Workshop “Analytical Sociology: Theory and Empirical Applications”, Venice, 20.11.2019

Asking Sensitive Questions: A Meta-Analysis of Experimental Survey Studies on the Performance of the Item Count Technique

Justus Junkermann, Felix Wolter, Ingmar Ehler

Johannes Gutenberg University Mainz, University of Konstanz, TU Kaiserslautern

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Universität Konstanz

JG|U

Outline

1. Introduction
2. The Item Count Technique
3. Meta-Analysis Design
   - Data Collection
   - Methods
4. Results
   - Overall effect
   - Publication bias
   - Meta-regression
5. Discussion

Background

- DFG project:
  - „Sensitive Questions and Social Desirability: Theory and Methods“.
  - PIs: Felix Wolter (Konstanz/Mainz), Jochen Mayerl (Chemnitz).
  - Research assistants: Justus Junkermann (Mainz), Henrik Andersen (Chemnitz).

- Sensitive questions in surveys:
  - E.g., delinquency, drug abuse, health issues, sexuality, xenophoby, homophoby, voting, green behavior.
  - Main problems:
    - Misreporting
    - [Item nonsresponse]

DO THEY TELL THE TRUTH?

BY HERBERT HYMAN[1]

Do the answers of people really give a true picture of their behavior?

The answer to this question is obviously of basic importance to the whole scientific study of opinion. Some facts

It is unlikely that the distortion can be explained in terms of forgetting. If this were the case, the distortion probably would be equally distributed among all economic strata.

Hyman 1944.

Background

- Survey methodologists have proposed various special questioning techniques in order to tackle the problems of sensitive questions.

- The principle is to add statistical random noise to the respondents‘ answers in order to conceal their true response:
  - Randomized Response Technique (RRT; Warner 1965).
  - Crosswise Model (CM; Yu et al. 2008).
  - Item Count Technique (ICT; Miller 1984, Droitcour et al. 1991).

- Our hope is that the enhanced anonymity / reduced embarrassment makes repondents answer truthfully to sensitive questions.

Background

- But: Do these questioning techniques actually work?

- Two main types of studies evaluating their performance:
  - External validation studies.
    - RRT/CM/ICT estimates are compared to known true values.
    - Very rare.
  - „More is better“ / „Less is better“ studies.
    - Survey experiments comparing RRT/CM/ICT estimates to conventional direct questioning (DQ).
    - Higher estimates for negatively connoted things (e.g., stealing) are considered as more valid, lower ones for positively connoted things (e.g., green behavior).

Background

- This study: Meta-analysis of all experimental studies published comparing ICT with DQ for sensitive items.

- Existing literature:
  - Meta-analysis on RRT studies (Lensvelt-Mulders et al. 2005):
    - RRT better than DQ.
    - But only 6+32=38 studies.
  - Meta-analysis on ICT studies (Blair et al. 2019, unpublished).
    - 264 effects, studies published until 2017.
    - ICT better than DQ with a stronger ICT effect for overreporting as compared to underreporting.
    - No meta-regression.

Outline

Item Count Technique (ICT)

- Experimental design, random split of the sample into two groups:
  - short-list group: *k* binary non-key / filler items.
  - long-list group: *k* binary non-key / filler items plus one sensitive item.
- Respondents indicate only the number of items that apply.
- Individual answer to the sensitive item is not disclosed.
- Calculation of the prevalence estimate for the sensitive item ($\hat{\pi}_{ICT}$):

$\hat{\pi}_{ICT} = \bar{x}_{LL} - \bar{x}_{SL}$, with $\bar{x}_{SL}$ = mean of the short list
$\bar{x}_{LL}$ = mean of the long list

- Sampling variance:

$$Var(\hat{\pi}_{ICT}) = Var(\bar{x}_{LL}) + Var(\bar{x}_{SL})$$

Item Count Technique (ICT)

- Example (Wolter & Laier 2014), face-to-face survey:
  - INT: *„For the next questions, we are going to use a special technique that guarantees you complete anonymity. I am going to hand you lists with four [five] questions, which you should please read first. Then, please tell me only the number of questions that you answer with ‚yes', thus, a number between 0 and 4 [5]"*.

Short-list group

- Have you ever been abroad?
- Have you ever used a taxi?
- Have you been using a plane this week?
- Did you wash your car this week?

Long-list group

- Have you ever been abroad?
- Have you ever used a taxi?
- Have you been using a plane this week?
- Did you wash your car this week?
- <u>Have you ever been driving a car although you had drunk too much alcohol?</u>

Variants of ICT

- Person Count Technique (PCT):
  - Use other persons instead of non-key questions.
  - E.g.: „Please think of 3 people you know. How many of them, including yourself, have ever taken cocaine?
- Item Sum Technique (IST):
  - Use metric non-key and sensitive items.
  - Not included in this study.
- Double-list design:
  - Add a second different short list to the long-list group and the corresponding long list including the (same) sensitive item to the short-list group.
  - Increases statistical efficiency.

