.glmm.begs)

Rank correlation test of funnel plot asymmetry

Test result: z = -1.93, p-value = 0.0533
Bias estimate: -136.0000 (SE = 70.3729)

# > Egger's regression test
> prev.meta.glmm.egger <- metabias(prev.meta.glmm, method.bias = "linreg")
> print(prev.meta.glmm.egger)

Linear regression test of funnel plot asymmetry

Test result: t = -2.73, df = 33, p-value = 0.0102
Bias estimate: -5.5267 (SE = 2.0275)

Details:
- multiplicative residual heterogeneity variance (tau\^2 = 19.8776)
- predictor: standard error
- weight:    inverse variance
- reference: Egger et al. (1997), BMJ
```

Doi Plot and LFK Index

The Doi plot, along with the Luis Furuya-Kanamori (LFK) index, offers a sensitive approach to detecting publication bias by assessing plot asymmetry and quantifying its magnitude. These can be generated (see Figure 3) using the commands shown here:

```
# Generating a Doi Plot and
# Calculating the LFK Index
pdf("doi.prev.meta.glmm.pdf",
    width = 15,
    height = 10)
par(mar = c(7, 7, 2.1, 2.1),
    cex.lab = 2,
    cex.axis = 2)
doiplot(prev.meta.glmm, main = "Doi Plot
    for Prevalence of Disease X")
dev.off()

lfkindex.pre.meta.glmm <- lfkindex(
   prev.meta.glmm)
lfkindex.pre.meta.glmm
```

The Doi plot provides an alternative graphical method for assessing publication bias by plotting effect sizes against a transformed measure of precision, offering greater sensitivity than the conventional funnel plot. Symmetry in the Doi plot suggests the absence of publication bias, while asymmetry indicates potential small-study effects or selective reporting (Cheema et al., 2022; Furuya-Kanamori et al., 2018). The Luis Furuya-Kanamori (LFK) index quantifies this asymmetry, with values within ±1 indicating no asymmetry, between ±1 and ±2 suggesting minor asymmetry, and values beyond ±2 indicating major asymmetry (Cheema et al., 2022; Furuya-Kanamori et al., 2018). Thus, the Doi plot, together with the LFK index, provides both a visual and quantitative assessment of publication bias and should be interpreted alongside other methods in meta-analysis of prevalence.

Subgroup and Sensitivity Analyses

Subgroup analysis allows researchers to examine differences in proportions between groups (for example, community vs. hospital setting; Patole, 2021). Sensitivity analysis tests the robustness of results by excluding certain studies (e.g., low-quality studies). Both are detailed here:

```
# Performing Subgroup and
# Sensitivity Analyses
site.prev.meta.glmm <- update(
    prev.meta.glmm, subgroup = setting)
summary(site.prev.meta.glmm)
males.prev.meta.glmm <- update(
  prev.meta.glmm, subset = males == "y")
summary(males.prev.meta.glmm)
quality.prev.meta.glmm <- update(
  prev.meta.glmm, subset = quality >= 6)
summary(quality.prev.meta.glmm)
```

Subgroup analysis using a random-effects model to pool estimates separately for each category (e.g., hospital and community settings), along with their 95% confidence intervals and heterogeneity measures (τ^2, τ, Q, and I^2) can be conducted using the first command (see instructions on the bottom of the previous page) with the argument: `subgroup = setting` (Hirst et al., 2025; Sánchez-Meca & Botella, 2024). These results allow comparison of effect sizes across subgroups (see Listing 4) while accounting for within-group variability (Richardson et al., 2019; Sánchez-Meca & Botella, 2024). The I^2 statistic within each subgroup reflects the extent of residual heterogeneity, with higher values indicating substantial variability even after stratification (Richardson et al., 2019). The test for subgroup differences (between-group Q test) evaluates whether the observed differences in pooled estimates across subgroups are statistically significant; a p-value < 0.05 suggests a meaningful difference between subgroups, whereas a non-significant p-value indicates that any observed variation may be due to chance (Hirst et al., 2025; Richardson et al., 2019). Overall, subgroup analysis helps explore potential sources of heterogeneity and assess whether study-level characteristics explain differences in effect estimates (Hirst et al., 2025; Richardson et al., 2019). Importantly, subgroup analyses are observational in nature and therefore prone to confounding, ecological bias, and low statistical power, particularly when the number of studies within each subgroup is small; consequently, findings should be interpreted cautiously and ideally be pre-specified rather than data-driven (Hirst et al., 2025; Spineli & Pandis, 2020).

Sensitivity analyses assess the robustness of the findings by examining how the pooled estimate changes when specific studies are excluded or restricted. If results remain similar, the findings are considered robust; substantial changes indicate sensitivity to study selection and warrant cautious interpretation (Patole, 2021). In the line of code above, the second and third commands show how to conduct sensitivity analyses. For example, when the argument `subset = males == "y"` is used, the analysis is restricted to studies with male participants, and the summary output shows whether excluding non-male studies changes the overall conclusions. Although this is not a sensitivity analysis, the coding process is similar, which is why it is described here. The third command (with the argument `subset = quality >= 6`) excludes studies with a quality score below 6, providing a summary of how the results differ when only higher-quality studies are included. These analyses help assess the consistency and reliability of the meta-analysis findings across different groups and scenarios.

Meta-Regression

Meta-regression helps identify the relationship between study-level covariates and effect sizes, allowing the exploration of how factors such as age or setting influence prevalence (Patole, 2021):

