Recommendations for Reviewers, Editors, and Authors in Systematic Reviews and Meta-Analyses

Ernest H. O'Boyle
Justin A. DeSimone
Michael T. Brannick
Ji Woon (June) Ryu
February 1, 2019

Need for Transparency

- Replication crisis in social science (Hunter, 2001; Ioannidis, 2005a; Ioannidis, 2005b)
- Call for transparency across all areas of scholarship (Christensen & Miguel, 2017; Nosek et al., 2015)
- Authors focused on research synthesis and meta-analysis are among the greatest proponents of transparent reporting (APA; 2008; Aytug et al., 2012; Moher et al., 2009)

Transparency in Meta-analysis

- Judgment calls (Aguinis et al., 2011; Geyskins et al., 2009; Wanous et al., 1989)

- Lack of transparency in reporting judgment calls can lead to difficulty in interpreting meta-analytic results (Carlson & Ji, 2011; DeSimone et al., in press)

- A lack of transparency can lead to confusing or conflicting results (Bobko & Stone-Romero, 1998; Ioannidis, 2016; Van Iddekinge et al., 2012)

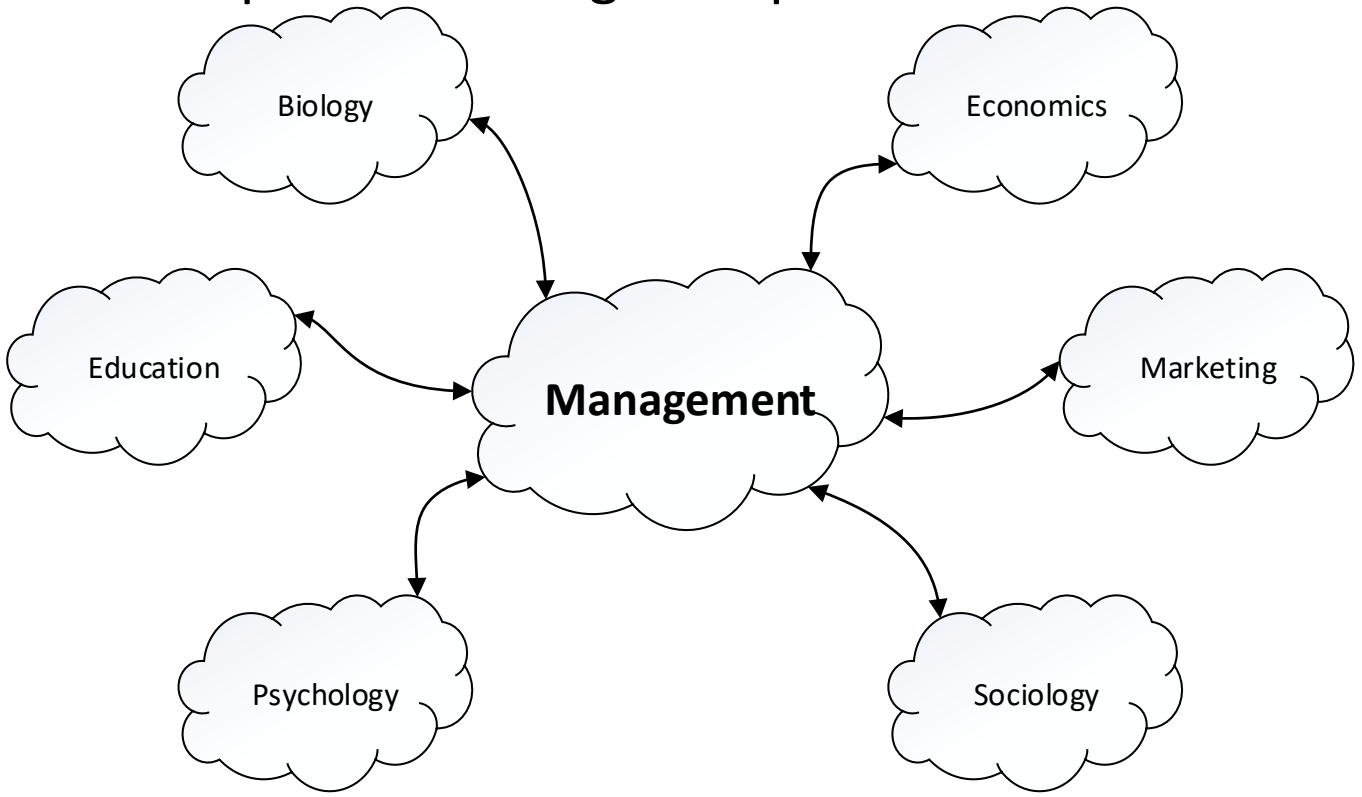
This Presentation

- Identify best practices and areas for improvement in reporting meta-analytic methodology and results
  - Focus on areas that are particularly relevant or unique to the field of management
- Organized into three sections
  - Search and Coding
  - Data
  - Analysis
- Encourage authors, editors, reviewers, and readers to follow and demand transparent reporting practices in systematic reviews and meta-analyses