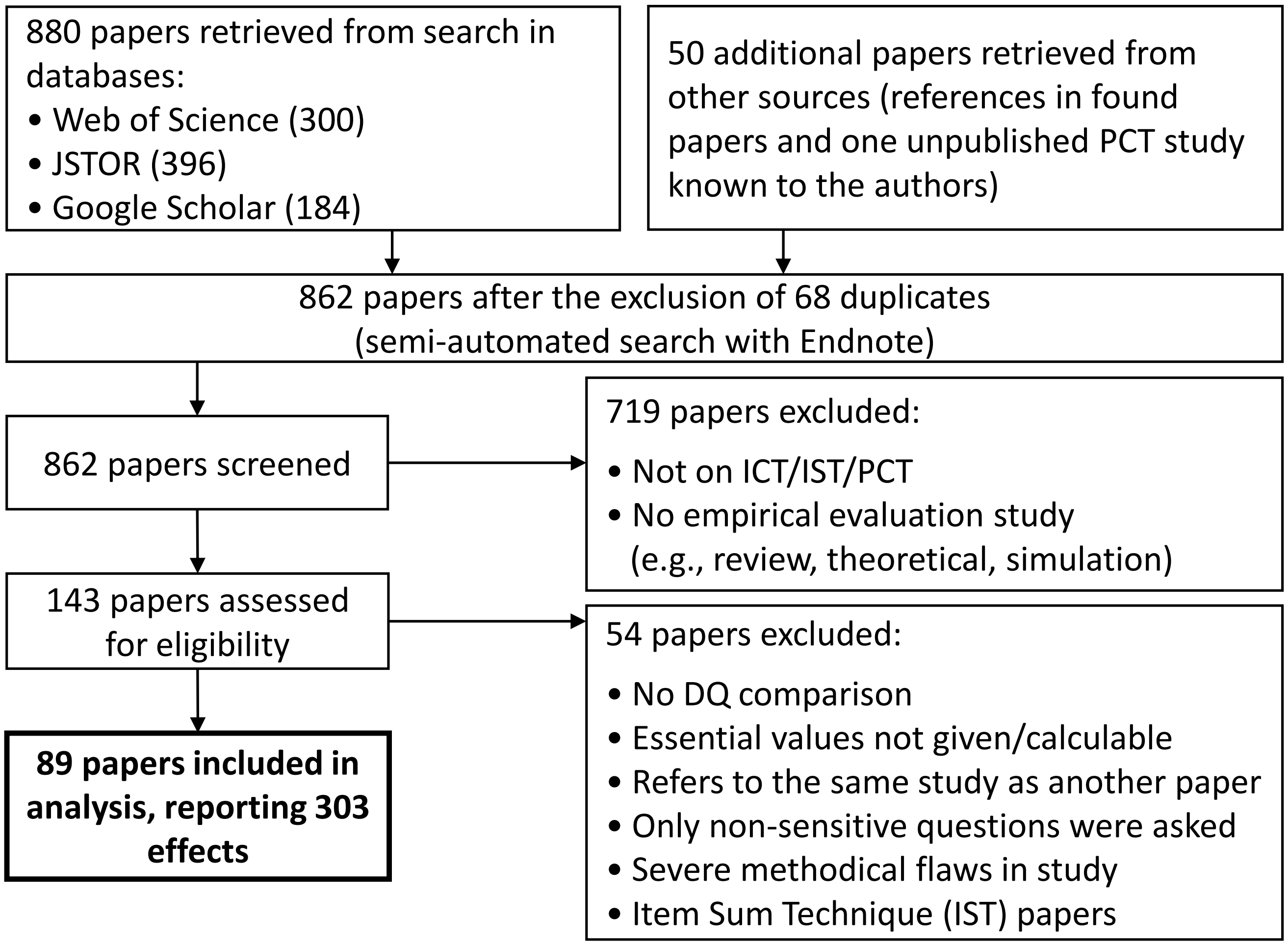
Outline

Data Collection

- Only studies included with:
  - Item Count Technique (ICT) or Person Count Technique (PCT) used.
  - Same item asked with Direct Questioning (DQ).
  - Sensitive items (i.e., no „test items“, SD bias expected).
  - Reported or computable standard errors.

- Systematic search for literature
  - With standardized search string in:
    - Jstor
    - Web of Science
    - Google Scholar
  - Bibliographical references of further studies, grey literature, unpublished studies etc.

