Methods for meta-analysis and meta-regression of proportions: concepts and tutorial with Stata command metapreg

Victoria Nyawira Nyaga (victoria.nyawiranyaga@sciensano.be)
Sciensano (Belgium)

Marc Arbyn
Sciensano (Belgium)

Method Article

Keywords: Meta-analysis, Meta-regression, network meta-analysis, Stata, logistic regression, Binomial, Metapreg

Posted Date: October 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2142518/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Background

Meta-analyses offer an efficient way to synthesize information facilitating evidence-based decision-making. This is achieved through a statistical procedure that synthesizes results from separate studies addressing the same research question and with a similar design. Unless sound statistical procedures are used, inference from a poorly conducted meta-analysis can lead to erroneous conclusions.

The accelerated technological development of software to perform complex and computer-intensive statistical procedures have extended the capabilities for pooling data. While this has expanded the scope of statistical problems that can be solved, the use of optimal methods for meta-analysis and meta-regression of proportions are yet to become routine in statistical practice.
In 2014, we published two procedures in Stata for meta-analysis of proportions. The first one metaprop[1], implemented classical and simple methods including the Freeman-Tukey double arcsine transformation[2]. The second was metaprop_one[3] which implemented the logistic regression and logistic-normal regression. Metaprop has over 2000 downloads and over 1300 citations by May 2022 (Google Scholar) since its publication. The methods in metaprop are based on approximation to the normal distribution. Both procedures perform pooling of a single set of proportions and allow subgroup meta-analysis by one categorical covariate but do not have capabilities for meta-regression.

We built further on our work and developed metapreg: a user-friendly and flexible tool based on the well-established logistic regression to perform meta-analysis and meta-regression of proportions in Stata. The tool has optimized and advanced statistical procedures including marginal standardization[4, 5, 6] to estimate the adjusted proportions and probability ratios from the fitted logistic regression model and arm-based network meta-analysis.

The framework is more powerful than metaprop since it allows the inclusion of continuous and categorical explanatory variables as well as simple interactions. In contrast to subgroup analyses, a unified modeling approach improves the estimation process by sharing information between studies yielding more precise estimates. It is a more efficient use of data and decreases the chance of finding spurious significant effects.

This article has four focus points; 1. Explaining the key concepts and rationale in methods for meta-analysis, 2. Discussion of the theoretical and methodological flaws in the traditional simplistic methods for meta-analysis of proportions, 3. Description of the framework of metapreg, and 4. illustrating its use on data derived from three published meta-analyses.

**Methodology**

The concept of linear regression

From a statistical point of view, a model is a mathematical expression formulated to decently describe the behavior of a random variable.

The simplest model splits the observed proportions into two additive components; the overall mean and a study-specific sampling error. When covariate information is available, a regression model expresses the statistical relation between the outcome response \( Y \) and the covariates \( X \). Consider \( I \) studies and let \( X_0 \) be a constant equal to 1, then a simple linear regression model without covariates is as follows:

\[
Y_i = X_0 \beta_0 + \epsilon_i \equiv \beta_0 + \epsilon_i \text{ for } i=1, \ldots, I
\]

(1)

where \( Y_i \) is the outcome in the \( i^{th} \) study, \( \beta_0 \) a parameter denoting the mean, and \( \epsilon_i \) a random error term. The error terms are identical, independent, centrally located around zero i.e. \( E(\epsilon_i) = 0 \), and assumed to have constant variance i.e. \( \sigma^2(\epsilon_i) = \sigma^2 \). These statements imply that:

\[
E(Y_i) = \beta_0
\]

\[
\sigma^2(Y_i) = \sigma^2 \text{ i.e variance of } Y_i \text{ in each study is } \sigma^2
\]

\[
\sigma(Y_i, Y_i') = 0 \text{ for any } i \neq i'.
\]

(2)
Estimation of parameters $\beta_0$ and $\sigma^2$ in the regression model requires no specification for the distribution for the error terms $\epsilon_i$. However, to make inferences i.e. compute confidence and prediction intervals and perform hypothesis tests, it is necessary to assume a distribution of $\epsilon_i$. A normal distribution i.e. $\epsilon_i \sim N(0, \sigma^2)$ is conventionally the standard assumption. This translates to; $Y_i$ is a normally distributed random variable centered around $\beta_0$ and with a spread equal to $\sigma^2$ i.e.

$$Y_i \sim N(\beta_0, \sigma^2)$$

The normality assumption greatly simplifies the theory of regression analysis. Another justification for the normal distribution is that even if $Y_i$ is not exactly normal, its distribution approaches normality with sufficiently large sample sizes. As a consequence of this assumption, most statistical software packages have been developed with standard capabilities to obtain parameter estimates for the normal regression model.

**Rationale of meta-analysis**

A meta-analysis is a model that estimates the mean and variance of the distribution of effect sizes (in our case, proportions) from multiple studies assessing the same condition and similar patient characteristics. The analysis also attempts to explain the present between-study heterogeneity, if any. Usually, the studies do not have the same reliability i.e. they have different variances dependent on the estimated proportion and the sample size. What this means is that the assumption of constant variance no longer holds i.e. $\sigma^2(\epsilon_i) \neq \sigma^2$. It is therefore necessary to modify distribution of $Y_i$ in 3 to allow for the variable variances. The modified distribution is expressed as follows:

$$Y_i \sim N(\beta_0, \sigma^2_i)$$

where $\sigma^2_i$ denotes the study-specific variance, the so-called within-study variance. The modified model then uses a procedure that utilizes weights to obtain the parameter estimates, the so-called weighted analysis. The weights $w_i$ assigned to each study are inverse to the within-study variance $\sigma^2_i$.

The within-study variance $\sigma^2_i$ in the distribution of $Y_i$ in 4 is a random variable though it is common to treat it as known, which is inappropriate as discussed by Jackson and White.[7] Ideally, a variance function should be estimated by regressing the squared residuals or the sample variances against an “appropriate” predictor variable. The fitted values from the variance function are then used to obtain $\sigma^2_i$.[8] Unfortunately, information on the “appropriate” predictor variable is never available.

Conventionally, the study sample variances are directly used so that the distribution of $Y_i$ in 4 now becomes:

$$Y_i \sim N(\beta_0, s^2_i)$$

such that precise and/or larger studies with smaller variances (more reliable information) get more weight. When the number of studies in the meta-analysis is large
enough, the direct use of sample variances to estimate the unknown within-study variances is justified. This is because the weights become essentially irrelevant.

Alternative weighting schemes use a function of the study size only. The arguments for not using the within-study variance are; 1) to avoid giving large weights to small but precise studies especially when there are few studies, 2) to avoid the estimation error in the within-study variance,[9] and 3) assign uniform weight regardless of the metric.[10]

**Random- vs fixed-effects model**

The distribution specified in 4 dictates that all studies have an identical (fixed) mean (or a mean that only varies as a function of known study characteristics i.e. covariates). Furthermore, it dictates that all the observed variation is purely due to within-study sampling error. This is referred to as the fixed-effects model.

