Reliability Generalization Meta-Analysis:

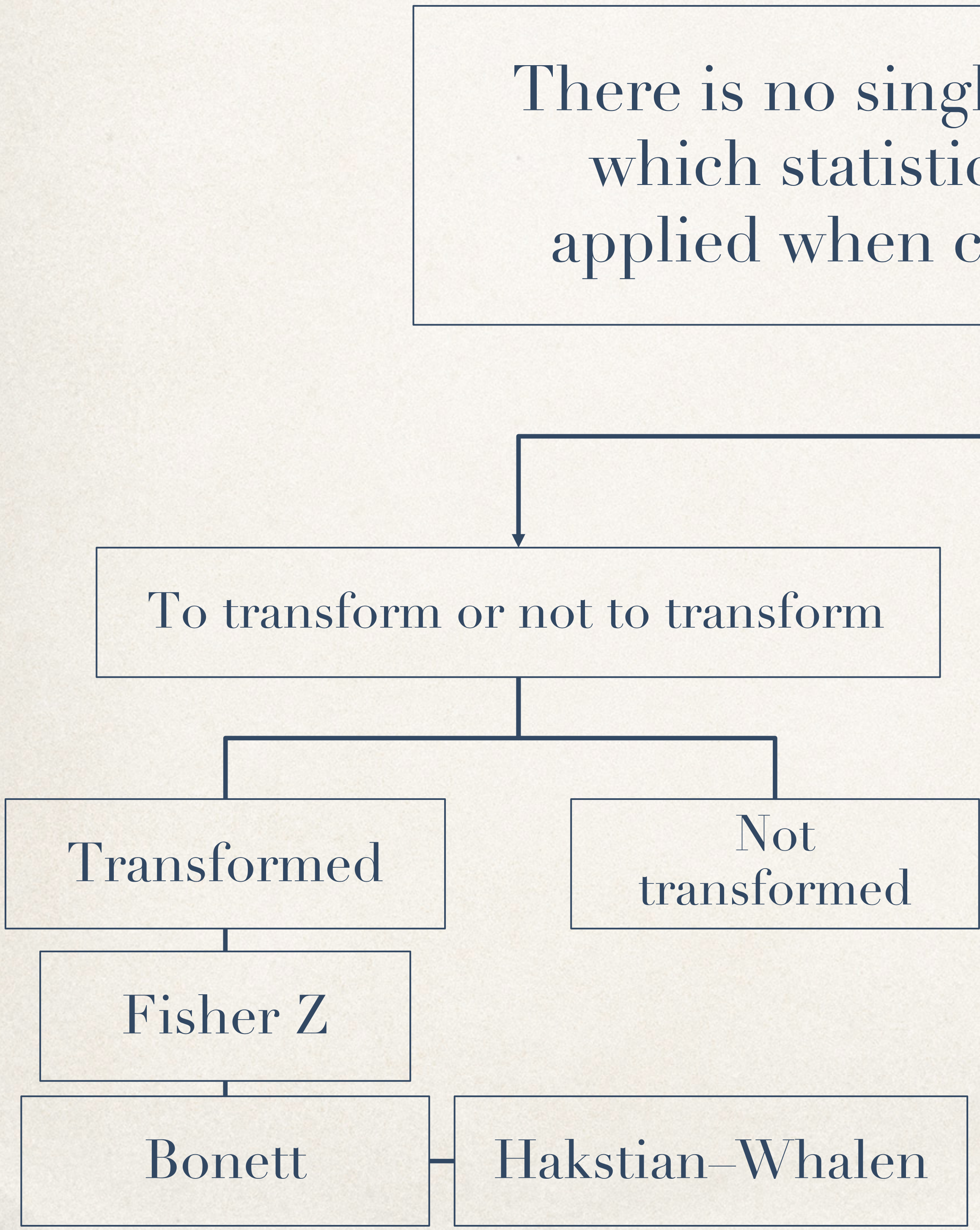
A comparison of statistical analytic strategies

Carmen López-Ibáñez, Desirée Blázquez-Rincón & Julio Sánchez-Meca

Department Basic Psychology and Methodology
University of Murcia, Spain.