Data Collection



Methods

- Effect size:
  - Raw mean difference
  - ICT−DQ, if affirmative answer is socially undesirable, e.g. delinquency.
  - DQ−ICT, if affirmative answer is socially desirable, e.g. voting.
- Data Structure:
  - 89 papers.
  - Containing 124 studies/samples.
  - Containing 303 estimates.
- Modelling:
  - Robust random effects models.
  - DerSimonian and Laird Model (DL) (Veroniki et al 2016; Borenstein et al 2009).
  - Multilevel REML (3lvl) (Veroniki et al 2016).
  - Multilevel Robust Variance Estimation (RVE) (2lvl) (Fisher & Tipton 2015).

Outline

1. Introduction
2. The Item Count Technique
3. Meta-Analysis Design
   - Data Collection
   - Methods
4. **Results**
   - **Overall effect**
   - **Publication bias**
   - **Meta-regression**
5. Discussion

Results

- Overall effect size of ICT vs. DQ:

Model	Estimate	SE	τ^2	I^2
Unweighted Mean	0.085	0.009	--------------	-------------------
DL	0.086	0.009	0.008	94.1%
2lvl RVE	0.085	0.010	0.010	92.0%
3lvl REML	0.085	0.011	0.008	94.8%

- High heterogeneity between studies (I^2>90%)
- 95%-prediction interval = [−0.11; 0.28]
  – You can‘t be sure, that ICT works as intended in a future study
- **ICT works, but it‘s not reliable.**

Prediction Interval

- The prediction interval is the interval in which 95% of new effect sizes will fall.

$$PI = M \mp t_{df}^{\alpha}\sqrt{T^2 + V_M}$$

- With

  M = mean effect size;

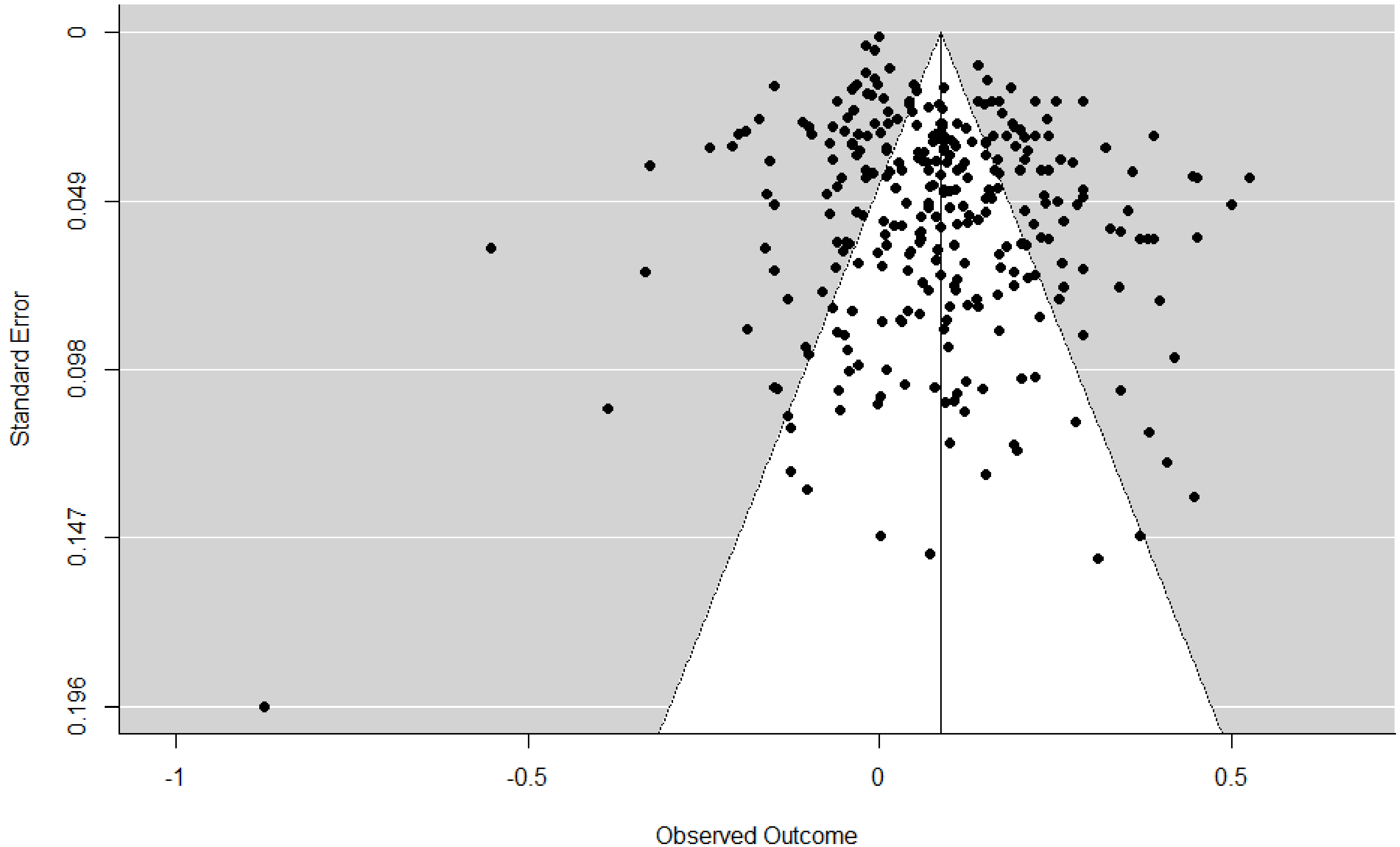
  t_{df}^{α} with α = 0.05 and *df* = k−2, with k = number of studies;

  T^2 = sample estimate of the variance of true effect sizes;

  V_M = sampling variance of M.