```
# Performing Meta-Regression Analysis
agemr.prev.meta.glmm <- metareg(
    prev.meta.glmm, formula = age)
summary(agemr.prev.meta.glmm)
metareg.prev.meta.glmm <- metareg(
    prev.meta.glmm,
    formula = setting + age)
summary(metareg.prev.meta.glmm)
```

In this code, the first command, with the argument `formula = age`, runs a meta-regression with age as a single predictor, and the `summary(agemr.prev.meta.glmm)` provides details on how age influences prevalence across studies. This includes regression coefficients, standard errors, and p-values, which help assess whether age significantly affects the effect size. The second command, with the argument `formula = setting + age`, extends the analysis by including multiple predictor variables, i.e., setting and age. The output from the `summary(metareg.prev.meta.glmm)` provides insights into how these variables together influence prevalence, enabling evaluation of the combined effects of multiple factors (see Listing 5). Meta-regression helps uncover potential sources of heterogeneity and provides a deeper understanding of how study characteristics influence the meta-analysis results (Patole, 2021).

Meta-regression analysis examines the relationship between study-level covariates and the outcome while accounting for between-study heterogeneity (Sánchez-Meca & Botella, 2024). The mixed-effects model output provides estimates of residual heterogeneity (τ^2, τ, I^2, H^2), where a reduction in these values compared to the overall model suggests that included covariates explain part of the variability (Choi & Kang, 2025; Sánchez-Meca & Botella, 2024). The tests for residual heterogeneity (Wald and likelihood ratio tests) assess whether significant unexplained heterogeneity remains after including moderators (Cordero &

Listing 4 ■ Results of Subgroup Analyses

```
# Results of Subgroup Analyses
Results for subgroups (random effects model):
                         k  events            95%-CI  tau^2    tau       Q   I^2
setting = Hospital      27 12.5680 [9.7860; 16.0006] 0.5289 0.7272 545.80 95.2%
setting = Community      8  8.3106 [4.4008; 15.1437] 0.8933 0.9452 250.15 97.2%