This research was funded by a grant from the Ministerio de Economía y Competitividad of the Spanish Government and Fondo Europeo de Desarrollo Regional (FEDER) (Project No. PSI2016-77676-P)

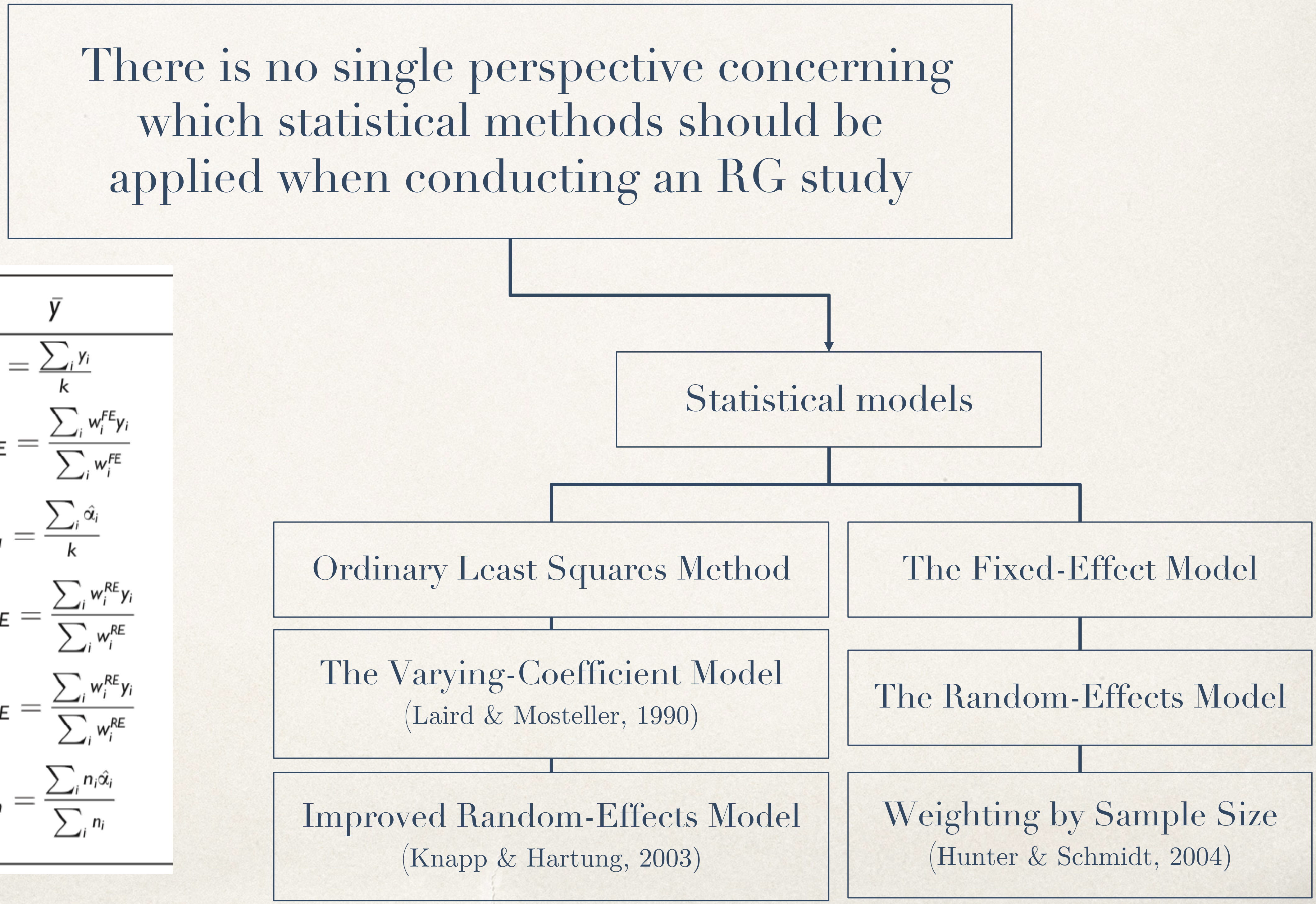
Background: RG Meta-Analysis



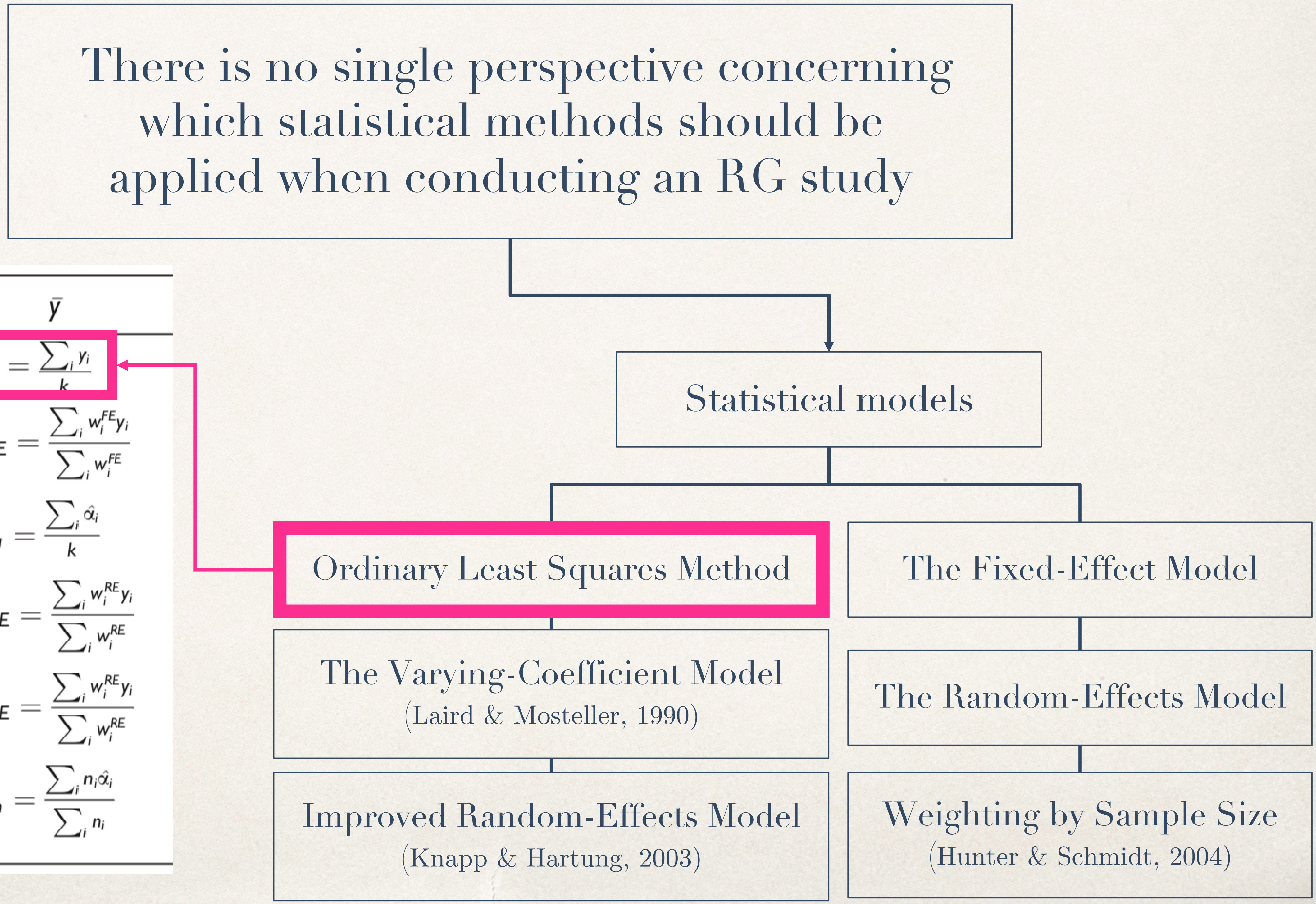
Transformation	Coefficient	Back-transformation
Not transformed	$\hat{\alpha}_i$	–
Fisher z	$Z_i = \frac{1}{2}\ln\left(\frac{1+\hat{\alpha}_i}{1-\hat{\alpha}_i}\right)$	$\hat{\alpha}_i = \frac{e^{2Z_i}-1}{e^{2Z_i}+1}$
Hakstian–Whalen	$T_i = \sqrt[3]{1-\hat{\alpha}_i}$	$\hat{\alpha}_i = 1 - T_i^3$
Bonett	$L_i = \ln(1-\lvert\hat{\alpha}_i\rvert)$	$\hat{\alpha}_i = 1 - e^{L_i}$