(Borenstein et al 2009, 131)

Publication Bias?



Eggers regression test p=0.44; rank test Kendall‘s tau = −0.1**; trim and fill = 0

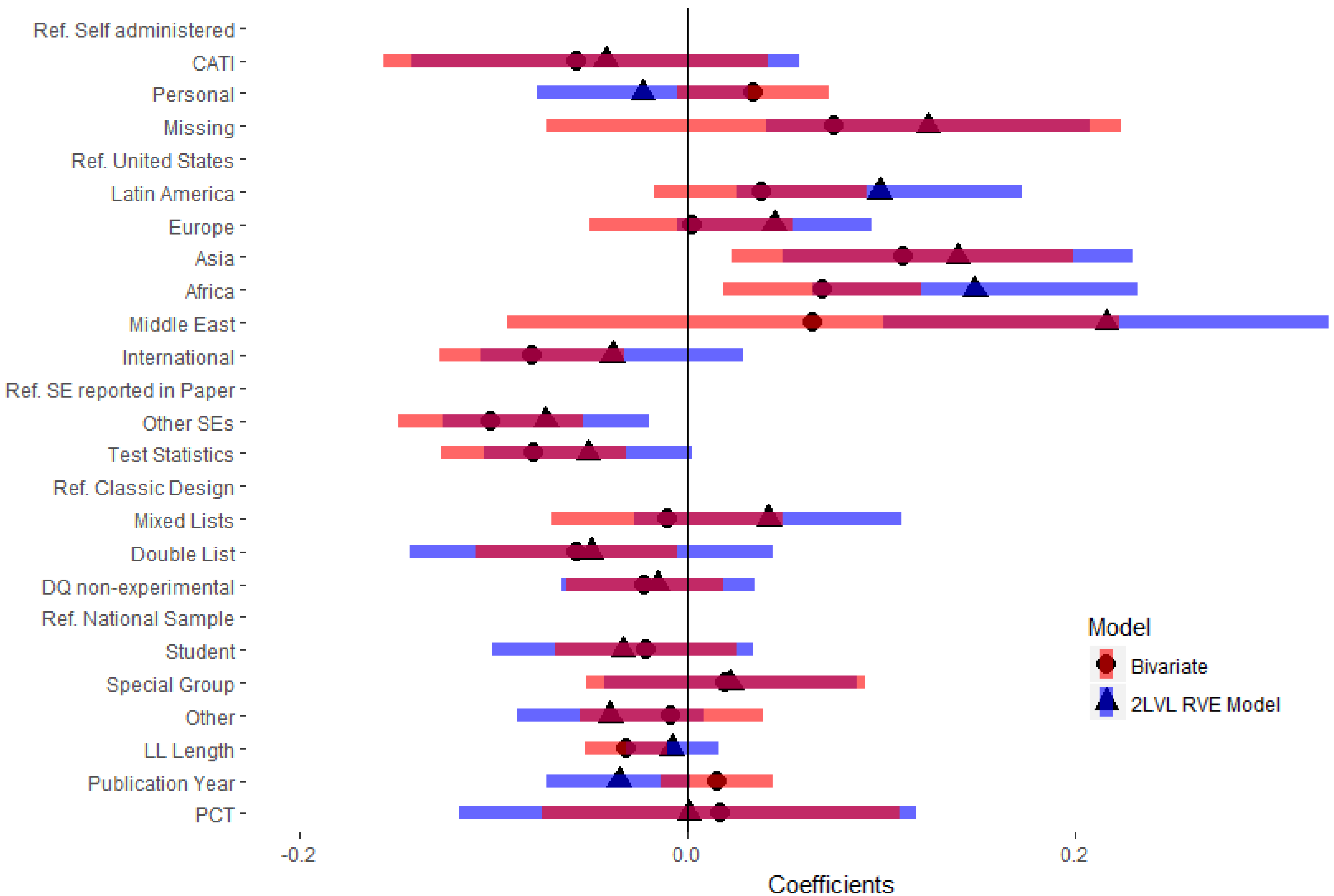
Meta-Regression

- Study characteristics:
  - Survey mode
  - Geographic region
  - Source of standard errors
  - ICT design (e.g. double list)
  - Sample (e.g. convenience vs. national)
  - ICT or PCT
  - Publication year
- Item characteristics:
  - Illegal behaviour
  - Sensitivity
  - Direction of social desirability
  - Opinion/trait/behaviour