In reality, it is not possible to attribute all variation to the sampling error and the known study characteristics. To capture the extra (unexplained) variation between the studies, a random component $\delta_i$ specific to each study is added to the model. The component captures the study-specific deviation from the overall mean i.e.

$$Y_i \sim N(\beta_0 + \delta_i, \sigma^2_i) \quad (6)$$

Conventionally, the study-specific random components come from a normal distribution centered around zero and whose spread is the between-study variance ($\tau^2$) i.e. $\delta_i \sim N(0, \tau^2)$. This specification dictates that the study-specific mean is random i.e. $\beta_i = \beta_0 + \delta_i$ (or a random function of the known study characteristics and the random study-specific component $\delta_i$). This extended model is commonly referred to as the random-effects model.

Because the two random components are uncorrelated, it becomes automatic that $\sigma^2(Y_i) = \sigma^2_i + \tau^2$. Consequently, the new weights $w_i$ assigned to each study are inverse to ($\sigma^2_i + \tau^2$). Analogous to the fixed-effects model, the weighting function uses $s^2_i + \tau^2$; the sum of the within-study sample variance and the estimated between-study variance. There are different methods to obtain an estimate of $\tau^2$ including the method of moments (MOM), maximum likelihood, restricted maximum likelihood, profile likelihood, Bayes model, and the permutations model.[11]

**Binomial random variables**

We refer to a study $i$ with a fixed number of binary responses generically labeled “success” (alive/healthy/cured) and “failure” (dead/sick/not cured). Let $n_i$ the number of “successes” and $N_i$ the sum of ‘successes’ and ‘failures’. It is natural to assume that $n_i$ is a binomially distributed random variable with parameters $N_i$ and $p_i$; the probability of “success”. The distribution is denoted by $n_i \sim bin(p_i, N_i)$. An estimate of $p_i$ is the proportion that “succeeded” in study $i$. A meta-analysis of binomial data using the distribution of $Y_i$ in 5 and/or 6 typically proceeds as follows:

1. Obtain an estimate of $\hat{p}_i = \frac{n_i}{N_i}$.
2. Obtain the within-study variance $s^2_i = \frac{\hat{p}_i(1-\hat{p}_i)}{N_i}$.
3. Via weighted analysis:
(a) Obtain “an initial” \( \hat{\beta}_0 \) assuming distribution 5 i.e \( \hat{p}_i \sim N(\beta_0, s_i^2) \).

(b) Obtain an estimate of \( \tau^2 \).

(c) Obtain “the final” \( \hat{\beta}_0 \) assuming distribution 6; \( \hat{p}_i \sim (\beta_0 + \sigma_i, s_i^2), \sigma_i \sim N(0, \tau^2) \).

There are two main violations in the procedure above. First, we have used the derived statistics \( \hat{p}_i \) from the binomial distribution and treated them as known while in reality, they are latent/unobserved.

Secondly, the symmetricity of the normal distribution allocates equal probability to each tail and this is reasonable whenever the proportions are all around 0.5. However, when the underlying distribution is skewed, the normality assumption violates the natural boundaries \([0, 1]\) of the proportions.

These violations manifest in several ways. 1) When \( n_i = 0 \) or \( n_i = N_i \), \( s_i^2 = 0 \) implying \( w_i = \infty \) leading to the exclusion of the study from the analysis. 2). When \( \beta_0 \) is near zero or one, it is possible to have confidence intervals outside the admissible interval \([0, 1]\). 3). When \( K \) is small; which is common in meta-analyses and/or \( \hat{p}_i \) near 0/1, the distribution of \( \hat{p}_i \) is likely to be skewed and discrete. 4). The within-study variance \( s_i^2 = \frac{\hat{p}_i \cdot (1 - \hat{p}_i)}{N_i} \) is a function of mean \( \hat{p}_i \). As \( \hat{p}_i \) moves towards 0 or 1, \( s_i^2 \) moves towards 0. This means that when the mean changes, the variance changes, and therefore ignoring this correlation by treating the variance \( s_i^2 \) as independent, known, and constant is inappropriate and may bias the summary estimates and their variances.[12]

Remedial measures

These problems occur because methods for normal data are applied to non-normal data. Several actions are taken in an attempt to make the proportions close to normal. Transformations such as the logit, arcsine, and the Freeman-Tukey double arcsine transformation[2] e.g in metaprop[1] have been the obvious recourse for the longest time simply due to their mathematical simplicity.

New tests and treatments are being developed because they offer advantages such as easier administration, lower cost, or better safety profile while maintaining accuracy and efficacy similar to those of standard treatments and tests. In this case, the focus of the meta-analysis is mainly to evaluate the ‘success’ probability ratios of the index vs. the comparator test(s)/treatment(s).

Different studies can evaluate different tests/treatments on similar patient groups in the same study. In this case, the type of test/treatment is a covariate. Another common scenario is when two or more tests/treatments are evaluated on the same patient groups in a study. Rather than make full use of the outcome information, a simplistic procedure performs subgroup analysis to compare the different proportions.

Alternative procedures use the log-transformed estimated probability ratios (relative risks) or the odds ratios and their standard errors e.g in metan,[13] and mvmeta.[14] When there are at least two covariates, a subgroup analysis is further performed to compare the different ratios.

Subgroup analyses ignore the covariance among variables which can lead to spurious significant effects, confounded effects,[15] and invalid standard errors.

However easy and popular these remedies may be, transformations often create inappropriate regression relationships and obscure the true nature of the data.
Schwarzer et al.[16] showed that using the inverse of the Freeman-Tukey double arcsine transformation yielded seriously misleading results when the study sizes are skewed. Hamza et al.[17] already showed a downward bias in the parameter estimates after the logit transformation. These shortcomings are discussed in detail elsewhere.[18, 7]

Given these limitations, it is best to completely abandon the procedures based on the distribution of $Y_i$ in 5 or 6 and use/develop a better modeling approach that is more appropriate for the natural distribution of the proportions.

Hierarchical modeling

The proportions in a meta-analysis are themselves not observed but parameters of an observed binomial ‘experiment’. As such, it is natural to model the problem hierarchically. In the first level of the hierarchy, given the probability of success, a binomial distribution describes the variation in the number of ‘successes’.

In presence of between-study variation, there is more variability in the data than would be predicted by the fixed-effect model. Thus in the second level, the probabilities (the unobserved proportions) are modeled using a given distribution expressed in terms of other parameters.

The unobserved proportions are bound between 0 and 1. The beta distribution is ideal because it describes the distribution of a continuous variable in the interval $[0, 1]$. Furthermore, the beta distribution is naturally conjugate to the binomial distribution which greatly simplifies the computations in estimating the model parameter estimates and their interpretation. However, fitting the beta distribution outside the Bayesian setting is computationally complex.

The alternative is to employ the logit or the log transformation on the latent parameters and approximate their distribution with the normal distribution. Unlike the beta distribution, the normal distribution is non-conjugate to the binomial distribution. This approach does not pose any conceptual problems except that it makes the interpretation of the model parameters less transparent and requires more complex algorithms to estimate the model parameters.