Reporting Issues in Search and Coding

- Management is multidisciplinary by nature
  - Management borrows ideas from a diverse set of disciplines
  - Meta-analyses in management often require searching multiple databases

- Transparent documentation:
  - Construct definition and scope
  - Databases
  - Keywords (and combinations)
  - Intentional omissions

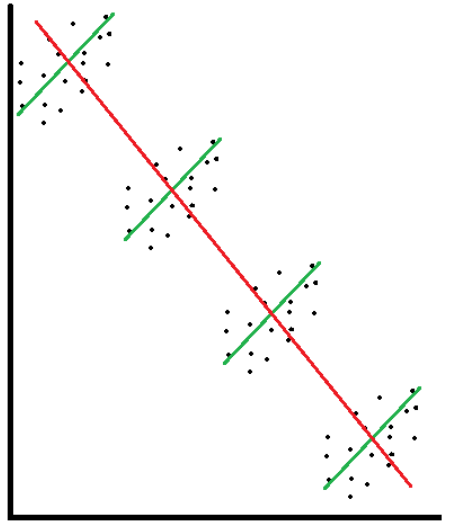


Construct Clarity

- Context plays an important role in identifying boundary conditions for many management theories

- Some constructs in management exist at different levels
  - They may be conceptualized and operationalized in different ways
  - Clear definitions of constructs and inclusion/exclusion criteria can help avoid confusion or misinterpretation (e.g., ecological fallacy)

- Consider context when reporting inclusion criteria
  - Anticipate potential misperception by readers



Inclusion and Exclusion Criteria

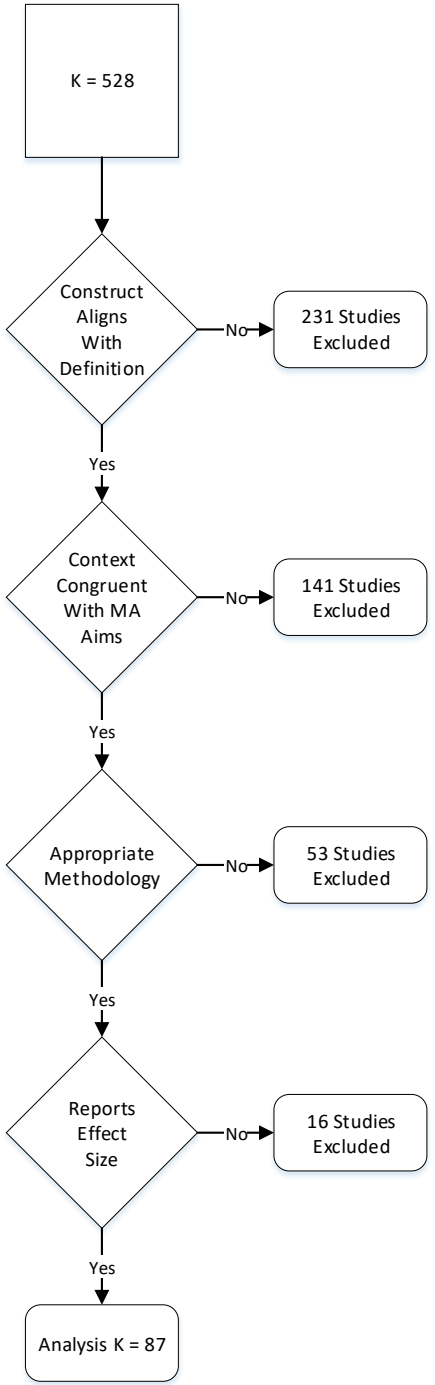
- Existing reporting standards and best practices (e.g., MARS, PRISMA) ask meta-analysts to carefully detail decision rules and inclusion/exclusion criteria

- Most meta-analyses in management use the Hunter & Schmidt procedure (Aguinis et al., 2011)
  - Favors inclusion over scrutiny of primary study quality