Meta-Regression

- Study characteristics:
  - Survey mode
  - **Geographic region**
  - **Source of standard errors**
  - **ICT design (e.g. double list)**
  - Sample (e.g. convenience vs. national)
  - ICT or PCT
  - Publication year
- Item characteristics:
  - Illegal behaviour
  - Sensitivity
  - **Direction of social desirability**
  - Opinion/trait/behaviour

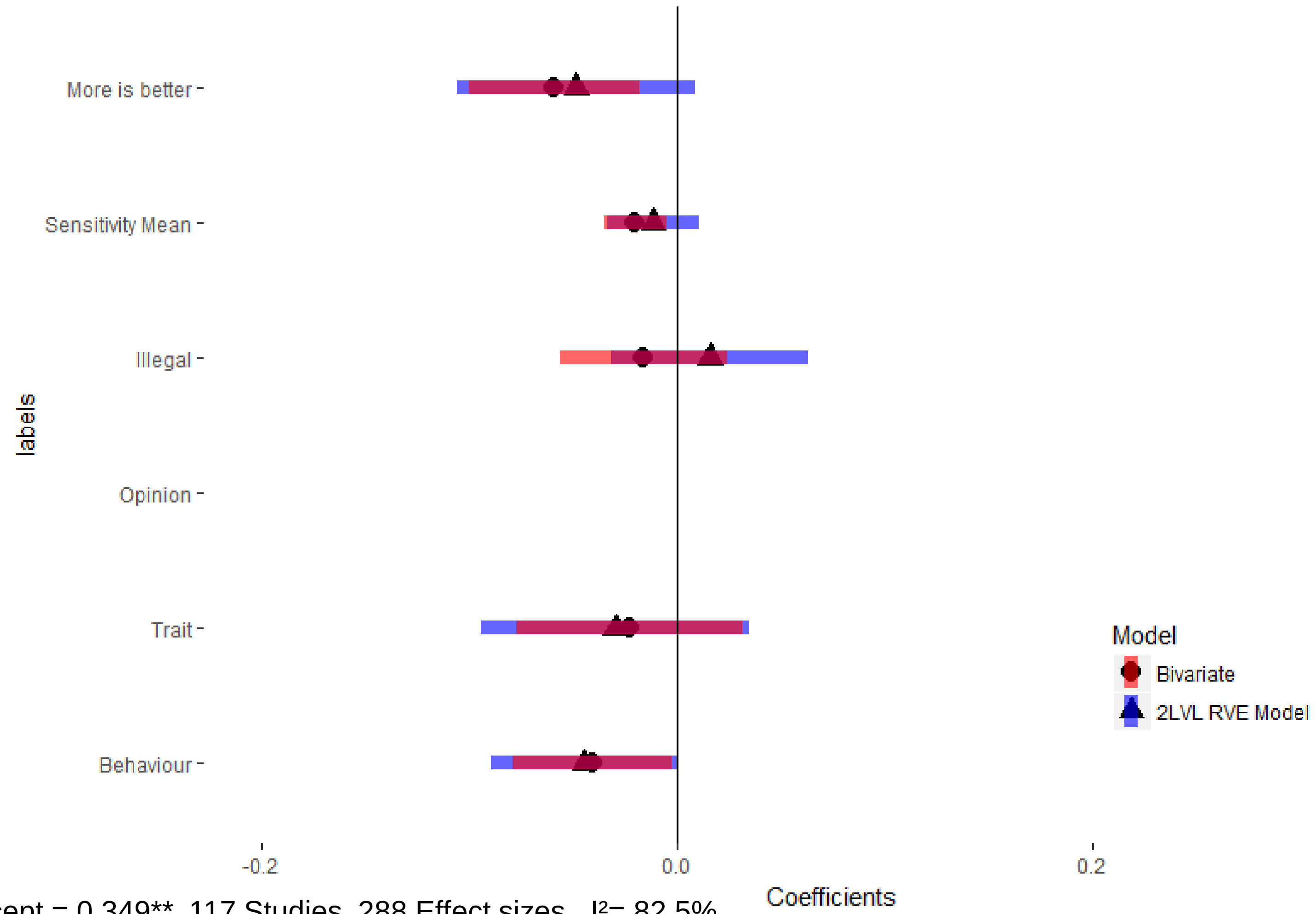
Meta Regression – Study Characteristics



Intercept = 0.349**, 117 Studies, 288 Effect sizes , I² = 82.5%

Junkermann, Wolter, Ehler
Mainz, Konstanz, Kaiserslautern

Meta Regression – Item Characteristics



Intercept = 0.349**, 117 Studies, 288 Effect sizes , I²= 82.5%

Junkermann, Wolter, Ehler
Mainz, Konstanz, Kaiserslautern

Summary of Meta-Regression Results

- PCT performs as well as ICT.
- ICT performs better in non-western cultures (Africa, Asia and Arab World).
- ICT performs slightly better if the answer is socially desirable.
- The double-list design seems to yield worse results than standard ICT.