Using the log transformation allows direct estimation of the probability ratios by simple exponentiation of the regression coefficients. On the downside, the log transformation permits obtaining predicted probabilities greater than 1 while proportions must fall within the bounds $[0,1]$. Computationally, the log-binomial model often also has convergence problems.[19, 4]

The logit transformation is better than the log transformation for several reasons. First, the odds ratio is a constant independent of the baseline probabilities. In contrast, the probability ratio depends on the baseline probability. Secondly, the odds have an unlimited range of probabilities such that any positive odds ratio still yield a valid probability.

Nonetheless, the probability ratios are a more intuitive and relevant measure than the odds ratio. While this may be the biggest motivation for the use of the log transformation, the focus of meta-analysis is on the average across studies and rarely on the direct model estimates which are study specific. Thus, whatever the choice of the transformation, the integration of the random effects is necessary to yield the more useful marginal estimates.
The logistic regression model

The logistic regression model is a part of the generalized linear models (GLM), an extension of the ordinary regression model to linearly model the transformed mean of a binomial variable. It is a linear model for the logit transformation of the binomial parameter that leaves the data as it is. The calculation of the population-averaged (adjusted) proportions and probability ratios from the fitted logistic regression model is straightforward through marginal standardization.[4, 5, 6]

In the context of metapreg, we distinguish between four inferential analyses based on the type of studies from which the data for the meta-analysis is generated. The first difference between the analyses is in the expression for the transformed mean of the binomial distribution. The second difference is in the presentation of the available dataset for the analysis. We present the random-effects models but whenever there are fewer than three studies in an analysis, the metapreg command fits a fixed-effects model by leaving out the random components from the predictor equation.

1 Independent studies

Each study contributes one set \((n_i, N_i, x_i)\) of data where \(x_i = x_{i0}, \ldots, x_{ip}\) denotes the values of \(p + 1\) covariates with \(x_{i0} = 1\). The logistic regression model with random study effects is:

\[
\begin{align*}
  n_i &\sim \text{bin}(p_i(x_i), N_i) \\
p_i(x_i) &= \frac{\exp(x_i\beta + \delta_i)}{1 + \exp(x_i\beta + \delta_i)} \\
\end{align*}
\]

where \(\delta_i \sim N(0, \tau^2)\) (7)

The variance of the \(\delta_i\) represents the unexplained variability of the proportions on the logit scale. \(\beta = (\beta_1, \ldots, \beta_p)\) denote the change in log odds of “success” for every 1-unit increase in \(x\). The \(\beta_0\) denotes the log odds of “success” when all covariates are set to 0. The absence of \(\delta_i\) in model 7 yields the fixed-effects model.

The \(p_i(x_i)\) can be extended to include any sensible interactions between the covariates but to remain simple, we implemented the metapreg program to only include the interaction terms between the first covariate and the rest of the covariates in the model.

We move on to methods for dependent studies where a study contributes more than one set of data. All sets of data generated from a certain study share similar study characteristics which can induce a positive correlation between the observations. The shared study-specific random effect in each set of data naturally captures this correlation. Taking into account this correlation can improve the precision of inferences for within-study estimates. Fitting a fixed-effects model in this setting ignores the dependence structure in the data and this may be inappropriate.

2 Comparative studies

Comparative meta-analysis where the ratio of two binomial proportions is the parameter of interest is rather common. For instance, the case-control studies where the proportion of ‘successes’ in the diseased or case group is compared
to the non-diseased or control group. Traditional meta-analyses use the log odds ratio or the log rate ratio as the outcome.

Let \((n_{i1}, N_{i1}), (n_{i2}, N_{i2}, x_{i})\) denote a pair of observations in a study \(i\). Let \(z_j\) be an indicator variable to identify the pairs. We identify the first set of data in the pair with \(z_1 = 0\) and the second set with \(z_2 = 0\). The model is expressed as follows:

\[
p_{ij}(z_j, x_i) = \frac{\exp(\gamma z_j + x_i \beta + \delta_i)}{1 + \exp(\gamma z_j + x_i \beta + \delta_i)}
\]

\(\delta_i \sim N(0, \tau^2)\) (8)

In a study \(i\), the odds of 'success' in the second pair are \(\exp(\gamma)\) times the odds in the first pair, i.e.

\[
p_{i1}(x_i) = \frac{\exp(x_i \beta + \delta_i)}{1 + \exp(x_i \beta + \delta_i)}
\]

\[
p_{i2}(x_i) = \frac{\exp(\gamma + x_i \beta + \delta_i)}{1 + \exp(\gamma + x_i \beta + \delta_i)}
\]

(9)

When \(\gamma = 0\) the 'success' probabilities are the same in the pairs. Here, the study-specific probability ratio is \(\frac{n_{i1} \cdot N_{i2}}{n_{i2} \cdot N_{i1}}\) and the asymptotic score confidence intervals for independent samples are obtained via the Koopman [20] method which relies on an iterative likelihood optimization.

3 Matched vs paired studies

Let \((a_i, b_i, c_i, d_i)\) denote the matched data from study \(i\). We refer to the representation in table 1. Suppose in study \(i\) there are \(Q_i\) 'case' groups compared to the 'control' group. For each study, there will be \(Q_i\) such tabulations as in table 1. From the four cells, we obtain the following two marginal distributions:

\[n_{i1} = (a_i + b_i),\]

\[n_{i2} = (a_i + c_i),\]

\[N_i = (a_i + b_i + c_i + d_i),\]

\[n_{i1} \sim bin(p_{i1}, N_i)\] and

\[n_{i2} \sim bin(p_{i2}, N_i)\]

(10)

When a study has data on the four cells, we refer to it as a "matched" study. In many studies, only the marginal data \((n_{i1}, n_{i2}, N_i)\) is available. We refer to such data as "paired". The terms matched and paired are often used synonymously, but in the context of metapreg, we use them to differentiate the two data setups.

Let \(z_j\) be an indicator variable to distinguish the case group from the control group. We assign \(z_1 = 1\) in the case group and \(z_2 = 0\) in the control group. With \(Q > 1\) case groups, let \(\phi = (\phi_2, \ldots, \phi_Q)\) be the case group effect with
$t_q = 1$ for $q = (2, \ldots, Q)$, and $t_q = 0$ if $q = 1$. The model is expressed as follows:

$$
\pi_{ijq}(z, t, x_i) = \exp(\gamma z_j + \phi q t_q + x_i \beta' + \delta_i) \frac{1}{1 + \exp(\gamma z_j + \phi q t_q + x_i \beta' + \delta_i)}
$$

\delta_i \sim N(0, \tau^2) \quad (11)

In the $i^{th}$ study, the odds of success with the $q^{th}$ case group are $\exp(\phi q)$ times the odds with the $1^{st}$ case group, i.e $\pi_{i1q}(x_i) = \frac{\exp(\gamma + x_i \beta' + \delta_i)}{1 + \exp(\gamma + x_i \beta' + \delta_i)}$ and $\pi_{i11}(x_i) = \frac{\exp(\gamma + x_i \beta' + \delta_i)}{1 + \exp(\gamma + x_i \beta' + \delta_i)}$. If $\gamma_j$ is zero, the log odds of success in the case and control group are similar, and if all $\phi_q$ are zero, then there are no differences between the different case groups.

