

Introduction to Meta- Analysis

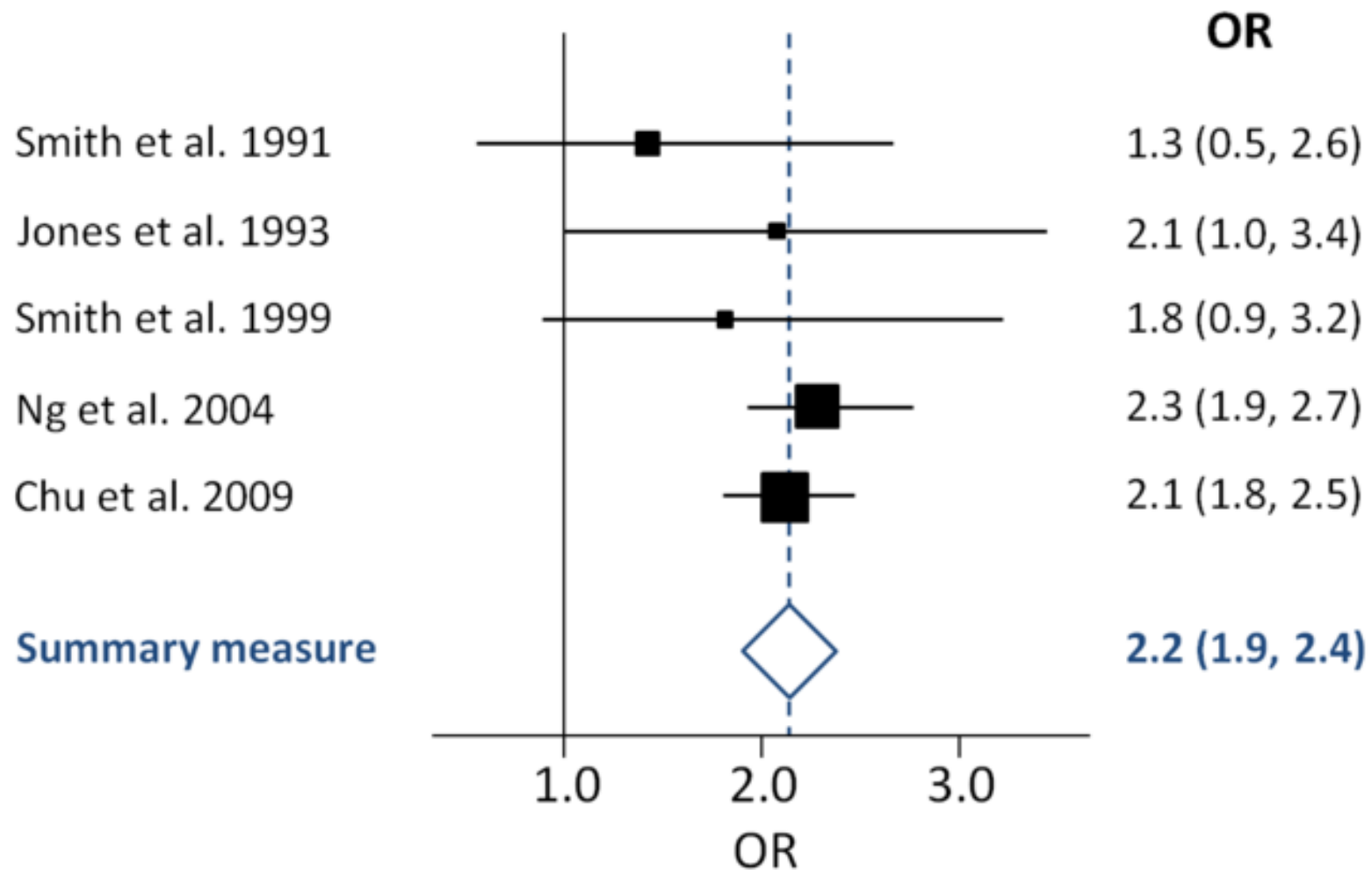
Caroline Andy, MS

Research Biostatistician II, WCM Division of Biostatistics

November 16, 2023

What is meta-analysis?

- Meta-analysis is a research method that combines the results of multiple studies to create a more precise estimate of an effect size.
- Meta-analysis is an essential tool for synthesizing evidence needed to inform clinical decision making and policy.



Terminology

Systematic Review (SR): Review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review.

Meta-analysis (MA): Use of statistical techniques in an SR to integrate the results of included studies to conduct statistical inference.