Outline

Overall Summary

- ICT works, but it is not reliable / under control of the researcher.
  - ICT performs 8.5 percentage points better than DQ, which is 28% better.
  - There is distinct heterogeneity between studies ($I^2 > 90\%$).
  - The prediction interval for a future ICT-DQ difference is [-0.11; 0.28], so you can‘t be sure that it will provide better estimates than DQ the next time you use it.
- We did not find publication bias.
- ICT works better in non-western societies.
- The double-list design performs poorer than the original ICT design.

Vielen Dank!

j.junkermann@uni-mainz.de
felix.wolter@uni-konstanz.de
ingmar.ehler@sowi.uni-kl.de

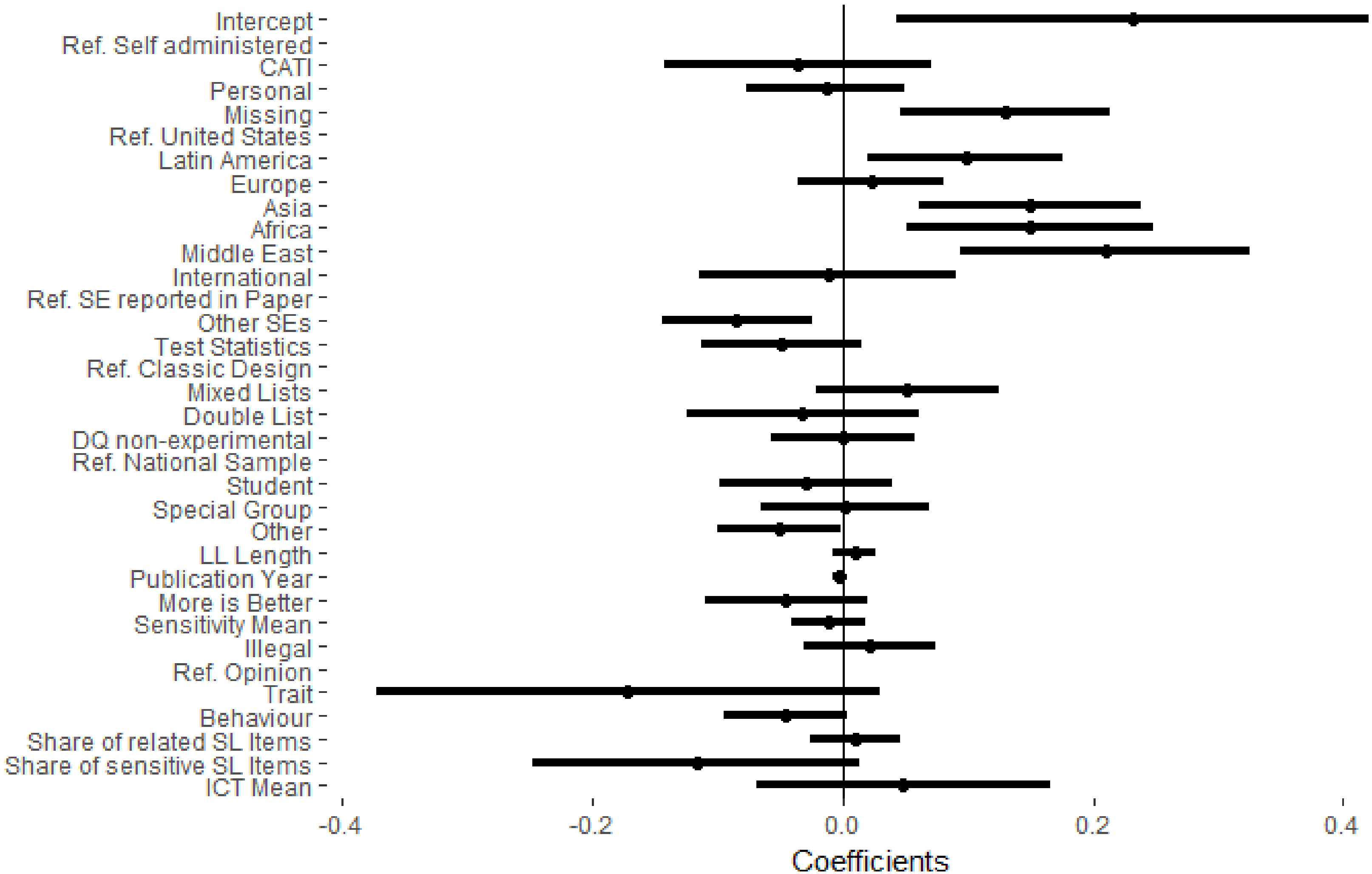
Additional Slides

Sensitivity Rating

- 8 experts rated the items on a likert scale from 1 (not sensitive at all) to 7 (very sensitive) in an online survey.

