

# metaselection

Selection models for meta-analyses of dependent effect sizes

James E. Pustejovsky and Martyna Citkowicz

2025-06-13

# Collaborators from the American Institutes for Research

Megha Joshi



Ryan Williams

Joshua Polanin

Melissa Rodgers

David Miller

## Acknowledgement

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305D220026 to the American Institutes for Research. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

# Selective reporting of primary study results

- Selective reporting occurs if affirmative findings are more likely to be reported and available for inclusion in meta-analysis
- Selective reporting distorts the evidence base available for systematic review/meta-analysis
  - Inflates average effect size estimates from meta-analysis
  - Biases estimates of heterogeneity  
([Augusteijn, van Aert, and van Assen 2019](#))
- Concerns about selective reporting span the social, behavioral, and health sciences.



# Many available tools for investigating selective reporting

- Graphical diagnostics
  - Funnel plots
  - Contour-enhanced funnel plots
  - Power-enhanced funnel plots (sunset plots)
- Tests/adjustments for funnel plot asymmetry
  - Trim-and-fill
  - Egger's regression
  - PET/PEESE
  - Kinked meta-regression
- Selection models
  - Weight-function models
  - Copas models
  - Sensitivity analysis
- p-value diagnostics
  - Test of Excess Significance
  - $p$ -curve /  $p$ -uniform /  $p$ -uniform\*



# But few that accommodate dependent effect sizes

Treatment	O P Q
Control	O P Q

$$\begin{aligned} d_{O1} \\ d_{P1} \\ d_{Q1} \end{aligned}$$

Treatment	$O_1$	$O_2$	$O_3$
Control	$O_1$	$O_2$	$O_3$

$$\begin{aligned} d_{12} \\ d_{22} \\ d_{32} \end{aligned}$$

Treatment T	$O$
Treatment U	$O$
Control	$O$

$$\begin{aligned} d_{T3} \\ d_{U3} \end{aligned}$$

- Dependent effect sizes are ubiquitous in education and social science meta-analyses.
- We have well-developed methods for modeling dependent effect sizes assuming no selection.
- But only very recent developments for investigating selective reporting in databases with dependent effect sizes ([Chen and Pustejovsky 2024](#)).

# Selection models have two parts

- Random effects model for the evidence-generating process (*before* selective reporting):