When data are 'matched' the study-specific probability ratio is $\frac{n_i + b_i}{n_i + c_i} = \frac{n_{i1}}{n_{i2}}$.

The corresponding confidence intervals are computed based on the score statistic[21] with constrained maximum likelihood (CML). When only the marginal data is available i.e ‘paired’ data, the Koopman CI’s for the study-specific probability ratios are computed. These intervals are expected to be wider than the former.

4 Multiple treatments - network meta-analysis

More often, more than two treatments are evaluated against different comparators for the same condition and patient characteristics. Suppose, there are $Q$ treatments in total but only $Q_i$ are evaluated while the other $Q - Q_i$ treatments are missing-at-random in study $i$. Let $[n_{i1}, N_{i1}], \ldots, [n_{iQ_i}, N_{iQ_i}], x_i$ denote the data from study $i$ where $x_i = (x_{i1}, \ldots, x_{ip})$ denotes the values of $p$ covariates. $x_{i0}$ is a categorical variable identifying the different treatments in study $i$.

Inspired by factorial analysis of variance models where studies are equivalent to blocks, we propose a logistic-normal regression model expressed as follows:

$$
\pi_{iq}(x_i) = \frac{\exp(\mu_q + x_i \beta' + \vartheta_{iq} + \delta_i)}{1 + \exp(\mu_q + x_i \beta' + \vartheta_{iq} + \delta_i)}
$$

$$
\delta_i \sim N(0, \tau^2 \delta)
$$

$$
\vartheta_{iq} \sim N(0, \tau^2 \vartheta)
$$

where $\mu_q$ is the log-odds of 'success' of the $q^{th}$ treatment. With the inclusion of random study effect ($\delta_i$) as well as random treatment effects nested within a study ($\vartheta_{iq}$), this model implies a conceptual extension of the previous models for pairwise meta-analysis. As such both models are based on similar assumptions.

The imposed random-effects structure induces a positive correlation between treatments from the same study ($\delta_i$) and another between studies evaluating the same treatment/test ($\vartheta_{iq}$) resulting in a compound symmetry variance-covariance structure between the measurements. The resulting $Q \times Q$ variance-
covariance matrix of log-odds of successes in study $i$ is expressed as follows:

$$\begin{bmatrix}
\tau_\delta^2 + \tau_\delta^2 & \cdots & \tau_\delta^2 \\
\vdots & \ddots & \vdots \\
\tau_\delta^2 & \cdots & \tau_\delta^2 + \tau_\delta^2
\end{bmatrix}$$

Also called the intra-study correlation coefficient, the correlation between any two treatments is $\rho_\delta = \frac{\tau_\delta^2}{\tau_\delta^2 + \tau_\delta^2}$. It also measures the proportion of the variability accounted for by the between-study variability. It is $\rho_\delta = 0$ when the study effects convey no information and close to 1 the more identical the studies are.

To fit model 12, there should be at least 2 treatments per study to be able to separate the two variance components. The specification in the model 12 assumes homogeneous (equal) variance $\tau_\delta^2$ between the treatments. A more flexible model allows the variances to differ by treatment i.e $\tau_\delta^2 = \tau_\delta^2_1, \ldots, \tau_\delta^2_Q$ however this requires more data to identify the extra $Q-1$ variance parameters. This model is analogous to the model proposed by Nyaga, Arbyn and Aerts[22] for network meta-analysis of diagnostic accuracy studies but applied within the frequentist inference framework.

By incorporating the induced correlation in the model, there is a potential gain in precision. Other advantages of synthesizing all the evidence in one single model are elegance, parsimony in parameter estimations, possibility to combine direct and indirect evidence and obtain any conceivable contrasts even when such contrasts do not exist from the head-to-head comparisons. Furthermore, it avoids the inflation of type I errors (multiplicity) introduced by performing a series of head-to-head comparisons.[23]

Implicit weighting

The hierarchical modeling in which the binomial distribution models the within-study variation implies no need for concern about the weights. This is because the variance is inherent in the distribution and depends on the covariates through $p(x_i)$ and $N_i$ i.e. $Var(Y_i) = \frac{p(x_i)(1-p(x_i))}{N_i}$. Thus, the quality of the information in each study is fully encapsulated by the parameters in $p(x)$ and $N_i$.

The weights assigned to the study are those that account for the most variability in the data in the most parsimonious fashion as when the likelihood function is optimal.

Goodness-of-fit

While the focus of meta-analysis is on the estimation of the population-averaged means and the between-study variability, a goodness-of-fit analysis might be worthwhile to examine the parsimony of the model fit.

If too many variables are included in a meta-regression at once, significant variables could be dropped due to low statistical power.[15]

Both the Wald test and the likelihood ratio (LR) test can be conducted on the hypothesis test $H_0 : \beta = 0$ vs $H_0 : \beta \neq 0$ i.e a covariate has no-effect vs has an effect on the logit scale. Should there be a conflict between the Wald and the LR test, the LR test result is a more powerful test.
The LR test can also be conducted to compare a 'complex' model with a 'reduced' model. Examples of 'reduced' models include model without the study-specific random-effects $\delta_i$, without interaction terms, or with one covariate less. The results of these tests provide evidence on whether the model fits the data reasonably well or not. Non-convergence and inestimable effects are other indications of lack of fit for which we recommend fitting a simpler model.

Besides the Wald and the LR tests, one can also use the Akaike Information Criterion (AIC) to select the optimal model with the lowest value.

Software development
The *metapreg* command was developed in Stata 14.1.[24] The random- and the fixed-effects models were fitted using Stata commands *meqrlogit* and *binreg*, respectively. The command *ci proportions* was used to compute the study-specific confidence intervals of the proportions. It allows the Wald, Wilson, Agresti-Coull, Jeffreys and the exact confidence intervals for the study-specific proportions. The confidence intervals for the study-specific probability ratios were programmed in *Mata*; the Stata matrix programming language. We used the Stata command *margins* to obtain the model predictions and *nlcom* to obtain functions of these predictions. The Stata command *testnl* was used to perform Wald-type tests for the obtained functions. The *metapreg* program, its help file and other accompanying data files are available for download at https://econpapers.repec.org/software/bocbocode/s458693.html.

With a connection to the internet, directly install the program into Stata by typing *ssc install metapreg*. After the installation is complete, typing *help metapreg* opens the help window. The help file contains a detailed description of the command options and demonstration examples. These accompanying datasets are available with a click in the help window. Table 2 highlights some important features of *metapreg*.