- The mean sensitivtity rating of each item was taken as an independent variable in the meta regression.

RVE with ICT Mean and SL Variables Included



Weights for Models

Estimate of ...	Fixed-effect	Random-effects
Weights	$w_i = \frac{1}{v_i}$	$w_i^* = \frac{1}{v_i + \hat{\tau}^2}$
Parameter	$M = \frac{\sum w_i T_i}{\sum w_i}$	$M^* = \frac{\sum w_i^* T_i}{\sum w_i^*}$
Variance	$var(M) = \frac{1}{\sum w_i}$	$var(M^*) = \frac{1}{\sum w_i^*}$
SE	$SE(M) = \sqrt{var(M)}$	$SE(M^*) = \sqrt{var(M^*)}$
95% CI	$M \pm 1.96 \times SE(M)$	$M^* \pm 1.96 \times SE(M^*)$

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Robust Variance Estimation

Weighted residual sum of squares

$$\mathrm{Q_E} = \sum_{j=1}^{m} \mathbf{T}_j'\mathbf{W}_j\mathbf{T}_j - \left(\sum_{j=1}^{m} \mathbf{T}_j'\mathbf{W}_j\mathbf{X}_j\right)\left(\sum_{j=1}^{m} \mathbf{X}_j'\mathbf{W}_j\mathbf{X}_j\right)^{-1}\left(\sum_{j=1}^{m} \mathbf{X}_j'\mathbf{W}_j\mathbf{T}_j\right)$$

Weights

$$\mathbf{W}_j = w_j\mathbf{I}_j = \{1/[k_j(\nu_{\cdot j} + \tau^2)]\}\mathbf{I}_j$$

Methods of moments estimator

$$\hat{\tau}^2 = \frac{\mathrm{Q_E} - m + \mathrm{tr}\left[\mathbf{V}\left(\sum_{j=1}^{m}\frac{w_j}{k_j}\mathbf{X}_j'\mathbf{X}_j\right)\right] + \rho\,\mathrm{tr}\left[\mathbf{V}\left(\sum_{j=1}^{m}\frac{w_j}{k_j}\left[\mathbf{X}_j'\mathbf{J}_j\mathbf{X}_j - \mathbf{X}_j'\mathbf{X}_j\right]\right)\right]}{\sum_{j=1}^{m}k_jw_j - \mathrm{tr}\left[\mathbf{V}\left(\sum_{j=1}^{m}w_j^2\mathbf{X}_j'\mathbf{J}_j\mathbf{X}_j\right)\right]}$$

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv preprint arXiv:1503.02220*.

DerSimoniaLaird (DL) Model

$$Q = \sum w_i T_i^2 - \frac{(\sum w_i T_i)^2}{\sum w_i} \text{ mit } (w_i = \frac{1}{v_i})$$

$$\tau^2 = \frac{Q - df}{C};\ df = k - 1$$

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i}$$

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, *7*(1), 55-79.

Restricted Maximum Likelihood (REML) Estimation

$$lnL(\tau^2) = -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\sum \ln(v_i + \tau^2) - \frac{1}{2}\sum \frac{\left(y_i - \hat{\mu}_{RE}(\hat{\tau}^2_{ML})\right)^2}{(v_i + \tau^2)} - \frac{1}{2} ln\left(\sum \frac{1}{(v_i + \tau^2)}\right)$$

$$\hat{\tau}^2_{REML} = \max\left\{0, \frac{\sum w^2_{i,RE}\left(\left(y_i - \hat{\mu}_{RE}(\hat{\tau}^2_{ML})\right)^2 - v_i\right)}{\sum w^2_{i,RE}} + \frac{1}{\sum w_{i,RE}}\right\}$$

($w_{i,RE}$ is obtained from DL model)

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, *7*(1), 55-79.