- Can use granular subject-level data or summary statistics from studies (i.e., odds ratio, risk ratio, hazards ratio, standardized mean difference, etc.)

Components of a meta-analysis

1. Conducting the meta-analysis to obtain a pooled estimate
 - Test for statistical heterogeneity
 - Selecting random or fixed effect model
2. Assessment of publication bias
 - Trim and fill method
3. Robust variance estimation meta-analysis
4. Investigate sources of heterogeneity using meta-regression
5. Advanced topics in meta-regression: clustered meta-forest methods



HEALTH | WELLNESS | YOUR HEALTH

Can Hormone Therapy Raise Your Risk for Alzheimer's?

The medication might lower—or at least not increase—your dementia risk if you take it in midlife



The New York Times

Do Hormone Treatments for Menopause Increase Dementia Risk?

A new study suggests a link, but experts caution that there's more to the story.

Share full article



CNN health

Sweet spot for HRT may reduce dementia risk by nearly a third, study says

By Sandee LaMotte, CNN

6 minute read · Updated 7:12 AM EDT, Thu November 2, 2023

(CNN) — Hormone replacement therapy may protect the female brain from Alzheimer's disease and dementia — if hormones are taken in the 40s and 50s when menopausal symptoms begin, according to a recent meta-analysis.

Capital Daily

How Hormone Replacement Could Help Prevent Alzheimer's

A new study including a sample of six million women suggests that midlife HRT could substantially cut the risk of developing Alzheimer's.



By Rozia Hussain | Published: Oct 24, 2023



Conducting the meta-analysis and testing statistical heterogeneity

Ref	midlife	latelife	estrogen-only	estrogen-progesterone	short_duration	long duration	RR	RR_L95	RR_U95
Heyman et al., 1984	NA	NA	NA	NA	NA	NA	1.5427249	0.8366600	2.792848
Amaducci et al., 1986	NA	NA	NA	NA	NA	NA	1.3038405	0.6324555	2.428992
Borenstein et al 1990	NA	NA	NA	NA	NA	NA	1.0723805	0.7071068	1.612451
Broe et al., 1991	NA	NA	NA	NA	NA	NA	0.8831761	0.6324555	1.264911
Graves et al, 1990	NA	NA	NA	NA	NA	NA	1.0954451	0.7071068	1.612451
Paganini-Hill et al., 1994	NA	NA	NA	NA	NA	NA	0.8366600	0.6324555	1.095445

Test for heterogeneity

```
1 # conducting meta-analysis and test for statistical heterogeneity
2 meta = meta::metagen(log(sample_dat$RR),
3     lower = log(sample_dat$RR_L95),
4     upper = log(sample_dat$RR_U95),
5     studlab = sample_dat$Ref,
6     sm = "RR", # summary measure
7     random = T, # random effects conducted
8     method.bias = "Egger",
9     method.tau = "DL")
```

Number of studies combined: k = 32

	RR	95%-CI	z	p-value
Common effect model	0.9879	[0.9754; 1.0005]	-1.89	0.0592
Random effects model	0.8110	[0.6973; 0.9432]	-2.72	0.0066

Quantifying heterogeneity:

$\tau^2 = 0.1473$ [0.0445; 0.2627]; $\tau = 0.3838$ [0.2109; 0.5126]
 $I^2 = 96.9\%$ [96.3%; 97.4%]; $H = 5.70$ [5.20; 6.25]

Test of heterogeneity:

Q	d.f.	p-value
1008.62	31	< 0.0001

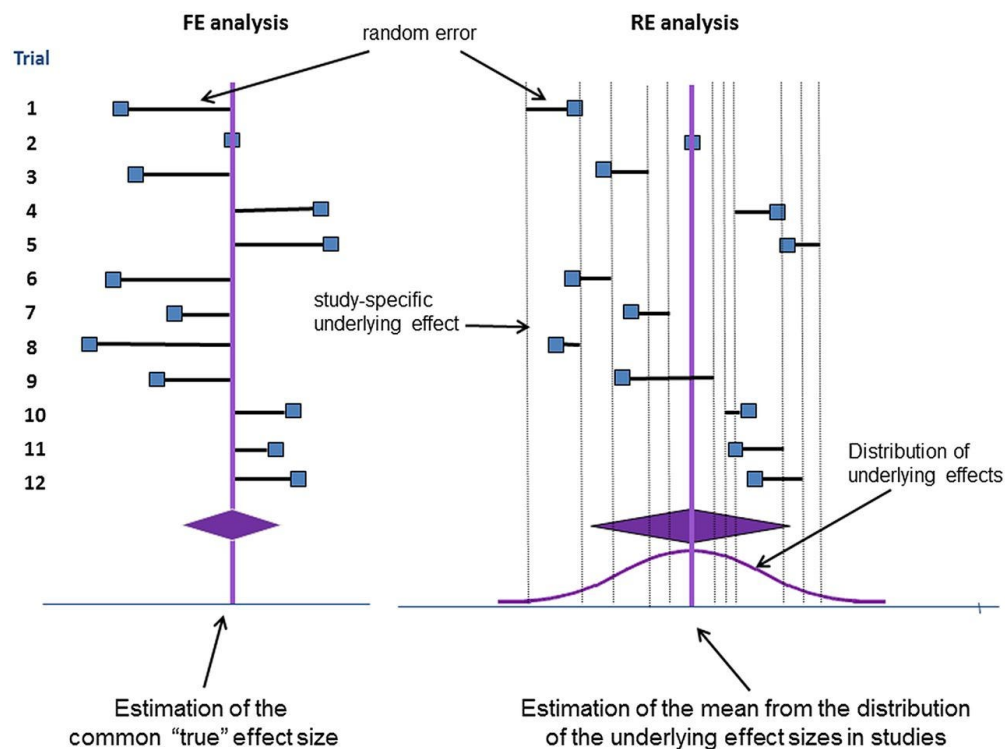
Details on meta-analytical method:

- Inverse variance method
- DerSimonian-Laird estimator for τ^2
- Jackson method for confidence interval of τ^2 and τ

Random vs. fixed effect models

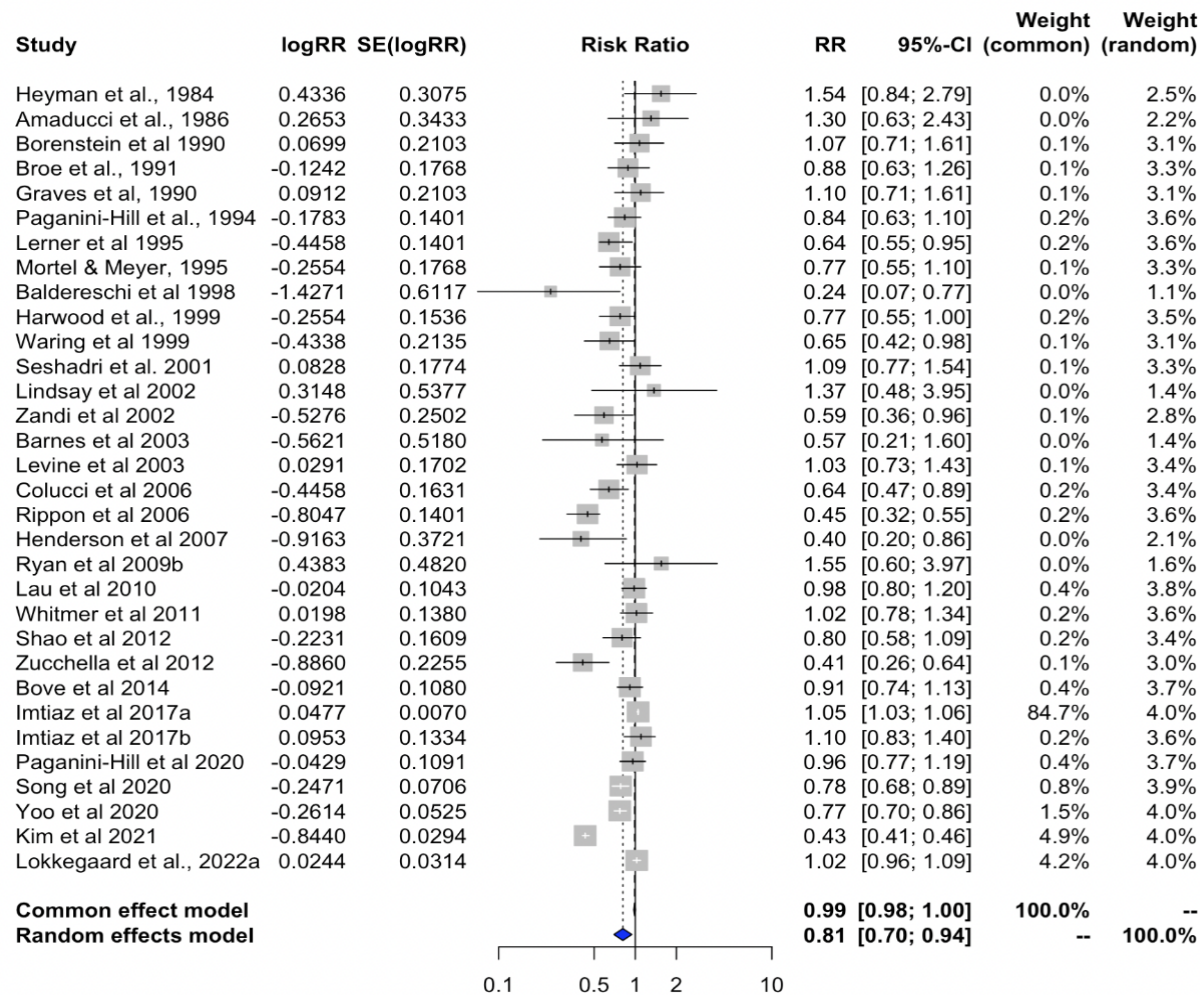
Fixed effect model: Assumes that the true effect of intervention is the same value in every study