Test for subgroup differences (random effects model):
                    Q d.f. p-value
Between groups 1.52    1  0.2180
```

Figure 4 ■ Baujat Plot for Identifying Influential Studies

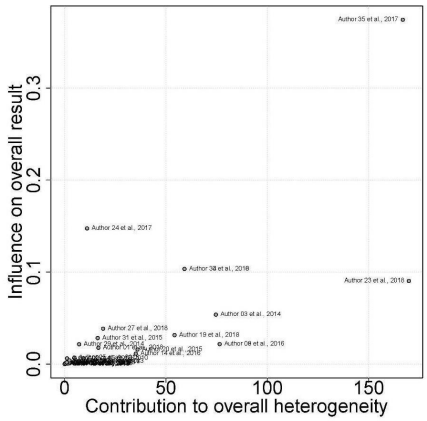
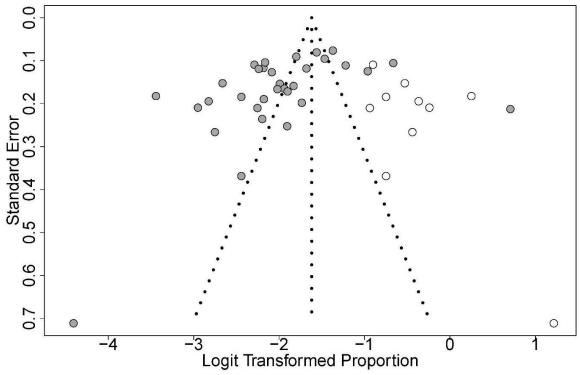


Figure 5 ■ Trim-and-Fill Funnel Plot with Adjusted Pooled Estimate



Dans, 2021; Higgins et al., 2003). The test of moderators (QM) evaluates whether the included covariates, individually or jointly, significantly explain variation in effect sizes; a p-value < 0.05 indicates that the moderator(s) have a statistically significant association with the outcome (Sánchez-Meca & Botella, 2024). The regression coefficients (estimates) indicate the direction and magnitude of association, with positive or negative values reflecting increasing or decreasing trends in the outcome per unit change in the covariate, and corresponding confidence intervals and p-values indicating statistical significance (Sánchez-Meca & Botella, 2024). In multiple meta-regressions, each coefficient represents the adjusted effect of that variable while controlling for others (Sánchez-Meca & Botella, 2024). Overall, meta-regression helps identify potential sources of heterogeneity and quantify the influence of study-level characteristics, although substantial residual heterogeneity may still persist even after adjustment (Sánchez-Meca & Botella, 2024; Spineli & Pandis, 2020).

Importantly, meta-regression is based on aggregated (study-level) data and is therefore susceptible to ecological bias (ecological fallacy), meaning that associations observed at the study level may not reflect true individual-level relationships; moreover, results are often underpowered and prone to spurious findings when the number of studies is small or multiple covariates are tested (Spineli & Pandis, 2020).

Additional Graphical Methods

To complement the conventional forest and funnel plots, several additional graphical methods can be employed to further explore heterogeneity, influence, and potential bias in meta-analysis. Among these, the Baujat plot, Trim-and-Fill funnel plot, and Galbraith (radial) plot are particularly useful yet less frequently reported in prevalence meta-analyses. These plots provide complementary perspectives by identifying influential studies, assessing the contribution of individual studies to between-study heterogeneity, and examining possible small-study effects or asymmetry.

Baujat plot

The Baujat plot (see code below) helps identify studies that contribute most to heterogeneity and have the greatest influence on the pooled estimate (Baujat et al., 2002). The x-axis represents each study's contribution to overall heterogeneity, while the y-axis reflects its influence on the sum-

mary effect size (see Figure 4). Studies appearing in the upper-right region of the plot are both highly influential and major contributors to heterogeneity, and may warrant further investigation through sensitivity analyses (Baujat et al., 2002).

```
# Generating a Baujat Plot for
# Identifying Influential Studies
pdf("Baujat.prev.meta.glmm.pdf", width =
     15, height = 10)
par(mar = c(7, 7, 2.1, 2.1), cex.lab = 2
    .5,
  cex.axis = 2.5)
baujat(prev.meta.glmm)
grid.text( "Baujat Plot",
  y  = unit(0.98, "npc"),
  gp = gpar(fontsize = 16,
            fontface = "bold"))
dev.off()
```

Trim-and-Fill funnel plot

The Trim-and-Fill funnel plot (see code below) is used to assess and adjust for potential publication bias in meta-analysis of prevalence (Duval & Tweedie, 2000). It augments the conventional funnel plot by imputing potentially missing studies to restore symmetry. Observed studies are displayed alongside imputed ones (see Figure 5), and any noticeable shift in the pooled estimate after adjustment suggests the possible influence of small-study effects or selective reporting (Duval & Tweedie, 2000).

```
# Generating a Trim-and-Fill Funnel
# Plot for Assessing Publication Bias
tnf.prev.meta.glmm <- trimfill(
    prev.meta.glmm)
tnf.prev.meta.glmm