Background: RG Meta-Analysis

Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$



Background: RG Meta-Analysis



Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Background: RG Meta-Analysis

There is no single perspective concerning which statistical methods should be applied when conducting an RG study

Statistical models

Ordinary Least Squares Method

The Fixed-Effect Model

The Varying-Coefficient Model
(Laird & Mosteller, 1990)

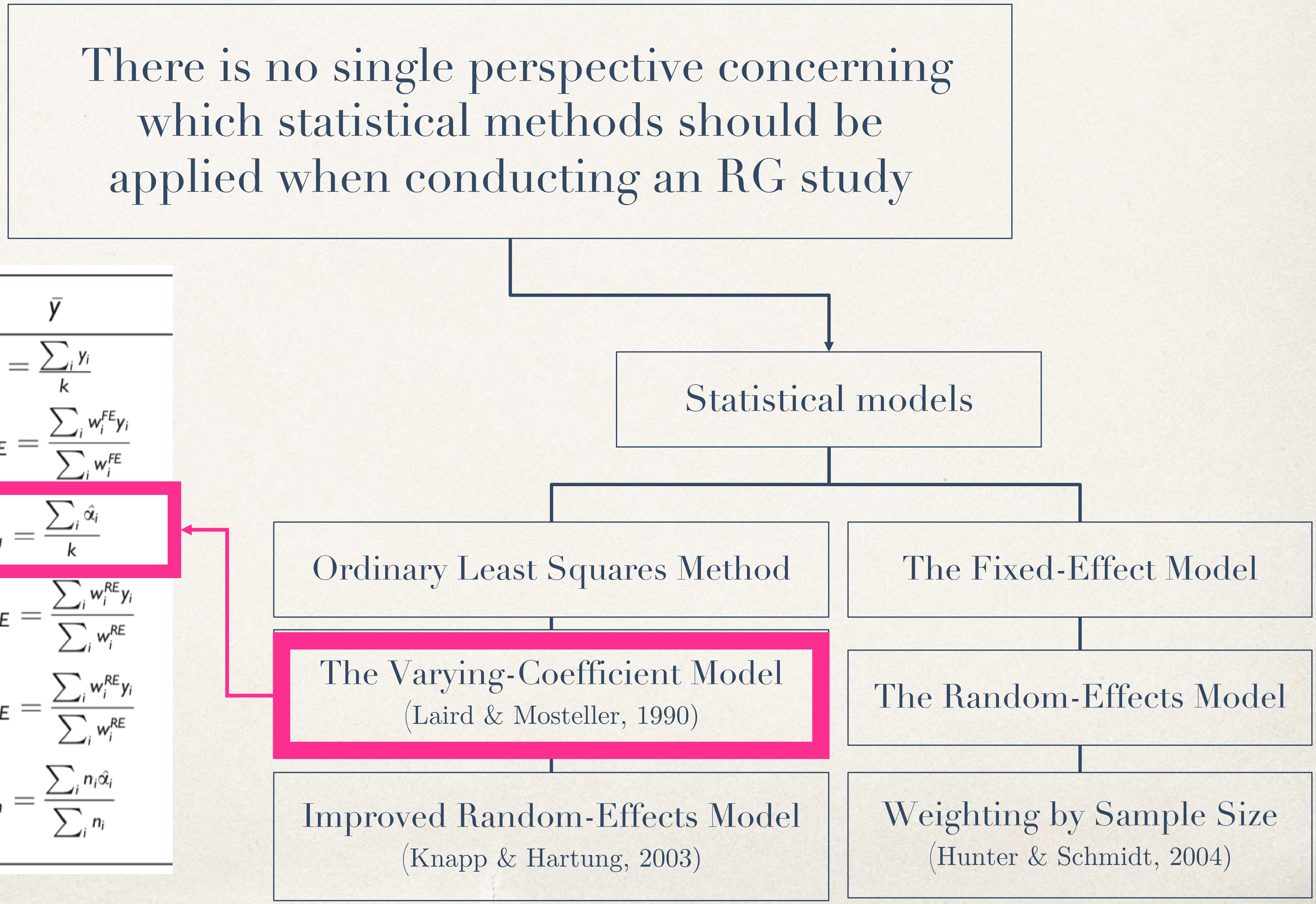