Random effect model: Involves an assumption that the effects being estimated in the different studies are not identical, but follow some distribution. The model represents our lack of knowledge about why real, or apparent, intervention effects differ by considering the differences as if they were random. The centre of this distribution describes the average of the effects, while its width describes the degree of heterogeneity



Forest plot

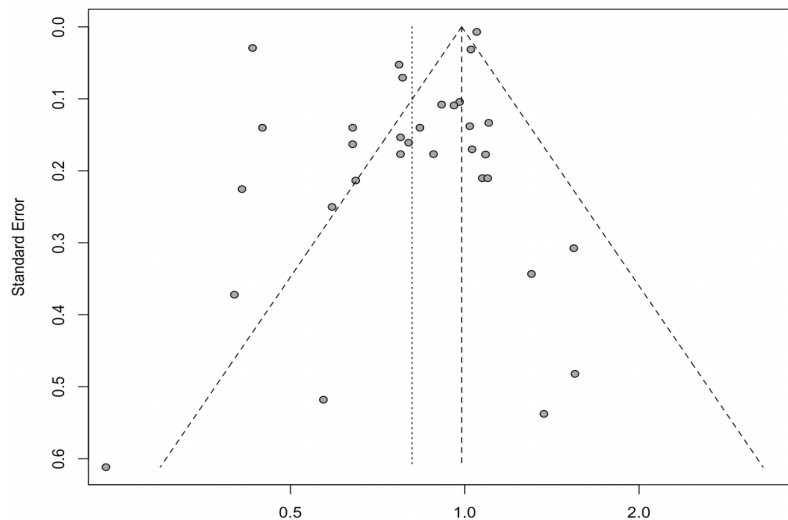
```
1 # Odds ratio meta-analysis plot
2 meta::forest(meta, comb.fixed = T, comb.random = T, weight.study = "random", col.diamond = "blue")
```



Publication bias

Of course, journals and researchers alike are biased to publish results that confirm the expectation or understanding of general stakeholders (including patients, physicians, scientists, decision makers, and sponsors). This can bias the effect estimates that we have access to when performing a meta-analysis. Therefore, it is important to always consider whether publication bias is present in your study sample.

```
1 # Funnel plot
2 funnel(meta)
3
4 # Rank test for publication bias p-value
5 metabias(meta)
```



Linear regression test of funnel plot asymmetry

Test result: $t = -1.89$, $df = 30$, $p\text{-value} = 0.0687$

Sample estimates:

bias	se.bias	intercept	se.intercept
-2.0392	1.0800	0.0207	0.0395

Details:

- multiplicative residual heterogeneity variance ($\tau^2 = 30.0496$)
- predictor: standard error
- weight: inverse variance
- reference: Egger et al. (1997), BMJ

Trim and fill method

When significant publication bias is detected, a trim and fill procedure can be applied which aims to estimate potentially missing studies due to publication bias and adjust the overall effect estimate.