pdf("tnf.prev.meta.glmm.pdf", width =
    15, height = 10)
par(mar = c(7, 7, 2.1, 2.1), cex.lab = 2
    .5, cex.axis = 2.5)
funnel(
  tnf.prev.meta.glmm,
  xlab = "Logit Transformed Proportion",
  ylab = "Standard Error",
  cex  = 2.5,
  lwd  = 7)
grid.text(
  "Trim-and-Fill Funnel Plot",
  y  = unit(0.98, "npc"),
  gp = gpar(fontsize = 16, fontface = "
    bold"))
dev.off()
```

Galbraith (radial) plot

The Galbraith (radial) plot (see code below) displays standardized effect sizes against the inverse of their standard errors, providing a clear visualization of heterogeneity and outliers. Studies that fall within the confidence bounds are generally consistent with the overall effect, whereas those lying outside these limits may be considered outliers or influential contributors to heterogeneity (see Figure 6). This plot is particularly useful for quickly identifying studies that deviate markedly from the overall trend (Bowden et al., 2018).

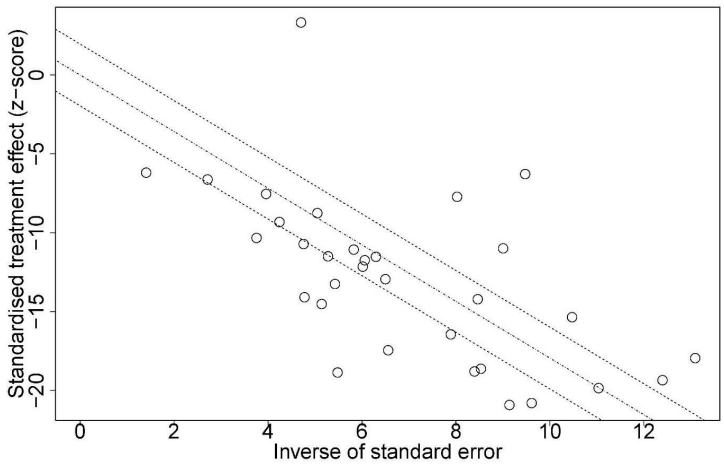
```
# Generating a Galbraith (Radial)
# Plot for Assessing Heterogeneity
pdf("radial.prev.meta.glmm.pdf", width =
     15, height = 10)
par(mar = c(7, 7, 2.1, 2.1), cex.lab = 2
    .5, cex.axis = 2.5)
radial(
  prev.meta.glmm,
  level = 0.95,
  cex   = 2.5,
  lwd   = 7)
grid.text(
  "Galbraith's Radial Plot",
  y  = unit(0.98, "npc"),
  gp = gpar(fontsize = 16, fontface = "
    bold"))
dev.off()
```

Discussion

This article provides step-by-step instructions, along with the included R scripts, making it easier for behavioral researchers, especially in low-resource settings, to estimate the pooled prevalence of a disease or a specific condition from smaller studies. By synthesizing data across studies, this method allows for more accurate prevalence estimates, helping inform health interventions and policies. The use of the GLMM is a significant strength, as it more robustly handles between-study variability compared to other common methods (Lin et al., 2022).

However, readers should be aware of the inherent assumptions and limitations in this article. This guide assumes that the readers are already familiar with the basic concepts and fundamentals of systematic review vis-à-vis meta-analysis. We also expect that our readers have a basic understanding of the R software environment. In this manuscript, we have chosen to limit the script to the 'meta' package of the R ecosystem (Balduzzi et al., 2019). This package is an easy-to-use package and is ideal for beginners for the direct acquisition of meta-analytical results (Balduzzi et al., 2019). However, multiple other packages

Figure 6 ■ Galbraith (Radial) Plot for Assessing Study-Level Contributions to Heterogeneity



in the R software environment can help users implement meta-analytical models. A review published in 2016 identified as many as 63 packages in R that are associated with meta-analysis (Polanin et al., 2016). For example, 'metafor' is another commonly used package that confers more complex analytical capabilities to the user (Viechtbauer, 2010). Interested readers may refer to the previously published articles comparing the 'meta' and 'metafor' packages and reference books (Harrer et al., 2021; Lortie & Filazzola, 2020). Readers should also note that the 'meta' package offers additional advanced features beyond the scope of this tutorial, such as performing meta-analyses of other types of effect outcomes. For more details, they can refer to previously published literature (Balduzzi et al., 2019).

Given the article's focus on practicality, it does not provide an in-depth discussion of the statistical theory underlying GLMMs or other techniques, so users may need to consult additional resources for a deeper understanding. In particular, although the benefits of using GLMM have been highlighted in this manuscript, it also has limitations that readers should be aware of, which will help them wisely choose to implement GLMM or its counterparts depending on the event counts and sample sizes of the included studies. A useful reference in this context is the comparison of GLMM with other transformation techniques published using datasets available in the Cochrane Library (Lin et al., 2022). Additionally, the focus of this guide was to demonstrate the technique of using GLMM for meta-analysis of prevalence studies. However, readers should be aware that GLMM is used in various types of meta-analyses, including advanced techniques like network meta-analysis and multilevel meta-analysis (Ades et al., 2013; Paul et al., 2010; Phillips et al., 2022; Riley et al., 2017). Despite these limitations, this guide remains a valuable resource, simplifying complex statistical methods for researchers with varying levels of expertise.