The Random-Effects Model

Improved Random-Effects Model
(Knapp & Hartung, 2003)

Weighting by Sample Size
(Hunter & Schmidt, 2004)

Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Background: RG Meta-Analysis



Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Background: RG Meta-Analysis

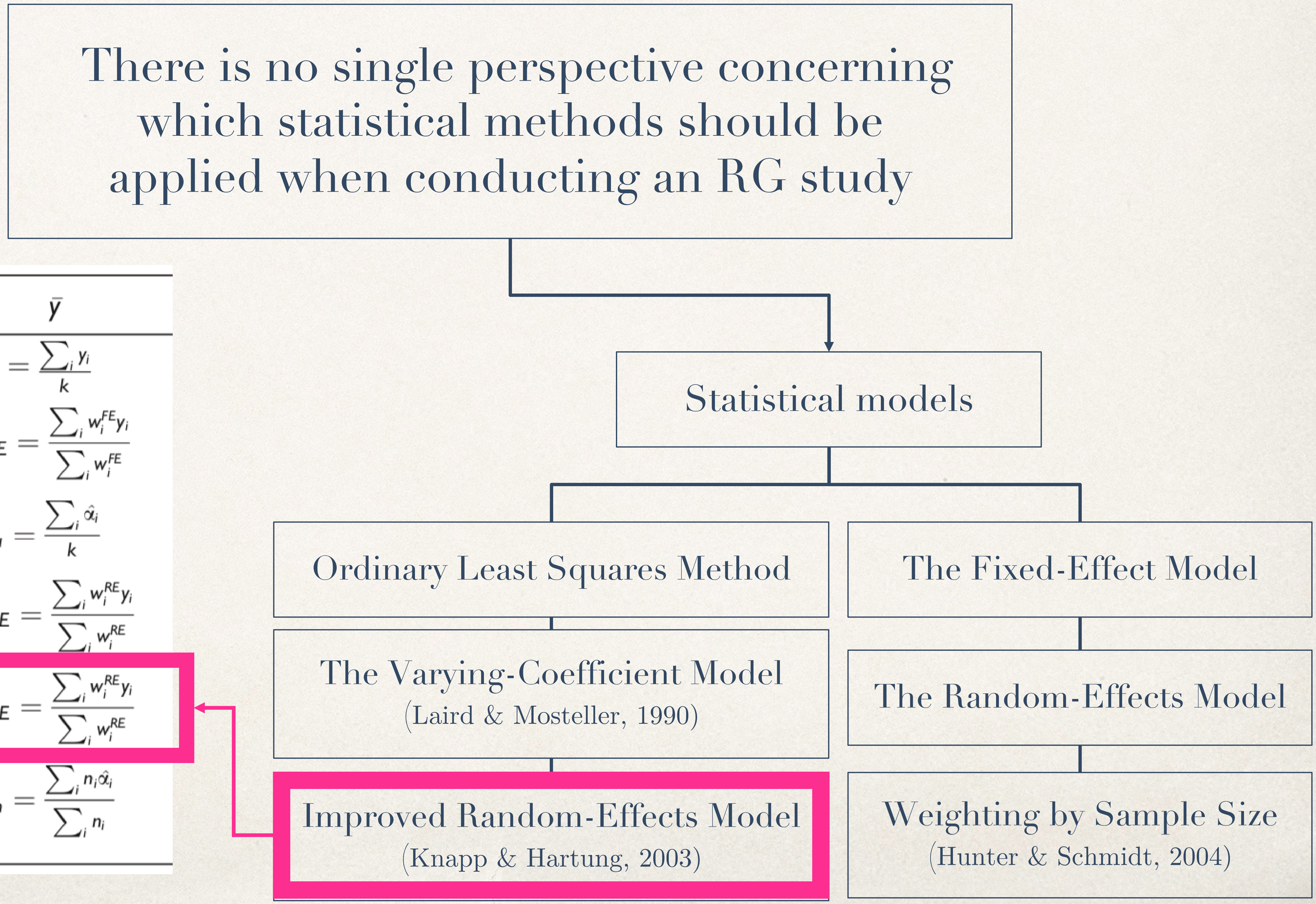
There is no single perspective concerning which statistical methods should be applied when conducting an RG study