- Document efforts to find unpublished studies, *a priori* quality standards (Slavin, 1986; 1995), and decision rules or cutoffs

Specify Decision Rules

- Construct definitions and constraints
- Inclusion/exclusion criteria
- Order of application
- Limits, cutoffs, and boundary conditions



Meta-Analysis Reporting Standards - Data

CARMA Feb 2019

Data Distribution – Report the Data

- Transparency
  - Error checking
- New methods
- Updates (new meta)

Appendix

Detailed Information on Primary Studies Included in Meta-Analysis

Study	Outcome	*N*	*r*	r_{xx}	r_{yy}	Balance measure	Cluster	Single vs. multi-item
Abendroth & den Dulk (2011)	LS	5,904	.43	.89		Valcour (2007)	1	Multi-item
Ali Omran (2016)	LS	276	.54	.85	.92	Carlson et al. (2009)	4	Multi-item
Aryee et al. (2005)	JS	267	−.26	.75	.82	Grzywacz and Marks (2000) - WFC	2	Multi-item
Aryee et al. (2005)	JS	267	.28	.75	.82	Grzywacz and Marks (2000) - WFE	2	Multi-item
Aryee et al. (2005)	JS	267	−.21	.76	.82	Grzywacz and Marks (2000) - FWC	2	Multi-item
Aryee et al. (2005)	JS	267	.25	.73	.82	Grzywacz and Marks (2000) - FWE	2	Multi-item
Aziz et al. (2010)	JS	199	.01	.93	.88	Fisher et al. (2009)	2	Multi-item
Bornstein et al. (2003)	LS	234	.35	.70		Macphee et al. (1986)	4	Multi-item
Bryant & Constantine (2006)	JS	133	.20	.76	.69	Marks & MacDermid (1996)	3	Multi-item
Bryant & Constantine (2006)	LS	133	.45	.76	.91	Marks & MacDermid (1996)	3	Multi-item
Cahill et al. (2015)	JS	2,495	.40	.97		Valcour (2007)	1	Multi-item
Carlson et al. (2009)	FS	685	.52	.93	.94	Carlson et al. (2009)	4	Multi-item
Carlson et al. (2009)	JS	685	.62	.93	.93	Carlson et al. (2009)	4	Multi-item
Carvalho & Chambel (2016)	LS	218	−.16	.88	.88	Carlson et al. (2009) - WFC	2	Multi-item
Carvalho & Chambel (2016)	LS	218	.27	.96	.88	Carlson et al. (2009) - WFE	2	Multi-item
Chan et al. (2016)	FS	234	.29	.94	.97	Brough et al. (2014)	3	Multi-item
Chan et al. (2016)	JS	234	.27	.94	.86	Brough et al. (2014)	3	Multi-item
Chen & Li (2012) Sample 1	FS	204	.37	.72	.95	Marks & MacDermid (1996)	3	Multi-item
Chen & Li (2012) Sample 2	FS	204	.47	.71	.96	Marks & MacDermid (1996)	3	Multi-item
Clark (2000)	JS	179	−.10	.86	.91	Bohen & Viveros-Long (1981) - WFC & FWC	6	Multi-item
Clark (2000)	LS	179	−.15	.86	.85	Bohen & Viveros-Long (1981) - WFC & FWC	6	Multi-item
Clarke et al. (2004)	JS	387	.22[b]			Milkie & Peltola (1999)	4	Single-item
Clarke et al. (2004)	FS	387	.12[a]			Milkie & Peltola (1999)	4	Single-item
DeHauw (2014) Sample 1	JS	395	.43	.93	.88	Own measure	4	Multi-item

Casper et al. (2018). The jingle-jangle of work-nonwork balance: A comprehensive and meta-analytic review of its meaning and measurement. *Journal of Applied Psychology, 103*, 182-214.