Conclusion

This manuscript provides a practical guide to conducting meta-analyses of single proportions in R. By leveraging the power of the 'meta' package (Balduzzi et al., 2019), this guide enables behavioral researchers to pool prevalence data from multiple studies, generate reliable prevalence estimates, and perform advanced statistical analyses such as meta-regression and subgroup analyses. However, we have not included a detailed interpretation of the output from these analyses, which is crucial for fully understanding the results. Readers may need to consult additional resources, as discussed in this manuscript, to effectively interpret and apply the findings. It is important to reiterate that this article is not intended as a detailed guideline for conducting meta-analyses of prevalence in R, but rather as a practical resource to help behavioral researchers perform the basic analyses required for meaningful evidence synthesis.

Authors' note

The R script (meta_analysis_script.R), the dataset (meta_analysis_data.csv), and the corresponding output files are provided as supplementary materials and are available online at https://github.com/ResearchCore/PrevalenceMetaR. Readers can enter their data using the variable names provided in the table or modify the script to accommodate custom headers. Output text and graphs

are also available as supplementary files.

References

Ades, A. E., Caldwell, D. M., Reken, S., Welton, N. J., Sutton, A. J., & Dias, S. (2013). Evidence synthesis for decision making 7: A reviewer's checklist. *Medical Decision Making*, *33*(5), 679–691. doi: 10.1177/0272989X13485156.

Afonso, J., Ramirez-Campillo, R., Clemente, F. M., Büttner, F. C., & Andrade, R. (2024). The perils of misinterpreting and misusing "publication bias" in meta-analyses: An education review on funnel plot-based methods. *Sports Medicine*, *54*(2), 257–269. doi: 10.1007/s40279-023-01927-9.

Asogwa, O. A., Boateng, D., Marza-Florensa, A., Peters, S., Levitt, N., van Olmen, J., & Klipstein-Grobusch, K. (2022). Multimorbidity of non-communicable diseases in low-income and middle-income countries: A systematic review and meta-analysis. *BMJ Open*, *12*(1), e049133. doi: 10.1136/bmjopen-2021-049133.

Balduzzi, S., Rucker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with r: A practical tutorial. *Evidence-Based Mental Health*, *22*(4), 153–160. doi: 10.1136/ebmental-2019-300117.

Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E., & Vos, T. (2013). Meta-analysis of prevalence. *Journal of Epidemiology and Community Health*, *67*(11), 974–978. doi: 10.1136/jech-2013-203104.

Barker, T. H., Migliavaca, C. B., Stein, C., Colpani, V., Falavigna, M., Aromataris, E., & Munn, Z. (2021). Conducting proportional meta-analysis in different types of systematic reviews: A guide for synthesisers of evidence. *BMC Medical Research Methodology*, *21*(1), 189. doi: 10.1186/s12874-021-01381-z.

Bashir, Y., & Conlon, K. C. (2018). Step by step guide to do a systematic review and meta-analysis for medical professionals. *Irish Journal of Medical Science*, *187*(2), 447–452. doi: 10.1007/s11845-017-1663-3.

Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, *21*(18), 2641–2652. doi: 10.1002/sim.1221.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. doi: 10.2307/2533446.

Bigna, J. J. (2026). Understanding heterogeneity in prevalence meta-analyses: From structural incompatibility to statistical variation. *Systematic Reviews*, *15*(1), 94. doi: 10.1186/s13643-026-03121-0.

Borenstein, M. (2023). How to understand and report heterogeneity in a meta-analysis: The difference between i-squared and prediction intervals. *Integrative Medicine Research*, *12*(4), 101014. doi: 10.1016/j.imr.2023.101014.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis* [John wiley & sons]. https://www.wiley.com/en-ie/Introduction+to+Meta-Analysis

Borges Migliavaca, C., Stein, C., Colpani, V., Barker, T. H., Munn, Z., Falavigna, M., & Prevalence Estimates Reviews - Systematic Review Methodology Group. (2020). How are systematic reviews of prevalence conducted? a methodological study. *BMC Medical Research Methodology*, *20*(1), 96. doi: 10.1186/s12874-020-00975-3.