$$T_{ij} \sim N(\mu, \tau^2 + \sigma_{ij}^2)$$

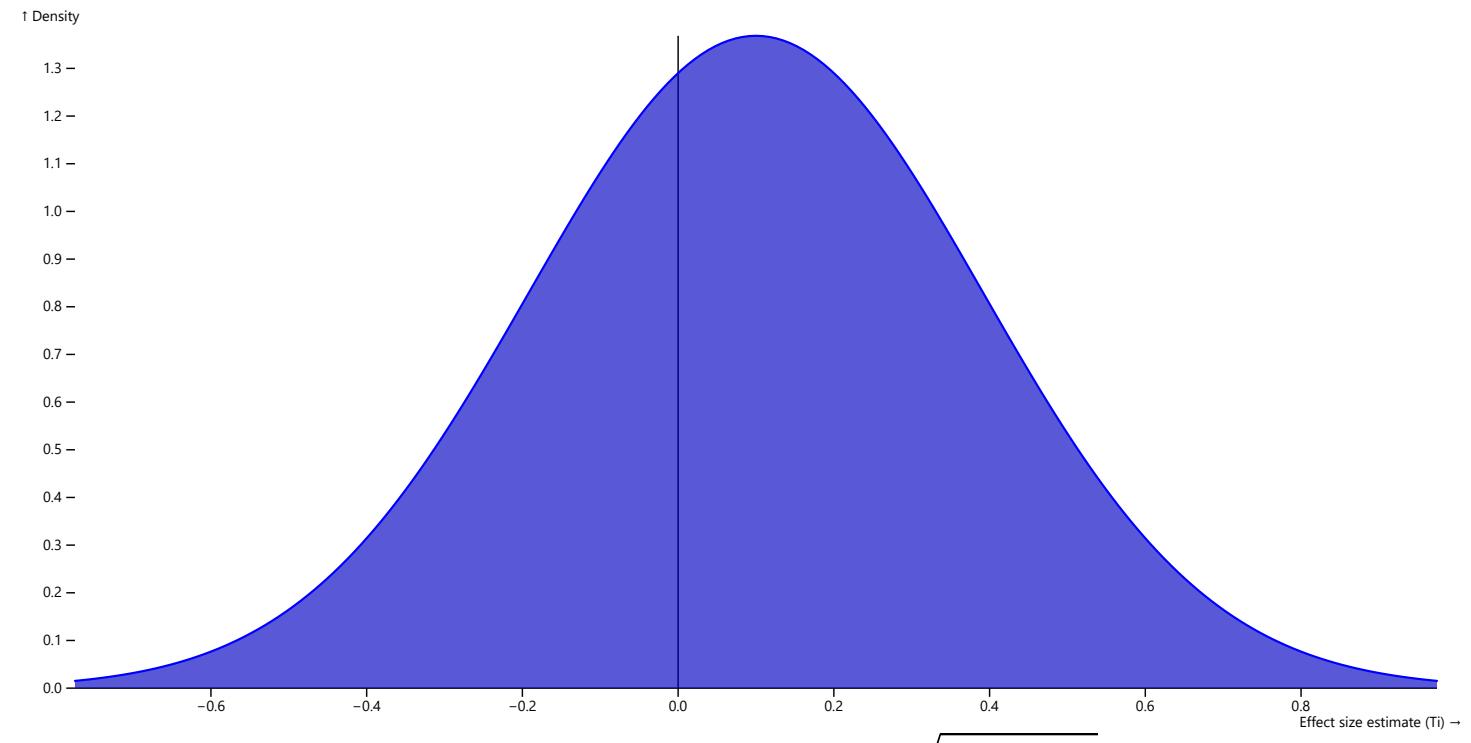
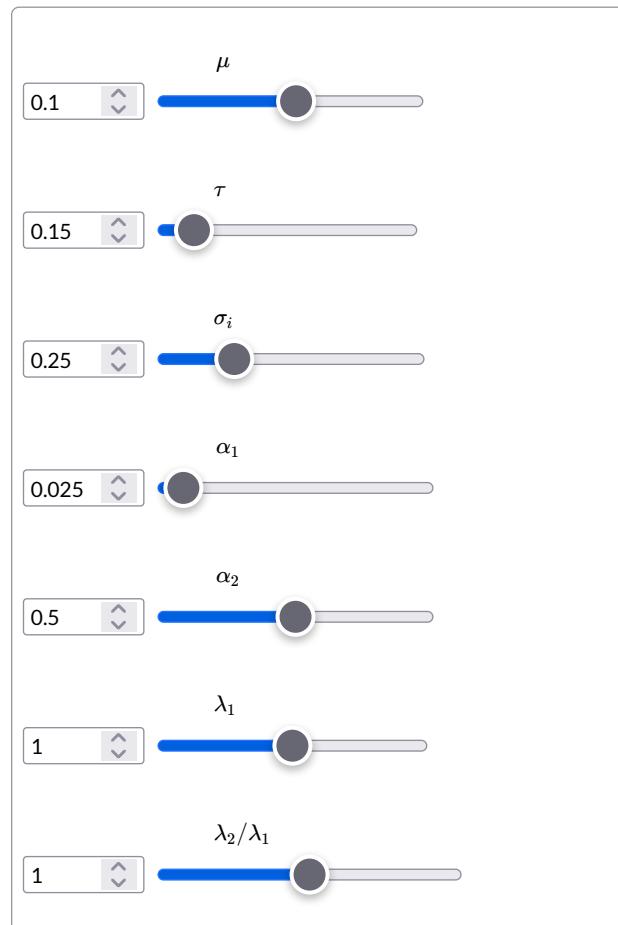
- A model describing  $\Pr(T_{ij} \text{ is observed})$  as a function of its  $p$ -value ( $p_{ij}$ )