Application
Random-intercept model - Proportion of women cured of CIN1 disease with cold coagulation.[25]
To demonstrate fitting of the intercept-only random-effects model and as a validation of metapreg, we reproduce table 3 in Nyaga, Arbyn and Aerts[1] which was generated by the command *metaprop one*.

The syntax used is;

```
. metapreg cured treated, ///
    studyid(study) cimethod(wilson) dp(4) ///
    nograph sumstat(ES) sumtable(abs)
```

The output is as follows;

```
********** Fitted model **********************
```
cured \sim \text{binomial}(\logit(p), \text{treated})
\logit(p) = \mu + \text{study}
\text{study} \sim N(0, \tau^2)
Number of observations = 7
Number of studies = 7

********************************************************************************
Study specific measures : ES
********************************************************************************

<table>
<thead>
<tr>
<th>Study</th>
<th>ES [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Javaheri, 1981</td>
<td>0.9565 0.7901 0.9923</td>
</tr>
<tr>
<td>Hussein &amp; Galloway, 1985</td>
<td>0.9091 0.6226 0.9838</td>
</tr>
<tr>
<td>de Cristofaro, 1990</td>
<td>1.0000 0.9162 1.0000</td>
</tr>
<tr>
<td>Rogstad, 1992</td>
<td>0.8000 0.5840 0.9193</td>
</tr>
<tr>
<td>Loobucycck &amp; Duncan, 1993</td>
<td>0.9695 0.9495 0.9817</td>
</tr>
<tr>
<td>Singh, 1998</td>
<td>0.8837 0.7552 0.9493</td>
</tr>
<tr>
<td>Joshi, 2013</td>
<td>0.9091 0.7219 0.9747</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9419 0.8855 0.9715</td>
</tr>
</tbody>
</table>

 Marginal summary: Absolute measures

<table>
<thead>
<tr>
<th>Effect</th>
<th>Prop. SE(logit) z(logit) P Lo Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.942 0.378 7.369 0 0.886 0.972</td>
</tr>
</tbody>
</table>

NOTE: H0: P = 0.5 vs. H1: P \neq 0.5
Test of heterogeneity - LR Test: RE model vs FE model

<table>
<thead>
<tr>
<th>DF</th>
<th>Chisq</th>
<th>p</th>
<th>tau2</th>
<th>I2tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1</td>
<td>4.0399</td>
<td>0.0222</td>
<td>0.4905</td>
</tr>
</tbody>
</table>

The equivalent syntax using \texttt{metaprop\_one} is
\begin{verbatim}
.metaprop\_one cured treated, ///
    random groupid(study)
    logit dp(4) lcols(study)
\end{verbatim}

Because the estimates of the study-specific proportions from \texttt{metaprop\_one} were similar to those from \texttt{metapreg} we omit most of the output except the following:

LR test: RE vs FE Model chi^2 = 4.0399
(d.f. = 5) p = 0.0222

Estimate of between-study variance

\[ \tau^2 = 0.4907 \]

Test of ES=0 : z = 7.3711 p = 0.0000

The significance test statistics were also similar with the following exceptions;
Independent studies meta-regression - Incomplete excision of cervical precancer as a predictor of treatment failure[26]

In 2017, Arbyn et al.[26] published a systematic review and meta-analysis to assess the risk of therapeutic failure associated with the histological status of the margins of the tissue excised to treat cervical precancer (CIN2+). They intended to assess the influence of covariates on the margin status, with the method of excisional treatment being one of them (cold-kife conisation (CKC), laser conisation (LC), large loop excision of the transformation zone (LLETZ), or mixed). We focus here on the effect of the excision procedure.

They performed a stratified analysis by treatment with metaprop. Their results are in column three of table 3. After computing the different summary proportions, metaprop also conducts a test of equality of those proportions. This test is merely an indication of the degree of evidence of no differences between the proportion but gives no information on the nature and the strength of the differences. This information can be obtained from the ratios of the proportions. The test statistics in this example were (chi = 6.99, d.f = 3, p-value = 0.07) indicating no differences in the pooled proportions by treatment.

After fitting a random-effects model, the test may be biased. Two possible sources of bias are; 1) the inefficiency of the method of moments in estimating the between-study variance which is required in computing the weights and consequently the variances of the overall and the sub-group proportions. 2) In calculating the heterogeneity statistic, the sub-group pooled estimates are treated as though they are fixed-effects estimates while they are random-effects estimates. To have a visual of the data, we call the metapreg command four times for each intervention group. The code is not presented here but the forest plots are displayed in figure 1. The following code ‘replicates’ the stratified analysis using metapreg.

```
metapreg tpos n, ///
    studyid(study) sumtable(abs) dp(2) ///
    power(2) plotregion(color(white)) ///
    graphregion(color(white)) ///
    by(treatment) stratify ///
    xlab(0, 25, 50, 75, 100) ///
    xtick(0, 25, 50, 75, 100) ///
    sumstat(Positive Margins (%)) ///
    olineopt(lcolor(red) ///
```
The results are presented in column four of table 3. The differences between columns three and four are; metaprop assumes normality and uses the MOM to obtain the parameter estimates while metapreg assumes a binomial-normal distribution and uses maximum likelihood. metaprop computes the $I^2$ statistic by Higgins and Thompson [27] which has been shown to lead to an incorrect conclusion of very high heterogeneity.[28] metapreg computes the $I^2$ statistic by Zhou and Dendukuri[28] which is more suitable for binomial-normal data.

It is quite surprising that metaprop yields a larger estimate of the variability between the studies that used cold knife conisation ($\tau^2 = 0.10$) than the studies that used mixed interventions ($\tau^2 = 0.07$) while a visual inspection of figure 1 suggests the opposite. In contrast, the estimates from metapreg ($\tau^2 = 0.66$ vs $\tau^2 = 0.99$) are in line with the observed variability in the forest plots. This discrepancy points to the statistical sub-optimality of the MOM in estimating the between-study variability.

Next, we fit a meta-regression model with treatment as a covariate to estimate the probability ratios with the following code.

```
. metaprop tpos n treatment, ///
    studyid(study) sumtable(abs rr) ///
    dp(3) plotregion(color(white)) ///
    graphregion(color(white)) ///
    lcol(location) ///
    xlab(0,.2, 0.4, 0.6, 0.8, 1) ///
    xtick(0,.2, 0.4, 0.6, 0.8, 1) ///
    sumstat(+ve Margins (%)) ///
    olineopt(lcolor(red) ///
    lpattern(shortdash)) ///
    diamopt(lcolor(red)) texts(1.2)
```

From the output below, the statistics of the LR test comparing the model with and without treatment were (chi = 6, d.f. = 3, p-value = 0.10). Hence, there are no differences between the summary proportion of positive margins between the treatment groups. Specific to when there is only one categorical variable, we came to the same conclusion by testing the equality of the positivity ratios. The p-value from this test was 0.52.