Bowden, J., Spiller, W., Del Greco, M. F., Sheehan, N., Thompson, J., Minelli, C., & Davey Smith, G. (2018). Improving the visualization, interpretation and analysis of two-sample summary data mendelian randomization via the radial plot and radial regression. *International Journal of Epidemiology*, *47*(4), 1264–1278. doi: 10.1093/ije/dyy101.

Brush, P. L., Sherman, M., & Lambrechts, M. J. (2024). Interpreting meta-analyses: A guide to funnel and forest plots. *Clin Spine Surg*, *37*(1), 40–42. doi: 10.1097/bsd.0000000000001534.

Chang, Y., Phillips, M. R., Guymer, R. H., Thabane, L., Bhandari, M., Chaudhary, V., Wykoff, C. C., Sivaprasad, S., Kaiser, P., Sarraf, D., Bakri, S., Garg, S. J., Singh, R. P., Holz, F. G., Wong, T. Y., & R. E. T. I. N. A. Study Group. (2022). The 5 min meta-analysis: Understanding how to read and interpret a forest plot. *Eye*, *36*(4), 673–675. doi: 10.1038/s41433-021-01867-6.

Cheema, H. A., Shahid, A., Ehsan, M., & Ayyan, M. (2022). The misuse of funnel plots in meta-analyses of proportions: Are they really useful? *Clinical Kidney Journal*, *15*(6), 1209–1210. doi: 10.1093/ckj/sfac035.

Choi, G. J., & Kang, H. (2025). Heterogeneity in meta-analyses: An unavoidable challenge worth exploring. *Korean J Anesthesiol*, *78*(4), 301–314. doi: 10.4097/kja.25001.

Cordero, C. P., & Dans, A. L. (2021). Key concepts in clinical epidemiology: Detecting and dealing with heterogeneity in meta-analyses. *Journal of Clinical Epidemiology*, *130*, 149–151. doi: 10.1016/j.jclinepi.2020.09.045.

Daly, C., & Soobiah, C. (2022). Software to conduct a meta-analysis and network meta-analysis. In E. Evangelou & A. A. Veroniki (Eds.), *Meta-research: Methods and protocols* (pp. 223–244). Springer US. doi: 10.1007/978-1-0716-1566-9_14.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for

publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. doi: 10.1111/j.0006-341X.2000.00455.x.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. doi: 10.1136/bmj.315.7109.629.

Evans, A. S. (2009). Epidemiological concepts. In P. S. Brachman & E. Abrutyn (Eds.), *Bacterial infections of humans* (pp. 1–50). Springer US. doi: 10.1007/978-0-387-09843-2_1.

Favorito, L. A. (2023). Systematic review and metanalysis in urology: How to interpret the forest plot. *Int Braz J Urol*, *49*(6), 775–778. doi: 10.1590/s1677-5538.Ibju.2023.9911.

Furuya-Kanamori, L., Barendregt, J. J., & Doi, S. A. R. (2018). A new improved graphical and quantitative method for detecting bias in meta-analysis. *JBI Evidence Implementation*, *16*(4), 1–19. doi: 10.1097/XEB.0000000000000141.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R*. Chapman & Hall. doi: 10.1201/9781003107347.

Higgins, J. P. T., Davey Smith, G., Altman, D. G., & Egger, M. (2022). Principles of systematic reviewing. In M. Egger, J. P. Higgins, & G. D. Smith (Eds.), *Systematic reviews in health research* (pp. 17–35). Wiley. doi: 10.1002/9781119099369.ch2.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557. doi: 10.1136/bmj.327.7414.557.

Hirst, L., Ntaga, I., Seehra, J., Mavridis, D., & Pandis, N. (2025). Reporting and interpretation of subgroup analyses in orthodontic meta-analysis; a meta-epidemiological study. *European Journal of Orthodontics*, *47*(4), 1–19. doi: 10.1093/ejo/cjaf053.

Lin, L., & Chu, H. (2020). Meta-analysis of proportions using generalized linear mixed models. *Epidemiology*, *31*(5), 713–717. doi: 10.1097/EDE.0000000000001232.

Lin, L., Xu, C., & Chu, H. (2022). Empirical comparisons of 12 meta-analysis methods for synthesizing proportions of binary outcomes. *Journal of General Internal Medicine*, *37*(2), 308–317. doi: 10.1007/s11606-021-07098-5.

Lortie, C. J., & Filazzola, A. (2020). A contrast of meta and metafor packages for meta-analyses in r. *Ecology and Evolution*, *10*(20), 10916–10921. doi: 10.1002/ece3.6747.

Migliavaca, C. B., Stein, C., Colpani, V., Barker, T. H., Ziegelmann, P. K., Munn, Z., Falavigna, M., & PERSyst Methodology Group. (2022). Meta-analysis of prevalence: I2 statistic and how to deal with heterogeneity. *Research Synthesis Methods*, *13*(3), 363–367. doi: 10.1002/jrsm.1547.

Noordzij, M., Dekker, F. W., Zoccali, C., & Jager, K. J. (2010). Measures of disease frequency: Prevalence and incidence. *Nephron Clinical Practice*, *115*(1), c17–20. doi: 10.1159/000286345.

Ntais, C., & Talias, M. A. (2024). Unveiling the value of meta-analysis in disease prevention and control: A comprehensive review. *Medicina*, *60*(10), 1–19. doi: 10.3390/medicina60101629.