This is a two-step method which 1) removes the smaller studies with large standard errors that cause funnel plot asymmetry, and 2) uses the trimmed funnel plot to estimate the true center of the funnel, replacing the omitted studies and their missing counterparts around the center.

```
1 taf_meta <- trimfill(meta)
```

Subgroup	Estimate	P-value
Raw estimates	0.81 (0.7, 0.94)	0.007
Trim and fill applied	1.01 (0.86, 1.18)	0.936

Robust variance estimation

A fundamental assumption of meta-analysis is that each observation (study estimate) is independent of the other study estimates included in the sample. This is often a reasonable assumption, however, you may encounter studies - like this HRT analysis - in which multiple estimates per study are presented. In this case, there may be induced correlation between within-study estimates. Thus you'll want to make use of robust variance estimation or "multi-level" meta-analysis modeling to account for this.

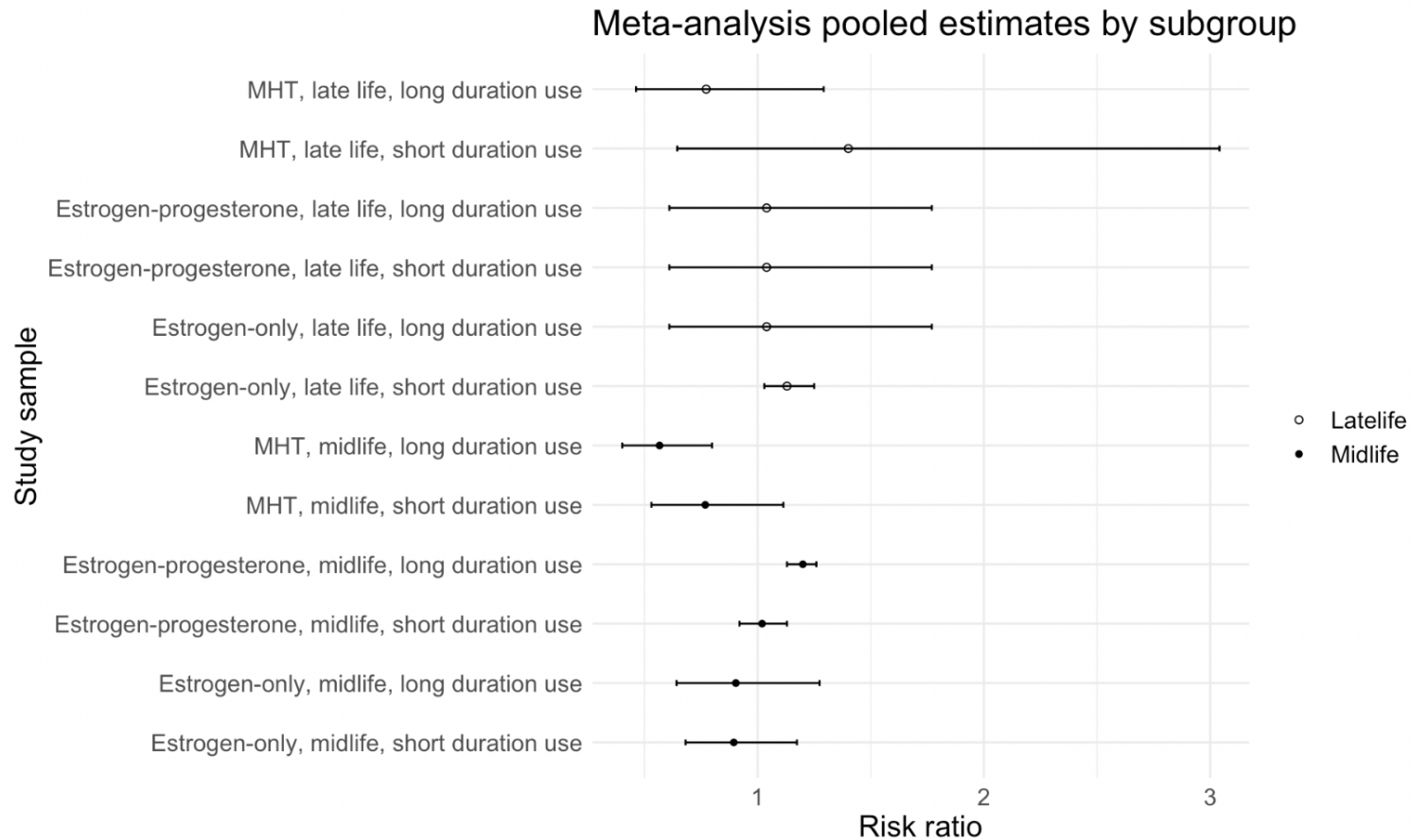
In the example up until now, we have considered a dataset which contains 1 estimate selected for each study. For the following steps, the entire dataset will be considered which includes all available estimates per study, where multiple estimates per study pertain to different subgroups within the study cohort.

```
1 meta_rve <- robumeta::robu(formula = RR ~ 1,  
2                             data = dat_rve,  
3                             studynum = Ref, # studynum tells R to aggregate by study  
4                             var.eff.size = var) # var.eff.size associates a variance with each effect size
```