From the meta-regression model, we learn that large loop excision was associated with 43% higher positive margins than cold knife conisation (RR = 1.43 95% CI = 1.06, 1.92).

```
Marginal summary: Relative measures
----------------------------------------------
Effect |RR SE(lor) z(lor) P>|z| Lower Upper
----------------------------------------------
treatment
  CKC |1.000 0.000 1.000 1.000
  LC  |1.369 0.179 1.756 0.079 0.964 1.945
  LLETZ |1.425 0.153 2.316 0.021 1.056 1.922
```

```
Comparative studies meta-regression - Effect of latitude on the protective effect of BCG vaccination against Tuberculosis.\cite{9}

Using the Stata command `metareg`\cite{30}, Sharp and Sterne\cite{9} investigated the effect of latitude (degrees north or south from the Equator) on the effectiveness of BCG vaccination. A variance-weighted least-squares linear regression model was fitted on the log odds ratios with latitude as a covariate. Their analysis showed a significant negative association between the log odds ratio and the absolute latitude. They concluded that the benefit of BCG vaccination was greater on higher absolute latitude.

We now fit a logistic-normal regression model with `bcg`, a categorical variable for the treatment group, and `lat`, a continuous variable with information on the absolute latitude. To allow the effects of BCG to vary by absolute latitude, we also include the interaction between the two variables. The syntax is;

```
. use https://github.com/VNyaga/Metapreg/blob/master/bcg.dta?raw=true, clear
. metapreg cases
tb population bcg lat, studyid(study) sortby(lat) sumtable(all) design(comparative) outplot(rr) interaction plotregion(color(white)) graphregion(color(white)) xlab(0.1, 1, 2) xtick(0.1, 1, 2) olineopt(lcolor(red) lpattern(shortdash)) diamopt(lcolor(red)) lcols(lat) rcols(cases tb population) astext(80) texts(2) logscale xline(1, lcolor(black))
```

`metapreg` stores the model estimates under the name `metapreg_modest`. The raw coefficients can be displayed by typing `estimates replay metapreg_modest`. The coefficient from the interaction between BCG and lat was \(-0.03334\) [95\% CI: \(-0.0388, 0.0074\)].
-0.02488]. This is comparable to the coefficient for lat -0.0320 [95% CI: -0.0417, -0.0223] when regressed against the log odds ratio of BCG vaccination as observed by Sharp and Sterne.[9]

Using the options `design(comparative), output(rr) and sortby(lat)`, we generated the forest plot (figure 2) of the rate ratios sorted by lat analogous to the forest plot of the log odds ratio in figure 2 of Sharp and Sterne.[9]

As observed in the original analysis, figure 2 shows that the effectiveness of BCG vaccination against tuberculosis was greater in higher altitudes.

**Multiple-treatment (network) meta-analysis -Response rate to the treatment of acute mania in adults with diverse drugs[31]**

In 2011, Cipriani et al.[31] systematically reviewed 47 randomised controlled trials (16 073 participants) from Jan 1, 1980, to Nov 25, 2010, which compared the proportions of patients who responded to 13 treatments of acute mania in adults. The treatments included placebo (PLA), aripiprazole (ARI), asenapine (ASE), carbamazepine (CARB), valproate (VAL), haloperidol (HAL), lamotrigine (LAM), lithium (LITH), olanzapine (OLA), quetiapine (QUE), risperidone (RIS), topiramate (TOP), and ziprasidone (ZIP) (see figure 3).

First, they used the Dersimonian-Laird [32] method to obtain the direct summary ORs in head-to-head comparisons of the antimanic drugs relative to placebo in Stata. They reported that all antimanic drugs were significantly more effective than the placebo except TOP.

They then performed a multiple-treatments meta-analysis to obtain the mixed summary ORs of the antimanic drugs relative to the placebo in Winbugs. They reported that ASE, ZIP, LAM, and TOP were not significantly more effective than the placebo. They reported further that the wide CIs from the multiple-treatment meta-analysis made it difficult to draw clear conclusions.

In 2013, Chaimani et al.[33] used this dataset to demonstrate the use of `mvmeta`[14] for network meta-analysis in Stata. The model implemented in `mvmeta`[14] assumes the following; in a network with T treatments, there is a reference treatment A present in all studies. The model has T – 1 basic parameters formed by contrasting the T treatments to A. In studies that do not report the reference treatment, the missing data are imputed. To perform the analysis, the program requires all study-specific pairwise contrasts (typically log OR), their respective variances, and covariances (in studies with more than two treatments). These pairwise contrasts are treated as outcomes that follow a multivariate normal distribution.

We will use this data to demonstrate network meta-analysis using `metapreg` in similar steps to Cipriani et al.[31] To explore the data, we first obtained the summary response rate by fitting an intercept-only random-effects model for each treatment separately. The syntax is as follows:

```
clear
.metapreg event total, stratify by(drug) ///
studyid(study) model(random) ///
sumtable(all) ///
```
The two options `stratify` and `by(drug)` together enable us to fit separate models for each treatment group and consolidate the results in one graph and table. In contrast, the prefix `by drug:` would generate separate graphs and tables.

A visual inspection of figure 4 suggests that TOP was less effective than the placebo, LAM and ASE were similar to the placebo, while the other treatments were better than the placebo.

While the number of studies evaluating a certain treatment differed, TOP, LAM, and ASE were evaluated only once (see table 4).

From table 4, the highest between-study variation was observed in studies that evaluated OLA and RIS. Some of the cells in the table for TOP, LAM, and ASE with "." are empty because whenever there less than three studies, the `metapreg` fits a fixed-effects model. As such, the automatic LR test comparing the random-effects model with the fixed-effects is not conducted. Further, the between-study variance and the $I^2$ are also absent.

Since all studies except one included the placebo, we can examine the significance of the contrasts in series of head-to-head comparative analyses of the active treatments relative to the placebo.

Before we fit the model, we need to put the data into the right shape for comparative analysis. This is necessary because some studies evaluated three treatments (and thus contribute three rows of data) while comparative analysis as performed with `metapreg` expects two rows per study. In the reshaping exercise, we need to identify studies that included the placebo, group the studies by the active treatment and sort the data such that the placebo is placed first in each set of "paired" data per study.

We do this as follows:

1. Identify the placebo treatment;
   ```
   gen placebo = 0
   replace placebo = 1 if drug == "PLA"
   ```
2. Sort the data such that in each study, the placebo is place first
   ```
   gsort study {placebo
   ```
3. Assign a sequence index to the treatments in each study;
   ```
   bys study: egen T = seq()
   ```
4. Count how many treatments are evaluated in each study and also how many observations are in the dataset.
   ```
   bys study: egen nt = count(drug)
   count
   global nobs = r(N)
   ```
5. In studies with three treatments, we replicate the data from the placebo treatment. This creates the comparative data for the third treatment.
   ```
   gen nexpand = 1
   ```
replace nexpand = 2 if placebo & nt == 3
expand nexpand
6 Rank the data from the third treatment forth.
replace T = 4 if nt == 3 & T == 3
7 Rank the new data from the placebo treatment third.
replace T = 3 if _n>$nobs
8 Sort the data once more so that the placebo appears second in each “paired”
   set of data per study.
gsort study -T
9 Identify which studies included the placebo
bys study : egen PLA = max(placebo)
10 Since the data is properly sorted, the first row in paired set of data should
   be data from the active treatment and the second row from the placebo. We
   generate a sequence to identify the first and second row.
egen 0 = seq(),f(1) t(2) b(1)
11 Identify the active treatment in each study by picking the drug in the first
   row of each pair.
gen treatment = drug if 0 == 1
replace treatment = treatment[_n - 1] if 0 == 2
12 Sort the data by the treatment group and place the data from the placebo
   first. This is important because the category that appears first is assigne
d by the reference category unless if instructed otherwise.
gsort treatment study -placebo
The data is now ready for a stratified comparative meta-analyses analogous to the
first analysis in Cipriani et al.[31] The syntax for this procedure is as follows;
   . metapreg event total drug if PLA, ///
      studyid(study) nomc ///
      model(random) design(comparative) ///
      sumtable(all) ///
      outplot(rr) stratify by(treatment) ///
      plotregion(color(white)) ///
      graphregion(color(white) margin(zero)) ///
      xlab(.5, 1, 2, 2.5) ///
      xtick(.5, 1, 2, 2.5) ///
      sumstat(Response Rate Ratio) ///
      diamopt(lcolor(red)) ///
      texts(1.5) logscale xline(1) ///
      astext(40) xsize(7) ysize(10)

Basically, the command above fits a logistic-normal regression model with drug,
a binary variable for treatment which is either placebo or an antimanic drug. This
is done for each drug as indicated by the options stratify by(treatment). The
option outplot(rr) requests for a forest plot of the study-specific response rate
ratios which is possible since the studies are comparative as indicated by the option
design(comparative). To save computational time we direct the command not to
do model comparison with the option nomc. Otherwise, the program would also fit
a logistic regression model without drug then perform a LR test to compare the fit
of the model with and without the covariate.
Though we requested for all summary tables with the option `sumtable(all)`, we will only present the forest plot.

The results presented in figure 5 indicated that all treatments were significantly more effective than placebo with the exception of TOP. The results in this analysis are congruent to the conclusion of Cipriani et al. [31] Though our earlier suspicions on the efficacy of the antimanic drugs relative to the placebo have been confirmed, we are missing the comparison between LAM and the placebo.

To have a complete and clearer picture, we will now proceed to network meta-analysis using the original data before the reshape. The syntax to perform the multiple treatment meta-analysis is:

```
use https://github.com/VNyaga/Metapreg/blob/master/maniacefficacy.dta?raw=true
metapreg event total drug, studyid(study) ///
  sumtable(all) outplot(rr) ///
  design(network, baselevel(PLA)) ///
  plotregion(color(white)) ///
  graphregion(color(white)) ///
  xlab(.5, 1, 1.5, 2) ///
  xtick(.5, 1, 1.5, 2) ///
  sumstat(Response Rate Ratio) ///
  texts(1.2) logscale xline(1)
```

The command above fits a logistic-normal regression model with `drug`, a categorical variable identifying the 14 treatments. We specify that we want a network meta-analysis and assign the placebo treatment as the reference category with the option `design(network, baselevel(PLA))`.

The first output of the program is the symbolic representation of the fitted model:

```
event ∼ binomial(logit(p), total)
logit(p) = mu.drug + drug + study
study ∼ N(0, sigma.sq)
drug ∼ N(0, tau.sq)
```

In addition to the study-specific random-effects, the model includes a second random-effect `drug` which enters the model as nested factor within a study. From the output tables (not presented here), the total variability in log-odds was 0.20 of which 90.17% could be attributed to between-study variability.

The mixed evidence on the relative efficacy of the evaluated antimanic drugs relative to placebo is summarized in figure 6. From the forest plot, all treatments except LAM and TOP were significantly more effective than placebo. Further scrutiny reveals that the confidence intervals are narrower than from the head-to-head comparisons presented in figure 6.

**Discussion**

The true distribution of a random variable can never be validated. However, if a statistical model is appropriate for the data at hand, the behavior of the observed data should reflect the assumed properties of the specified distribution. The logistic regression model provides a general framework for meta-analysis of proportions that respects the binomial nature of the data.
Explicitly defining weights in meta-analysis is perceived as a standard step in a sound meta-analysis. Without understanding the mechanics of the unorthodox implicit weighting within the logistic regression framework some may be critical of its validity in meta-analysis. This is not a surprise given the popularity of the Dersimonian-Laird [29] inverse-variance method for random-effects meta-analysis. This method has been widely used for over 20 years and often taught in meta-analysis courses. It uses empirical weights. Stating that methods using weighting scheme could be flawed might seem outrageous. But not treating the weights as random variables should be considered as a fundamental flaw in the procedures that treat weights as known.

Short-comings of the inverse-variance method include the bias in estimation of the pooled effect, underestimation of the between-study variance, and poor coverage of the obtained confidence intervals, especially for sparse data and/or small sample sizes.[34, 35] In 2010, Shuster [36] disapproved the use of empirically based weighted random-effects meta-analysis. A simulation study published in 2011 [37], showed that the inverse-variance weights methodology was by far the most unreliable in a fixed-effects meta-analysis model comparing two proportions.

The success of metaprop indicates how big the gap was (and is) between accessible procedures for meta-analysis of proportions and the end-users. It is an approximate method with limited scope as it does not attempt to explain the excess variability of the proportions.

metaprop was the first product of the OPSADAC project (Optimization of statistical procedures to assess the diagnostic accuracy of cervical cancer screening tests) [38] conceived to develop innovative and optimized statistical procedures for meta-analysis of proportions and diagnostic accuracy studies and finally disseminate them to the end users as user-friendly statistical programs.

In continuation of the project, we developed a model-based framework for meta-analysis of binomial data based on the logistic regression. The framework is flexible, efficient, and provides more precise estimates. To make the framework accessible, we developed metapreg, a statistical program in Stata.

The program has capabilities to perform meta-regression when studies are independent, comparative, paired, or matched, and assess multiple treatments. It allows both continuous and categorical covariates and interaction terms between the first covariate and the rest. When studies are comparative, paired, or matched, either the study-specific proportions or the relative probabilities can be plotted in a forest plot. Furthermore, the program can perform model comparisons in meta-regression.

We have demonstrated the use of metapreg with data from three published data to serve as a starting point in drawing users closer toward the use of better methods recommended for binomial data.

Acknowledgments
Funding
Financial support was received from Horizon 2020 Framework Programme for Research and Innovation of the European Commission, through the RISCC Network (Grant No. 847845), the Belgian Cancer Centre and the European Society of Gynaecological Oncology (ESGO).
Abbreviations

Availability of data and materials
The metadta program was developed in Stata 14.1. The code, the help files used herein are publicly available for download at https://econpapers.repec.org/software/bocode/s458693.htm. The code to reproduce the analysis herein can be downloaded at https://github.com/VNyaga/Metapreg/blob/master/metapreg-article-code.do.

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Authors’ contributions
V.N.N. conceived the need for guidance on this topic, developed metapreg, and drafted the manuscript. M.A. conceptualized the OPSADAC project (Optimization of statistical procedures to assess the diagnostic accuracy of cervical cancer screening tests) from which this publication springs from and edited the manuscript. Both authors reviewed and approved the final manuscript.

Author details
Unit of Cancer Epidemiology, Sciensano, Brussels, Belgium.