Patole, S. (2021). *Principles and practice of systematic reviews and meta-analysis*. Springer. doi: 10.1007/978-3-030-71921-0.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., & Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested laplace approximations. *Statistics in Medicine*, *29*(12), 1325–1339. doi: 10.1002/sim.3858.

Phillips, M. R., Steel, D. H., Wykoff, C. C., Busse, J. W., Bannuru, R. R., Thabane, L., Bhandari, M., Chaudhary, V., & Retina Evidence Trials International Alliance Study Group. (2022). A clinician's guide to network meta-analysis. *Eye*, *36*(8), 1523–1526. doi: 10.1038/s41433-022-01943-5.

Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2016). A review of meta-analysis packages in r. *Journal of Educational and Behavioral Statistics*, *42*(2), 206–242. doi: 10.3102/1076998616674315.

Posit Team. (2025). *Rstudio* [Posit Software Pbc Integrated Development Environment for R]. http://www.posit.co/

R Core Team. (2025). R: A language and environment for statistical computing. In *R foundation for statistical computing*. https://www.R-project.org/

Richardson, M., Garner, P., & Donegan, S. (2019). Interpretation of subgroup analyses in systematic reviews: A tutorial. *Clinical Epidemiology and Global Health*, *7*(2), 192–198. doi: 10.1016/j.cegh.2018.05.005.

Riley, R. D., Jackson, D., Salanti, G., Burke, D. L., Price, M., Kirkham, J., & White, I. R. (2017). Multivariate and network meta-analysis of multiple outcomes and multiple treatments: Rationale, concepts, and examples. *BMJ*, *358*, j3932. doi: 10.1136/bmj.j3932.

Sánchez-Meca, J., & Botella, J. (2024). Moderators analysis in meta-analysis: Meta-regression and subgroups analyzes [10.1016/j.cireng. 2024.03.012]. *Cirugía Española (English Edition)*, *102*(8), 446–447. doi: 10.1016/j.cireng.2024.03.012.

Sarkar, S., & Baidya, D. K. (2025). Meta-analysis - interpretation of forest plots: A wood for the trees. *Indian J Anaesth*, *69*(1), 147–152. doi: 10.4103/ija.ija_1155_24.

SAS Institute Inc. (2025). *SAS/STAT® 15.4 user's guide*.

Schwarzer, G. (2006). Meta: An r package for meta-analysis. *R News*, *7*(3), 40–45. doi: 10.32614/CRAN.package.meta.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2025). *Metasens: Statistical methods for sensitivity analysis in meta-analysis* [Comprehensive R Archive Network (CRAN)]. https://CRAN.R-project.org/package=metasens

Schwarzer, G., Chemaitelly, H., Abu-Raddad, L. J., & Rucker, G. (2019). Seriously misleading results using inverse of freeman-tukey double arcsine transformation in meta-analysis of single proportions. *Research Synthesis Methods*, *10*(3), 476–483. doi: 10.1002/jrsm.1348.

Schwarzer, G., & Rucker, G. (2022). Meta-analysis of proportions. *Methods in Molecular Biology*, *2345*, 159–172. doi: 10.1007/978-1-0716-1566-9_10.

Shimizu, I., & Ferreira, J. C. (2023). Losing your fear of using r for statistical analysis. *Jornal Brasileiro de Pneumologia*, *49*(3), e20230212. doi: 10.36416/1806-3756/e20230212.

Spineli, L. M., & Pandis, N. (2020). Problems and pitfalls in subgroup analysis and meta-regression. *American Journal of Orthodontics and Dentofacial Orthopedics*, *158*(6), 901–904. doi: 10.1016/j.ajodo.2020.09.001.

Spineli, L. M., & Pandis, N. (2021). Publication bias: Graphical and statistical methods. *American Journal of Orthodontics and Dentofacial Orthopedics*, *159*(2), 248–251. doi: 10.1016/j.ajodo.2020.11.005.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. doi: 10.1136/bmj.d4002.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. doi: 10.18637/jss.v036.i03.

Wang, N. (2023). Conducting meta-analyses of proportions in r. *Journal of Behavioral Data Science*, *3*(2), 64–126. doi: 10.35566/jbds/v3n2/wang.

Open practices

The *Open Data* badge was earned because the data of the experiment(s) are available on github.com/ResearchCore/PrevalenceMetaR
The *Open Material* badge was earned because supplementary material(s) are available on github.com/ResearchCore/PrevalenceMetaR

Citation

Yadav, V., Trushna, T., Goel, A. D., Mandal, U. K., & Sabde, Y. D. (2026). Conducting meta-analysis of single proportions using R: A practical tutorial for behavioral researchers. *The Quantitative Methods for Psychology*, *22*(2), 102–118. doi: 10.20982/tqmp.22.2.p102.

Received: 10/02/2026 ~ Accepted: 22/06/2026

Listing 5 follows

Listing 5 ■ Results of Meta-Regression Analysis