Statistical models

- Ordinary Least Squares Method
- The Fixed-Effect Model
- The Varying-Coefficient Model (Laird & Mosteller, 1990)
- The Random-Effects Model
- Improved Random-Effects Model (Knapp & Hartung, 2003)
- Weighting by Sample Size (Hunter & Schmidt, 2004)

Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Background: RG Meta-Analysis



Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Background: RG Meta-Analysis

There is no single perspective concerning which statistical methods should be applied when conducting an RG study

Statistical models

Ordinary Least Squares Method

The Fixed-Effect Model

The Varying-Coefficient Model
(Laird & Mosteller, 1990)

The Random-Effects Model

Improved Random-Effects Model
(Knapp & Hartung, 2003)

Weighting by Sample Size
(Hunter & Schmidt, 2004)

Model	$\bar{y}$
OLS	$\bar{y}_u = \frac{\sum_i y_i}{k}$
FE	$\bar{y}_{FE} = \frac{\sum_i w_i^{FE} y_i}{\sum_i w_i^{FE}}$
VC	$\bar{\alpha}_u = \frac{\sum_i \hat{\alpha}_i}{k}$
RE	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-c	$\bar{y}_{RE} = \frac{\sum_i w_i^{RE} y_i}{\sum_i w_i^{RE}}$
RE-n	$\bar{\alpha}_n = \frac{\sum_i n_i \hat{\alpha}_i}{\sum_i n_i}$

Objective

Different methods could involve different conclusions depending on the selected statistical model.

⬇

The main objective is to assess how different statistical methods employed can imply different results, both reliability coefficient and their corresponding confidence intervals, when they are applied to real RG meta-analyses.

Method: Selection Criteria

1. The study had to be written in **English or Spanish**.
2. The research had to present an **RG meta-analysis**.
3. The RG meta-analysis had to focus on **one or several measures** of a psychological construct.
4. The RG MA should report **alpha coefficients**.
5. The RG MA must provide **the whole database** with the individual alpha coefficients.

Method: Search Strategies

1. **Scopus & Google Scholar** databases.
2. Keywords: **“Reliability Generalization”**, **“Meta-Analysis of Internal Consistency”**, **“Meta-Analysis of Alpha Coefficients”**
3. Temporal Range: **1998-December** 2018.
4. Language Restrictions: studies written in English or Spanish.

Method: Data Extraction

1. Coefficients Alpha
2. Number of Items
3. Sample Size
4. Mean & Standard Deviation of test scores
5. If the article used several scales or subscales, we took them as independent scales.