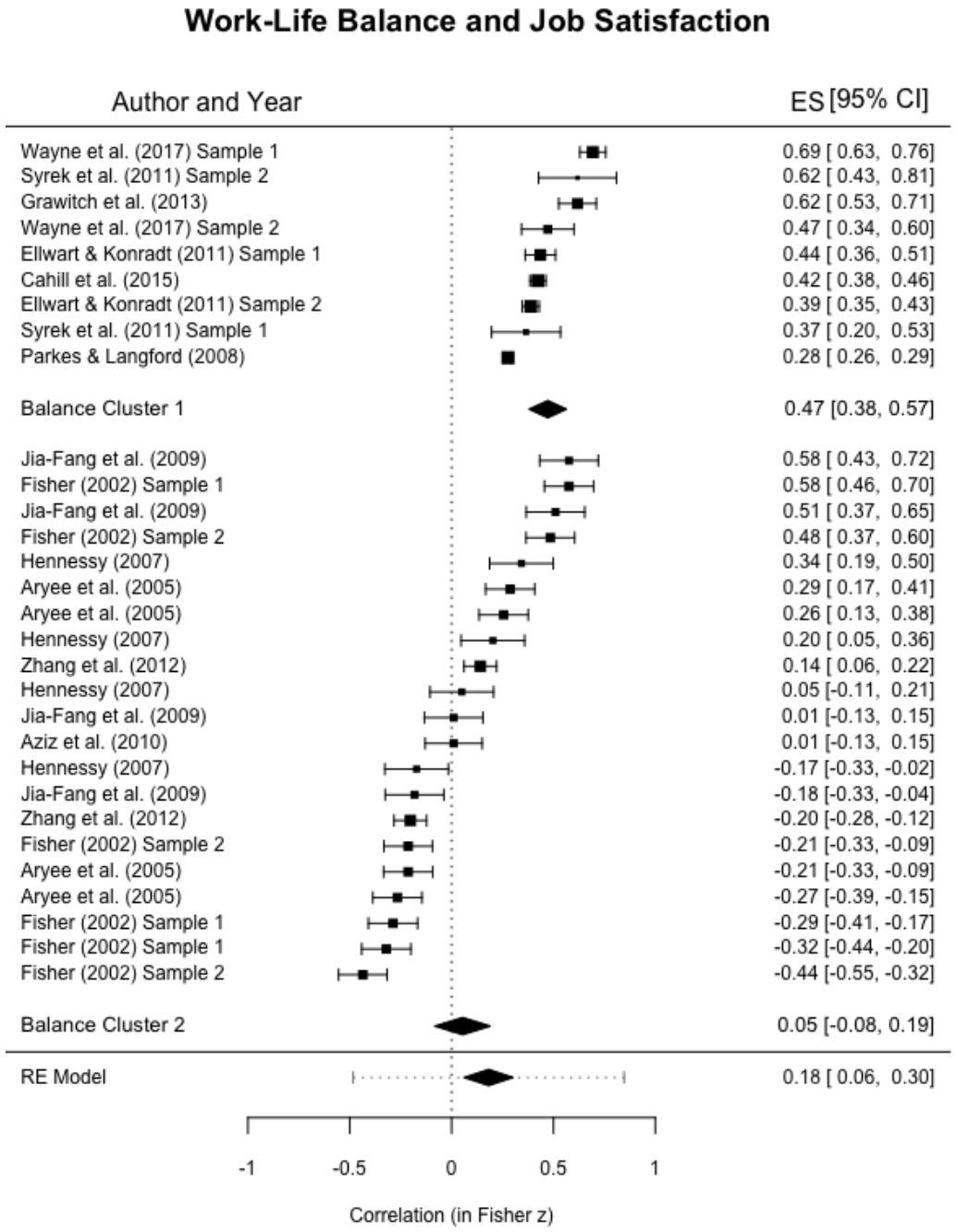
Dependent Effect Sizes

- Within Studies – Same People
  - Multiple times (e.g., pre-post)
  - Multiple measures
  - Multiple conditions
- Between Studies
  - Multiple reports of same study
  - Partial overlap between studies
  - Same sample, different authors
  - New meta contains old meta as one estimate; overlap in data
  - Mturk or similar where people may volunteer for multiple studies

Identify dependencies in data.
Describe how you addressed them.