```
> summary(agemr.prev.meta.glmm)

Mixed-Effects Model (k = 35; tau\^2 estimator: ML)

   logLik   deviance        AIC        BIC       AICc
-101.9592     8.9190   209.9184   214.5845   210.6926

tau\^2 (estimated amount of residual heterogeneity):     0.1547
tau (square root of estimated tau\^2 value):             0.3934
I\^2 (residual heterogeneity / unaccounted variability): 89.67%
H\^2 (unaccounted variability / sampling variability):   9.68

Tests for Residual Heterogeneity:
Wld(df = 33) = 209.6850, p-val < .0001
LRT(df = 33) = 225.3949, p-val < .0001

Test of Moderators (coefficient 2):
QM(df = 1) = 81.6915, p-val < .0001

Model Results:

         estimate      se      zval    pval    ci.lb    ci.ub
intrcpt   -4.4281  0.2784  -15.9042  <.0001  -4.9738  -3.8824  ***
age        0.0607  0.0067    9.0383  <.0001   0.0475   0.0738  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> \# Meta-regression with multiple covariates (e.g., setting and age)
> metareg.prev.meta.glmm <- metareg(prev.meta.glmm, formula = setting + age)
> summary(metareg.prev.meta.glmm)

Mixed-Effects Model (k = 35; tau\^2 estimator: ML)

   logLik   deviance        AIC        BIC       AICc
-101.7935     8.5875   211.5870   217.8084   212.9203

tau\^2 (estimated amount of residual heterogeneity):     0.1542
tau (square root of estimated tau\^2 value):             0.3927
I\^2 (residual heterogeneity / unaccounted variability): 89.56%
H\^2 (unaccounted variability / sampling variability):   9.58

Tests for Residual Heterogeneity:
Wld(df = 32) = 209.0538, p-val < .0001
LRT(df = 32) = 225.3664, p-val < .0001

Test of Moderators (coefficients 2:3):
QM(df = 2) = 82.2695, p-val < .0001

Model Results:

                estimate      se      zval    pval    ci.lb    ci.ub
intrcpt          -4.4950  0.2957  -15.2030  <.0001  -5.0745  -3.9155  ***
settingHospital   0.1207  0.1784    0.6766  0.4987  -0.2289   0.4703
age               0.0600  0.0068    8.8508  <.0001   0.0467   0.0733  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```