Vevea and Hedges (1995) step-function model

Citkowicz and Vevea (2017) beta-function model

# A piece-wise normal distribution

Under Vevea and Hedges (1995) step-function model, the distribution of observed effect size estimates is piece-wise normal.



$$\mu = 0.1$$

$$\mathbb{E}(T_{ij}) = 0.100$$

$$\Pr(T_{ij} \text{ is observed}) = 1.000$$

$$\sqrt{\tau^2 + \sigma_{ij}^2} = 0.292$$

$$\sqrt{\mathbb{V}(T_{ij})} = 0.292$$

# Estimation Strategy

- Model the marginal distribution of observed effects, ignoring the dependence structure
  - Maximum likelihood (composite marginal likelihood)
  - Augmented, reweighted Gaussian likelihood

## Two methods of handling dependence

- Cluster-robust variance estimation  
(sandwich estimators)
- Clustered bootstrap re-sampling



# Color priming

Lehmann, Elliot, and Calin-Jageman (2018) reported a systematic review of studies on **color-priming**, examining whether exposure to the color red influenced attractiveness judgments.

- Many published studies where selective reporting was suspected.

Coef.	Mean ES		Heterogeneity Variance	
	Est.	SE	Est.	SE
<b>(A) Summary meta-analysis</b>				
Overall	0.207	0.0571	0.103	0.0251
<b>(B) Moderation by design type</b>				
Between-Subjects	0.19	0.0642	0.104	0.0256
Within-Subjects	0.273	0.1456	0.104	0.0256

# Color-priming selection models

```
1 library(metaselection)
2
3 # load the data
4 data("dat.lehmann2018", package = "metadat")
5
6 # tidy up
7 dat.lehmann2018$study <- dat.lehmann2018$Full_Citation
8 dat.lehmann2018$sei <- sqrt(dat.lehmann2018$vi)
9 dat.lehmann2018$Design <- factor(dat.lehmann2018$Design, levels = c("Between Subjects", "Within
10
11 # fit a one-step selection model
12 sell1 <- selection_model(
13   yi = yi,                      # effect size est.
14   sei = sei,                     # standard error
15   cluster = study,              # identifier for independent clusters
16   data = dat.lehmann2018,        # dataset
17   selection_type = "step",       # type of selection model
18   steps = .025,                  # single threshold for step-function
19   estimator = "CML",            # estimation method
20   bootstrap = "none"           # large-sample sandwich standard errors
21 )
22
23 summary(sell1)
```

Step Function Model

Call:

```
selection_model(data = dat.lehmann2018, yi = yi, sei = sei, cluster = study,
  selection_type = "step", steps = 0.025, estimator = "CML",
  bootstrap = "none")
```

Number of clusters = 41; Number of effects = 81

Steps: 0.025

Estimator: composite marginal likelihood

Variance estimator: robust

Log composite likelihood of selection model: -44.46436

Inverse selection weighted partial log likelihood: 58.35719

Mean effect estimates:

	Large Sample				
Coef.	Estimate	Std. Error	p-value	Lower	Upper
beta	0.133	0.137	0.333	-0.136	0.402

Heterogeneity estimates:

	Large Sample				
Coef.	Estimate	Std. Error	p-value	Lower	Upper
tau2	0.0811	0.0845	---	0.0105	0.625

Selection process estimates:

Step: 0 < p <= 0.025; Studies: 16; Effects: 25	Large Sample				
Coef.	Estimate	Std. Error	p-value	Lower	Upper
lambda0	1	---	---	---	---

Step: 0.025 < p <= 1; Studies: 29; Effects: 56	Large Sample				
Coef.	Estimate	Std. Error	p-value	Lower	Upper
lambda1	0.548	0.616	0.593	0.0607	4.96