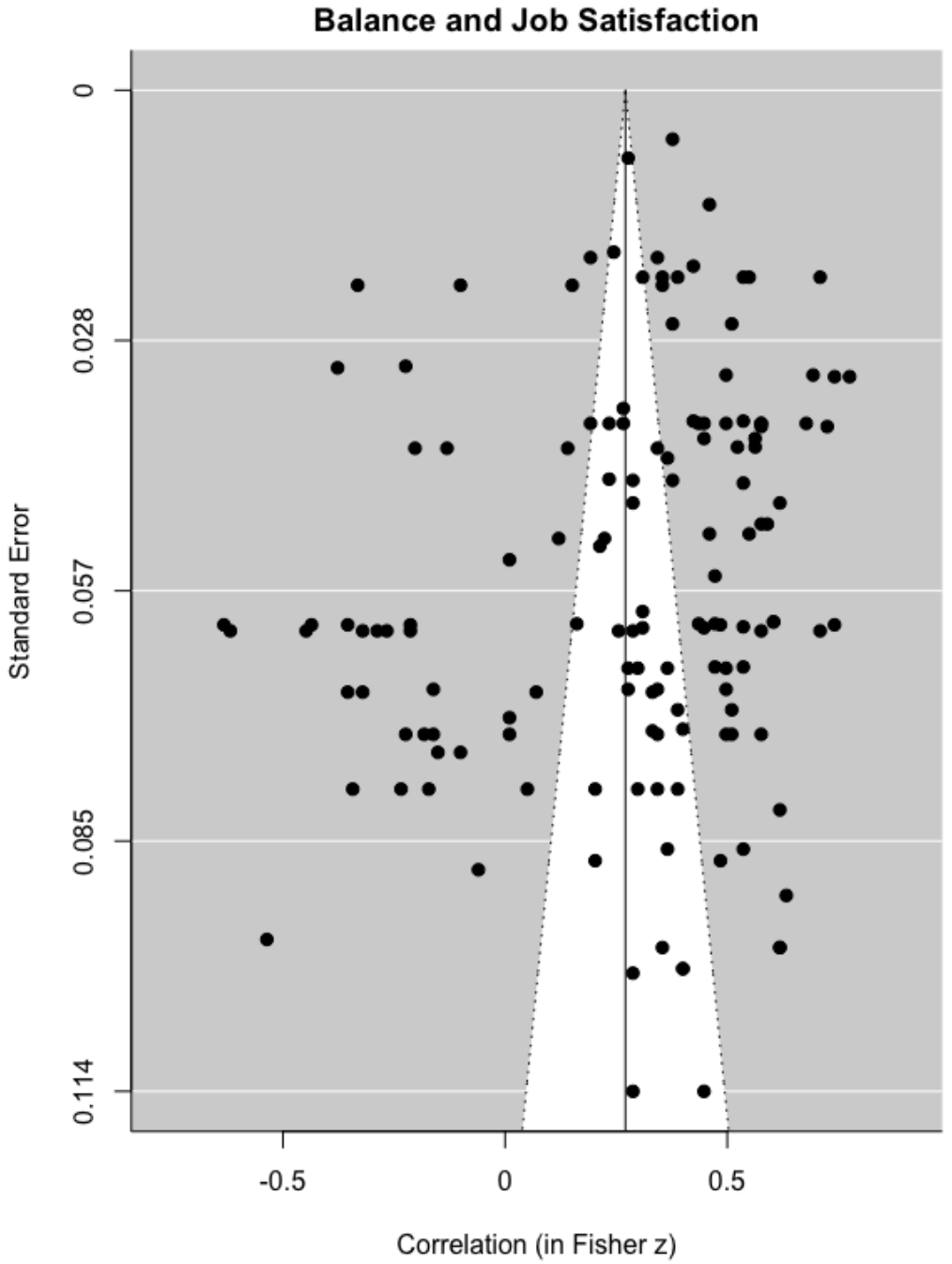
Effect Size Distribution – Visual displays

- Forest plots – good if there are few studies
- Can show outliers
- Notes
  - Sorted by effect size
  - RE summary of each level (definition of balance)
  - Cluster 2 has extreme range. Coding?
  - Prediction interval (95% PI) at bottom with overall estimate



Funnel Plot & Availability Bias

- Good if large number of studies
- Quick look for bias, outliers, heterogeneity
- Notes
  - If homogeneous, most dots inside white space
  - Huge heterogeneity in these data
  - No obvious association between ES & N (availability)
  - Streaks indicate dependent ES