Method: Statistical Analysis

To compare the different results, we established two comparison criteria

Differences between the **average alpha** values obtained with the different procedures

13 different statistical models

Repeated Measures ANOVA

Differences between the width of the **confidence interval** around the average reliability coefficient.

18 different statistical models

Repeated Measures ANOVA

Software: R v3.5.2, SPSS v24 & JASP v0.9.2

Method: Statistical Analysis

Another comparison criterion was the degree of heterogeneity found between the different transformation methods

The value of I^2 was calculated through the different transformations methods

4 different statistical models

Results: selection process

We obtained **29 studies** with one or several scales or subscales with individual alpha coefficients.

From this studies, we extracted **110 scales or subscales** with individual alpha coefficients.

3552 alpha coefficients completed this databases

Results: Average Alpha Coefficient

$F(12, 1308) = 67.861$ $p < .000$ $\eta^2 = .384$

Table 1. Average alpha coefficient

	OLS	FE	RE	REn
NT	$\alpha = .823$	$\alpha = .868$	$\alpha = .833$	$\alpha = .829$
Z	$\alpha = .823$	$\alpha = .829$	$\alpha = .827$	-
HW	$\alpha = .823$	$\alpha = .844$	$\alpha = .826$	-
B	$\alpha = .823$	$\alpha = .829$	$\alpha = .828$	-
Global Average	$\alpha = .831$			

OLS: Ordinary Least Squares, FE: Fixed-Effect model, RE: Random-Effects model, REn: Random-Effect model weighted by simple size

NT: Untransformed Method, Z: Fisher Z-Transformation, HW: Hakstian-Whalen Transformation, B: Bonett Transformation

Results: Average Alpha Coefficient

$F(12, 1308) = 67.861$ $p < .000$ $\eta^2 = .384$

		OLS				FE				RE			
		NT α = .823	**Z** α = .823	**HW** α = .823	**B** α = .823	**NT** α = .868	**Z** α = .829	**HW** α = .844	**B** α = .829	**NT** α = .833	**Z** α = .827	**HW** α = .826	**B** α = .828
OLS	**NT** α = .823												
	Z α = .823	–											
	HW α = .823	–	–										
	B α = .823	–	–	–									
FE	**NT** α = .868	*	*	*	*								
	Z α = .829	–	–	–	–	*							
	HW α = .844	*	*	*	*	*	*						
	B α = .829	–	–	–	–	*	–	*					
RE	**NT** α = .833	*	*	*	*	*	–	*	–				
	Z α = .827	–	–	–	–	*	–	*	–	*			
	HW α = .826	*	*	*	*	*	–	*	–	*	–		
	B α = .828	–	–	–	–	*	–	*	–	–	–	–	
REn	α = .829	–	–	–	–	*	–	*	–	–	–	–	–

Legend

- * Statistical Differences
- – No Statistical Differences

OLS: Ordinary Least Squares, FE: Fixed-Effect model, RE: Random-Effects model, REn: Random-Effect model weighted by simple size

NT: Untransformed Method, Z: Fisher Z-Transformation, HW: Hakstian-Whalen Transformation, B: Bonett Transformation

Results: Average Alpha Coefficient

The main statistically significant results

- Between **Fixed-Effect model** with **NT** (.868) & **HW** (.844) and all the other methods.

- Between **Random-Effects model** with **NT** (.833) and OLS model (.823), and RE model with Z (.827) & HW (.826).

- Between **Random-Effects model** with **HW** (.826) and OLS model (.823).