Analysis	Risk ratio	Lower 95 CI	Upper 95% CI	P-value	Significance
Robust variance estimation	0.82	0.742	0.898	<0.001	***

Stratified meta-analysis

It may be of interest to us to consider effect estimates within particular subgroups. For example, in the HRT-Alzheimer's study, we found that women who began HRT earlier in life around menopause saw a protective association of HRT use and the Alzheimer's outcome, whereas women who began in late life saw an increased risk of Alzheimer's associated with HRT use.



Clustered meta-regression

If we identify that our study sample is highly heterogeneous, it is usually of interest to learn what factors may be driving this heterogeneity. Meta-regression investigates the extent to which statistical heterogeneity between results of multiple studies can be related to one or more characteristics of the studies. It creates a model describing the linear relationship between study-level covariates and the effect size.

Meta-regression model using metafor package to apply RVE and therefore include multiple estimates per study (clustered meta-regression):

```
1 # Meta-regression model using metafor package to apply RVE / clustered meta-regression
2 m_multi <- metafor::rma.mv(RR, var,
3     mods = ~ Outcome + Exposure + `Patient population` + `Duration of use` +
4         study_design + samp_size + time_period + Measure,
5     random = list(~ 1 | effect_id, ~ 1 | Ref), data = dat_rve)
```

Meta-regression model using meta package when the study sample includes only 1 estimate per study (no RVE):

```
1 # Meta-regression model using meta package when only 1 estimate per study (no RVE)
2 m.gen.reg <- meta::metareg(meta, ~Exposure + `Patient population`)
```

Meta-regression results

	Estimate	Lower 95% CI	Upper 95% CI	P-value
Intercept	0.557	0.289	0.824	<0.001
Outcome: Dementia	-0.003	-0.028	0.021	0.778
Exposure: Estrogen-progesterone	0.031	0.010	0.052	0.003
Exposure: MHT	0.075	0.036	0.114	<0.001
Patient population: Late life	0.050	0.015	0.086	0.006
Patient population: Overall	0.097	0.063	0.132	<0.001
Overall duration of use	0.012	-0.015	0.039	0.374
Long duration of use	0.004	-0.020	0.027	0.768
Case-control study design	0.105	-0.110	0.320	0.34
Cross-sectional study design	0.080	-0.427	0.588	0.756
Sample size >= 500	0.031	-0.160	0.223	0.748
Time period: 1995-2010	-0.093	-0.294	0.107	0.362
Time period: Before 1995	0.029	-0.242	0.301	0.832
Effect estimate: OR	0.065	-0.217	0.348	0.651
Effect estimate: HR	0.231	-0.016	0.479	0.067

Meta-forest methods

MetaForest is an adaptation of the random forests algorithm for meta-analysis. Like all random forests, this approach has several important advantages:

1. Performs variable selection, identifying which moderators contribute most strongly to the effect size found. (This is also great when dealing with concerns of collinearity/association of covariates)
2. Is a non-parametric technique, which means that they can easily capture non-linear relationships between the moderator and effect size, or even complex, higher-order interactions between moderators. This is a main advantage when performing meta-analysis on a heterogeneous body of literature.
3. Is robust to overfitting. A main limitation of tree-based methods is overfitting, but as with other random forest regression methods, this limitation is overcome with the use of bootstrapping.

Importantly, in the context of meta-analysis, we can also perform clustered MetaForest regression which allows us to take into account hierarchical correlation structure which arises from including multiple estimates per study.

```
1 # Using 5000 trees based on the convergence plot
2 set.seed(50)
3 # Model with 10000 trees for replication
4 mf_rep <- metaforest::MetaForest(RR~., data = data, vi = "var",
5                                 study = "n_study",
6                                 whichweights = "random",
7                                 num.trees = 5000)
8 # Run recursive preselection, store results in object 'preselect'
9 preselected <- metaforest::preselect(mf_rep,
10                                    replications = 100,
11                                    algorithm = "recursive")
12 # Plot the results
13 plot(preselected)
```


Meta-forest variable importance plot