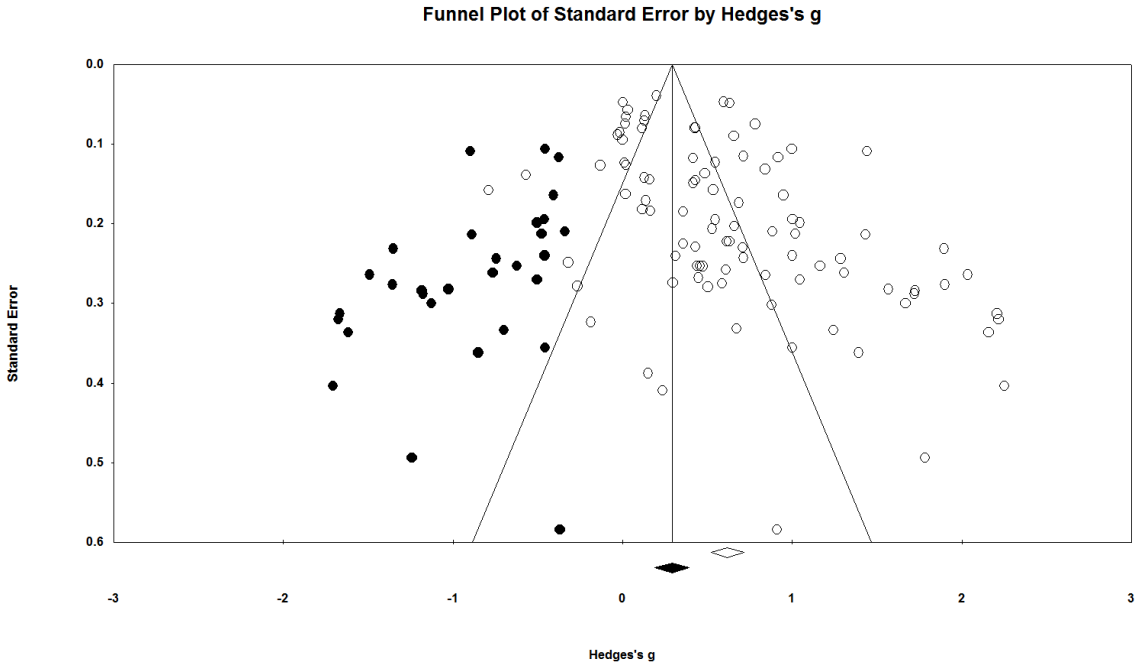
Trim & Fill

Note that the smaller studies show larger effect sizes (open dots)

Imputed studies by trim & fill algorithm are filled in black

Note the difference in the overall estimates

Other methods may be superior for this purpose



Effect Size Conversions

- Nature and Number of Conversions
  - Describe choices and rationale
- Sensitivity Analysis
  - Impact of study type that required conversion

Reporting issues in the analysis

Software

- A lot of variety in software packages
  - Some like R's metafor package are very flexible and can recreate the results from other software packages (e.g., CMA), but depending on settings, you can get different results (e.g., tau, SD-rho)
- Increasingly, a lot of options in these programs
  - Need for the reporting of estimators, thresholds used, etc.

Bivariate Results

- Hunter-Schmidt (psychometric MA) has been dominant paradigm in IO/MGMT
  - Corrected and uncorrected estimates
  - report the standard deviation and standard error of observed and corrected summary effect sizes.
    - These values are relevant to heterogeneity and meta-analytic moderation
  - Corrections used
    - Type(s) of reliability corrected for
    - Type/case of range restriction (if applicable)
    - other artifacts corrected for
    - Whether the corrections were computed based on data reported in the primary study or through artifact distributions
  - Order of corrections

Multivariate Results

- Increasingly common to combine Hunter-Schmidt with Lipsey-Wilson to conduct meta-regression/meta-ANOVA
- Not going to discuss the appropriateness of these combinations, just the reporting of their use
- Same goes for MASEM.

Meta-Regression/Meta-ANOVA

- How weight for artifact-corrected effect sizes were determined
- Whether meta-regression run on corrected or uncorrected effect sizes
- Whether assumptions were met/common sense employed
  - Scatterplots
  - Examination of the residuals
  - Number of included studies in each category

Meta-regression is not immune to regression violations

Mixed-Effects MARA Results

Variable	Model
Performance definition	
Accounting measure	.08 (.01)***
Related competitors	.08 (.08)
Family firm definition	
Ownership definition	−.00 (.04)
Ownership and management definition	.03 (.05)
Self-reported	−.16 (.07)
Study characteristics	
Published	.04 (.02)
Publication year	.01 (.01)*
Response rate	.23 (.12)*
Institutions	
Common-law country	.12 (.07)*
Rule of law	.05 (.03)*
R^2	.90
k	37

96% of data from Norway

MASEM

- MAs in management routinely combine population estimates from prior MAs with their own findings to test complex models
- Report choice of which past MA to use
- Discuss the independence and comparability of samples across multiple MAs
- Justify the appropriateness of combining meta-analytic findings, as different MAs may have relied on more or less relevant inclusion criteria
- Degree of effect size heterogeneity and the specific procedures used
  - How large were the SD-rho?
- How many elements in the correlation matrices were missing
  - Did it meet MCAR/MAR assumptions?