# Now with bootstrapping!

```
1 # turn on parallel processing
2 library(future)
3 plan(multisession, workers = 8)
4
5 set.seed(20250613) # for reproducibility
6
7 sell_boot <- selection_model(
8   yi = yi,                      # effect size est.
9   sei = sei,                      # standard error
10  cluster = study,               # identifier for independent clusters
11  data = dat.lehmann2018,        # dataset
12  selection_type = "step",       # type of selection model
13  steps = .025,                  # single threshold for step-function
14  estimator = "CML",            # estimation method
15  bootstrap = "two-stage",      # recommended type of bootstrapping
16  R = 1999,                      # number of bootstrap re-samples
17  CI_type = c("large-sample",    # keep the large-sample sandwich CI
18           "percentile")        # recommended type of bootstrap CI
19 )
20
21 summary(sell_boot)
```

Step Function Model with Cluster Bootstrapping

Call:

```
selection_model(data = dat.lehmann2018, yi = yi, sei = sei, cluster = study,
  selection_type = "step", steps = 0.025, estimator = "CML",
  CI_type = c("large-sample", "percentile"), bootstrap = "two-stage",
  R = 1999)
```

Number of clusters = 41; Number of effects = 81

Steps: 0.025

Estimator: composite marginal likelihood

Variance estimator: robust

Bootstrap type: two-stage

Number of bootstrap replications: 1999

Log composite likelihood of selection model: -44.46436

Inverse selection weighted partial log likelihood: 58.35719

Mean effect estimates:

Coef.	Estimate	Std. Error	p-value	Large Sample		Percentile		Bootstrap	
				Lower	Upper	Lower	Upper	Lower	Upper
beta	0.133	0.137	0.333	-0.136	0.402	-0.0174	0.435		

Heterogeneity estimates:

Coef.	Estimate	Std. Error	p-value	Large Sample		Percentile		Bootstrap	
				Lower	Upper	Lower	Upper	Lower	Upper
tau2	0.0811	0.0845	---	0.0105	0.625	1.73e-17	0.238		

Selection process estimates:

Step: 0 < p <= 0.025; Studies: 16; Effects: 25

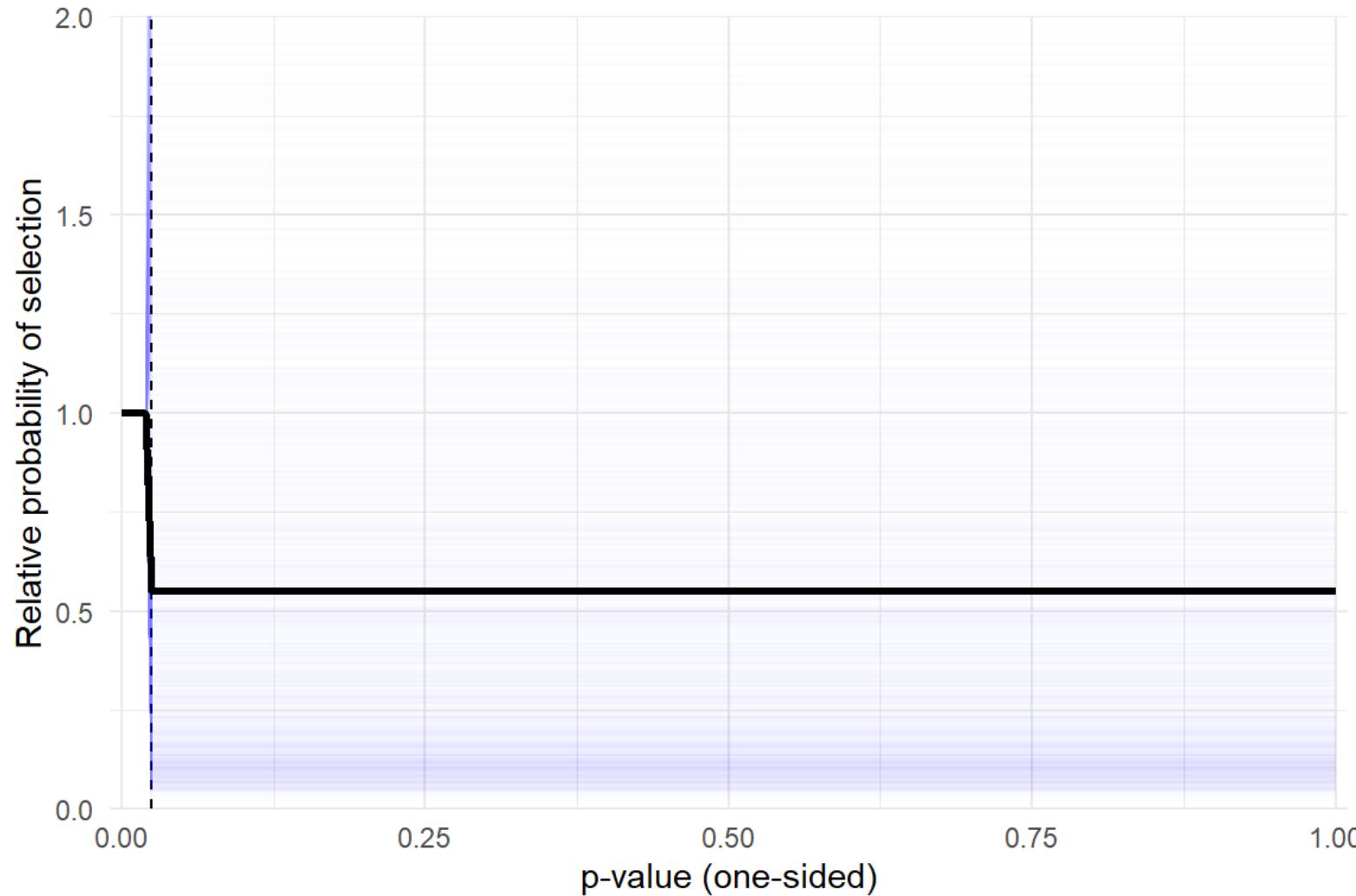
Coef.	Estimate	Std. Error	p-value	Large Sample		Percentile		Bootstrap	
				Lower	Upper	Lower	Upper	Lower	Upper
lambda0	1	---	---	---	---	---	---	---	---