Results: Confidence Width

F(17, 1853)= 76.893 p<.000 η^2=.414

Table 2. Average confidence width

	OLS	FE	RE	REn	REc	VC
NT	CW= .0889	CW= .0123	CW= .0575	CW= .0612	CW= .0799	-
Z	CW= .0889	CW= .056	CW= .0801	-	CW= .0825	-
HW	CW= .0889	CW= .0148	CW= .0662	-	CW= .0863	-
B	CW= .0889	CW= .0845	CW=.0958	-	CW= .08	CW= .1425
Global Average			CW= .0753			

OLS: Ordinary Least Squares model, FE: Fixed-Effect model, RE: Random-Effect model, REn: Random-Effect model Weighted by Sample Size, REc: Improve Random-Effect model, VC: Varying-Coefficient model

NT: Untransformed Method, Z: Fisher Z-Transformation, HW: Hakstian-Whalen Transformation, B: Bonett Transformation

Results: Confidence Width

$F(17, 1853) = 76.893$ $p<.000$ $\eta^2=.414$

		OLS				FE				RE				REn	REc			
		NT .0889	Z .0889	HW .0889	B .0889	NT .0123	Z .056	HW .0148	B .0845	NT .0575	Z .0801	HW .0662	B .0958	.0612	NT .0799	Z .0825	HW .0863	B .08
OLS	NT .0889																	
	Z .0889	–																
	HW .0889	–	–															
	B .0889	–	–	–														
FE	NT .0123	*	*	*	*													
	Z .056	*	*	*	*	*												
	HW .0148	*	*	*	*	*	*											
	B .0845	–	–	–	–	*	*	*										
RE	NT .0575	*	*	*	*	*	–	*	*									
	Z .0801	–	–	–	–	*	*	*	–	*								
	HW .0662	*	*	*	*	*	–	*	*	*	*							
	B .0958	–	–	–	–	*	*	*	–	*	*	*						
REn	NT .0612	*	*	*	*	*	–	*	*	–	*	–	*					
REc	NT .0799	*	*	*	*	*	*	*	–	*	–	*	*	*				
	Z .0825	*	*	*	*	*	*	*	–	*	–	*	–	*	–			
	HW .0863	*	*	*	*	*	*	*	–	*	–	*	–	*	*	*		
	B .08	*	*	*	*	*	*	*	–	*	–	*	*	*	–	–	*	
VC	B .1425	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Legend

* Statistical Differences

– No Statistical Differences

OLS: Ordinary Least Squares model, FE: Fixed-Effect model, RE: Random-Effect model, REn: Random-Effect model Weighted by Sample Size, REc: Improve Random-Effect model, VC: Varying-Coefficient model

NT: Untransformed Method, Z: Fisher Z-Transformation, HW: Hakstian-Whalen Transformation, B: Bonett Transformation

Results: Confidence width

The main statistically significant results

- Between **Fixed-Effect model** with **NT** (.0123) & **HW** (.0148) and all the other methods.

- Between **Varying-Coefficient model** with **B** (.1425) and all the other methods.

- Between **Improved Random-Effects model** with **NT** (.0799), **Z** (.0825), **HW** (.0863), & **B** (.08) and almost the other methods **except**:
    - FE model with B (.0845), RE model with Z (.0801).

Results: Heterogeneity

Table 3. Descriptive Statistics of I^2

	NT	**Z**	**HW**	**B**
Mean	**91.43**	26.56	**89.55**	8.589
SD	8.986	30.42	10.47	19.21
Min	38.18	.000	28.90	.000
Max	99.45	94.80	99.25	84.45

In all cases, both when the scores were not transformed and when the scores were transformed by HW method, we found the highest values of I^2.

Conclusions

Average reliability coefficient

Statistically significant differences appear between different methods (Repeated Measures ANOVA)

Conclusions

Average reliability coefficient	Statistically significant differences appear between different methods (Repeated Measures ANOVA)
Width of the confidence interval	Statistically significant differences appear between different methods (Repeated Measures ANOVA)

Conclusions

Average reliability coefficient

Width of the confidence interval

As a consequence, the results can change drastically depends on the transformation methods and the statistical model assumes.

Conclusion

The choice of both the statistical model and the transformation of the scores is a decision of the researcher.

One way to decide the model to use is through the heterogeneity of the alpha coefficients. If we assume the presence of heterogeneity, the best option is to choose a random effects model.

Our recommendation is to use Bonett Transformation since it normalizes the distribution of alpha coefficients and stabilizes their variances.

Thank you for your attention

carmen.li@um.es