References
14. White, I.: Mvmeta: Stata module to perform multivariate random-effects meta-analysis (2022)


doi:10.1111/jebm.12443


doi:10.1016/S1470-2045(17)30700-3


doi:10.1002/sim.1186


doi:10.1002/sim.6115


doi:10.1016/j.cct.2006.04.004


doi:10.1177/1536867X0800800403


doi:10.1016/S0140-6736(11)60873-8


doi:10.1016/0197-2456(86)90046-2


doi:10.1371/journal.pone.0076654


doi:10.1111/j.1751-5823.2013.00187.x


doi:10.1002/sim.7233


doi:10.1002/sim.3607


doi:10.1002/jrsm.20


Figures

Figure 1: Forest plot - Proportion of cones with positive resection margins, by treatment procedure (CKC – cold knife conisation, LC – laser conisation, LLETZ - large loop excision of the transformation zone).

Figure 2: Forest plot - Meta-analysis on the protective effect of BCG against tuberculosis.
Figure 3: Network plot - Eligible comparisons for the multiple treatments meta-analysis for efficacy in acute mania.

Figure 4: Forest plot - Stratified independent meta-analyses of proportions of patients with acute mania who responded to 3-week anti-mania treatment with diverse drugs.

Figure 5: Forest plot - Stratified comparative meta-analyses of comparisons between antimanic drugs and placebo.

Figure 6: Forest plot - The network meta-analysis of the treatment response rate ratio with placebo as the reference treatment. Response rate > 1 indicate higher efficacy than placebo.

Table 1: Cross-tabulation of successes in the case and control group.

<table>
<thead>
<tr>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Success)</td>
<td>$a_i$</td>
</tr>
<tr>
<td>0 (Failure)</td>
<td>$b_i$</td>
</tr>
<tr>
<td>1 (Success)</td>
<td>$c_i$</td>
</tr>
<tr>
<td>0 (Failure)</td>
<td>$d_i$</td>
</tr>
<tr>
<td>Option in metapreg</td>
<td>Descriptions</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>design()</td>
<td>Specifies the shape of data or the design of meta-analysis to perform.</td>
</tr>
<tr>
<td>cimethod()</td>
<td>Specifies the type of confidence intervals for the study-specific estimates as displayed in the forest plot.</td>
</tr>
<tr>
<td>nomc</td>
<td>Instructs the program not to do model comparison in meta-regression.</td>
</tr>
<tr>
<td>by()</td>
<td>Requests the marginal estimates at unique values of the by variable.</td>
</tr>
<tr>
<td>stratify</td>
<td>Together with the option by(), the stratify option makes it possible to fit separate models in each group of data in the by variable, but present the results in one table and plot.</td>
</tr>
<tr>
<td>sumtable()</td>
<td>Indicates the type of marginal estimates to display.</td>
</tr>
<tr>
<td>outplot()</td>
<td>specifies which statistics to display on the forest plot.</td>
</tr>
</tbody>
</table>
Table 3: Pooled proportions of incomplete excisions by treatment procedure as synthesized by metaprop and metapreg.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Studies</th>
<th>metaprop</th>
<th>metapreg (strati-</th>
<th>metapreg (unified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>fied analysis)</td>
<td>(stratified analysis)</td>
</tr>
<tr>
<td>cold-knife conisation</td>
<td>17</td>
<td>20·17%</td>
<td>18·77%</td>
<td>17·2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CI[14·34 – 26·71]</td>
<td>CI[13·49 – 25·52]</td>
<td>CI[13·30 – 21·90]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>τ² = 0.10</td>
<td>τ² = 0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I² = 98·35%</td>
<td>I² = 94·91%</td>
<td></td>
</tr>
<tr>
<td>laser conisation</td>
<td>13</td>
<td>17·76%</td>
<td>16·82</td>
<td>23·5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>τ² = 0.05</td>
<td>τ² = 0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I² = 95·35%</td>
<td>I² = 89·94%</td>
<td></td>
</tr>
<tr>
<td>large loop excision of the</td>
<td>42</td>
<td>25·89%</td>
<td>24·99%</td>
<td>24·4%</td>
</tr>
<tr>
<td>transformation zone</td>
<td></td>
<td>CI[22·32 – 29·62]</td>
<td>CI[21·29 – 29·10]</td>
<td>CI[20·30 – 29·10]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>τ² = 0.07</td>
<td>τ² = 0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I² = 95·76%</td>
<td>I² = 91·11%</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>22</td>
<td>23·72%</td>
<td>22·04%</td>
<td>21·2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CI[18·90 – 28·89]</td>
<td>CI[15·61 – 30·17]</td>
<td>CI[17·10 – 25·90]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>τ² = 0.07</td>
<td>τ² = 0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I² = 96·68%</td>
<td>I² = 95·55%</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>94</td>
<td>23·13%</td>
<td>22·08%</td>
<td>22·10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>τ² = 0.10</td>
<td>τ² = 0.65</td>
<td>τ² = 0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I² = 97·63%</td>
<td>I² = 93·67%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Heterogeneity statistics.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Studies</th>
<th>P (RE vs FE)</th>
<th>sigmaq</th>
<th>I²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARI</td>
<td>7</td>
<td>0</td>
<td>0.07</td>
<td>73.73</td>
</tr>
<tr>
<td>PLA</td>
<td>36</td>
<td>0</td>
<td>0.18</td>
<td>78.22</td>
</tr>
<tr>
<td>HAL</td>
<td>8</td>
<td>0</td>
<td>0.16</td>
<td>60.76</td>
</tr>
<tr>
<td>QUE</td>
<td>7</td>
<td>0</td>
<td>0.12</td>
<td>76.57</td>
</tr>
<tr>
<td>LITH</td>
<td>8</td>
<td>0</td>
<td>0.35</td>
<td>70.03</td>
</tr>
<tr>
<td>ZIP</td>
<td>5</td>
<td>0</td>
<td>0.08</td>
<td>75.77</td>
</tr>
<tr>
<td>OLA</td>
<td>13</td>
<td>0</td>
<td>0.43</td>
<td>88.19</td>
</tr>
<tr>
<td>DIV</td>
<td>8</td>
<td>0</td>
<td>0.15</td>
<td>55.45</td>
</tr>
<tr>
<td>RIS</td>
<td>5</td>
<td>0</td>
<td>0.21</td>
<td>86.08</td>
</tr>
<tr>
<td>CARB</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LAM</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>PAL</td>
<td>2</td>
<td>0.19</td>
<td>0.02</td>
<td>54.07</td>
</tr>
<tr>
<td>TOP</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>ASE</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>