Step: 0.025 < p <= 1; Studies: 29; Effects: 56

Coef.	Estimate	Std. Error	p-value	Large Sample		Percentile		Bootstrap	
				Lower	Upper	Lower	Upper	Lower	Upper
lambda1	0.548	0.616	0.593	0.0607	4.96	0.0537	2.88		

# Selective reporting of non-significant results

```
1 selection_plot(sel1_boot) +  
2 ggplot2::coord_cartesian(ylim = c(0,2))
```



# Add a moderator

```
1 # turn on parallel processing
2 library(future)
3 plan(multisession, workers = 8)
4
5 set.seed(20250613) # for reproducibility
6
7 sell_mod <- selection_model(
8   yi = yi,                      # effect size est.
9   sei = sei,                     # standard error
10  cluster = study,              # identifier for independent clusters
11  mean_mods = ~ 0 + Design,    # design type moderator
12  data = dat.lehmann2018,       # dataset
13  selection_type = "step",     # type of selection model
14  steps = .025,                 # single threshold for step-function
15  estimator = "CML",           # estimation method
16  bootstrap = "two-stage",     # recommended type of bootstrapping
17  R = 1999,                     # number of bootstrap re-samples
18  CI_type = c("large-sample",  # keep the large-sample sandwich CI
19    "percentile")             # recommended type of bootstrap CI
20 )
```

```
1 summary(sell_mod)
```

Step Function Model with Cluster Bootstrapping

Call:

```
selection_model(data = dat.lehmann2018, yi = yi, sei = sei, cluster = study,
  selection_type = "step", steps = 0.025, mean_mods = ~0 +
  Design, estimator = "CML", CI_type = c("large-sample",
  "percentile"), bootstrap = "two-stage", R = 1999)
```

Number of clusters = 41; Number of effects = 81

Steps: 0.025

Estimator: composite marginal likelihood

Variance estimator: robust

Bootstrap type: two-stage

Number of bootstrap replications: 1990

Log composite likelihood of selection model: -44.12226

Inverse selection weighted partial log likelihood: 61.14273

Mean effect estimates:

	Coef.	Estimate	Std. Error	p-value	Large	Sample	Percentile	Bootstrap
					Lower	Upper	Lower	Upper
beta_DesignBetween		0.113	0.117	0.333	-0.116	0.343	-0.0484	0.339
beta_DesignWithin		0.196	0.234	0.400	-0.261	0.654	0.0104	0.985

Heterogeneity estimates:

	Coef.	Estimate	Std. Error	p-value	Large	Sample	Percentile	Bootstrap
					Lower	Upper	Lower	Upper
tau2		0.0785	0.081	---	0.0104	0.593	1.14e-17	0.197

Selection process estimates:

Step: 0 < p <= 0.025; Studies: 16; Effects: 25

	Coef.	Estimate	Std. Error	p-value	Large	Sample	Percentile	Bootstrap
					Lower	Upper	Lower	Upper
lambda0		1	---	---	---	---	---	---

Step: 0.025 < p <= 1; Studies: 29; Effects: 56

	Coef.	Estimate	Std. Error	p-value	Large	Sample	Percentile	Bootstrap
					Lower	Upper	Lower	Upper
lambda1		0.533	0.601	0.577	0.0584	4.86	0.042	2.66

# Discussion

- Other supported models:
  - Step-functions with multiple steps (e.g.,  $\alpha_1 = .025, \alpha_2 = .500$ )
  - beta-function models ([Citkowicz and Vevea 2017](#))
  - Location-scale meta-regression ([Viechtbauer and López-López 2022](#))
  - Predictors of selection ([Coburn and Vevea 2015](#))
- Marginal step-function selection models are worth adding to the toolbox ([Pustejovsky, Citkowicz, and Joshi 2025](#)).
  - Low bias compared to other selective reporting adjustments (including PET-PEESE)
  - Bias-variance trade-off relative to regular meta-analytic models
  - Two-stage clustered bootstrap percentile confidence intervals work tolerably well

# R package metaselection

- Currently available on Github at <https://github.com/jepusto/metaselection>
- Install using

```
1 remotes::install_github("jepusto/metaselection", build_vignettes = TRUE)
```

- Under active development, suggestions welcome!

# References

- Augusteijn, Hilde E. M., Robbie C. M. van Aert, and Marcel A. L. M. van Assen. 2019. "The Effect of Publication Bias on the Q Test and Assessment of Heterogeneity." *Psychological Methods* 24 (1): 116–34. <https://doi.org/10.1037/met0000197>.
- Chen, Man, and James E. Pustejovsky. 2024. "Adapting Methods for Correcting Selective Reporting Bias in Meta-Analysis of Dependent Effect Sizes." <https://doi.org/10.31222/osf.io/jq52s>.
- Citkowicz, Martyna, and Jack L Vevea. 2017. "A parsimonious weight function for modeling publication bias." *Psychological Methods* 22 (1): 28–41. <https://doi.org/10.1037/met0000119>.
- Coburn, Kathleen M, and Jack L Vevea. 2015. "Publication bias as a function of study characteristics." *Psychological Methods* 20 (3): 310–30. <https://doi.org/10.1037/met0000046>.
- Lehmann, Gabrielle K, Andrew J Elliot, and Robert J Calin-Jageman. 2018. "Meta-Analysis of the Effect of Red on Perceived Attractiveness." *Evolutionary Psychology* 16 (4): 1474704918802412.
- Pustejovsky, James E., Martyna Citkowicz, and Megha Joshi. 2025. "Estimation and Inference for Step-Function Selection Models in Meta-Analysis with Dependent Effects." [https://doi.org/10.31222/osf.io/qg5x6\\_v1](https://doi.org/10.31222/osf.io/qg5x6_v1).
- Vevea, Jack L, and Larry V Hedges. 1995. "A General Linear Model for Estimating Effect Size in the Presence of Publication Bias." *Psychometrika* 60 (3): 419–35. <https://doi.org/10.1007/BF02294384>.
- Viechtbauer, Wolfgang, and José Antonio López-López. 2022. "Location-scale Models for Meta-analysis." *Research Synthesis Methods*, April, jrsm.1562. <https://doi.org/10.1002/jrsm.1562>.