Short Summary

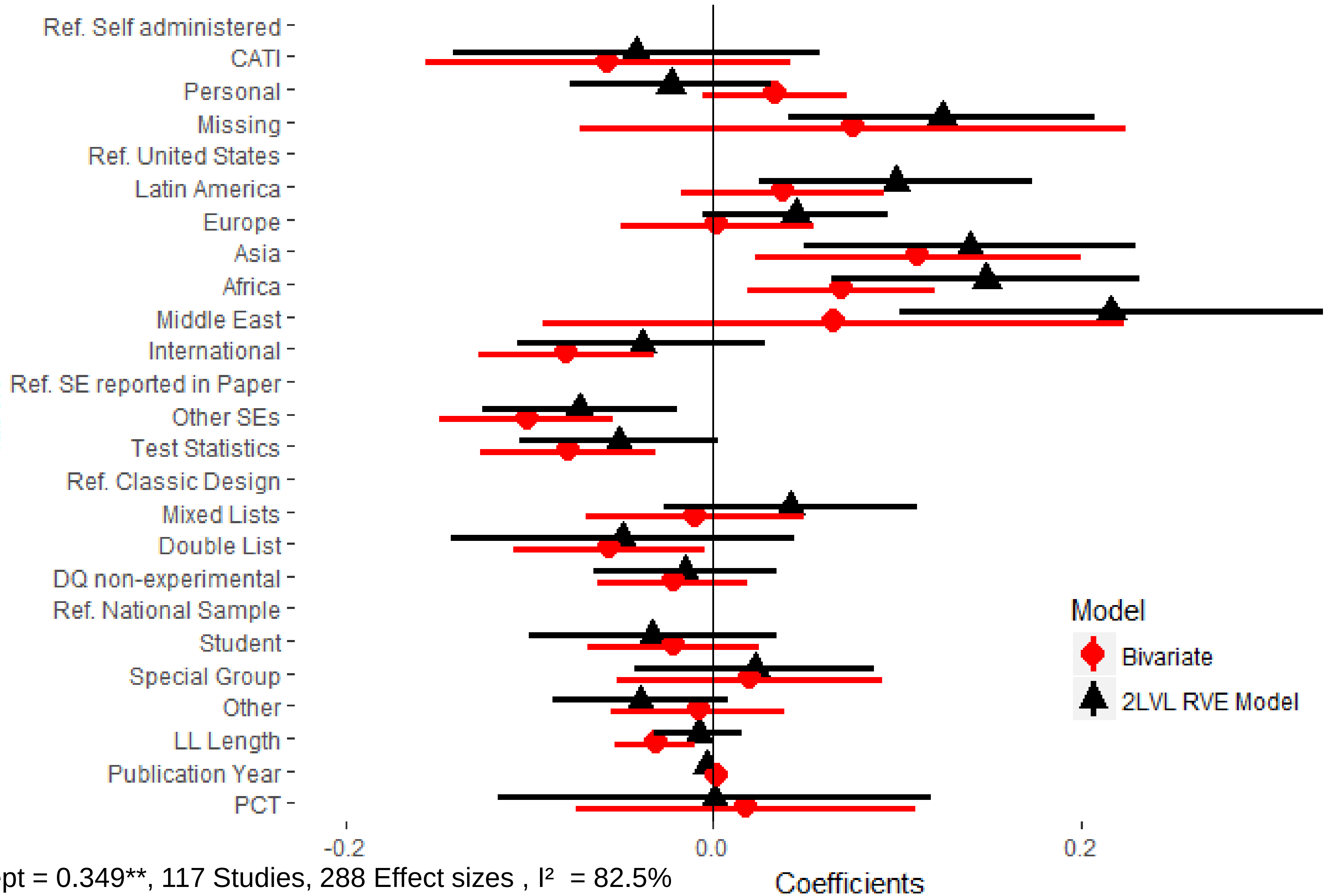
Meta-Analysis
303 effect estimates in 124 samples from 89 papers
Criteria: Comparative prevalence estimates for Direct Questioning (DQ) Standard errors are reported or can be computed
Results
ICT performs ~8.5 percentage points better than DQ ($p<0.001$)
This is ~28% better than DQ
High heterogeneity between studies ($I^2 > 90\%$)

Blair et al. 2019 Results

		Average Sensitivity Bias $\widehat{\mu}$		Predictive Intervals $\mathcal{N}(\widehat{\mu}, \widehat{\tau})$		
	Prediction	Estimate (s.e.)	95% C.I.	50%	95%	N studies
Vote buying	Underreporting	-0.08 (0.03)	[-0.13, -0.03]	[-0.14, -0.02]	[-0.25, 0.09]	19
Turnout	Overreporting	0.07 (0.04)	[-0.01, 0.14]	[0.00, 0.13]	[-0.12, 0.25]	10
Racial prejudice	Underreporting	0.04 (0.03)	[-0.03, 0.09]	[0.01, 0.07]	[-0.05, 0.13]	9
Religious prejudice	Underreporting	-0.01 (0.03)	[-0.08, 0.05]	[-0.06, 0.03]	[-0.14, 0.11]	12
Sexual orientation prejudice	Underreporting	0.02 (0.02)	[-0.03, 0.06]	[-0.02, 0.05]	[-0.09, 0.12]	16
	Overreporting	-0.02 (0.09)	[-0.19, 0.16]	[-0.12, 0.09]	[-0.31, 0.28]	5
Support for authoritarian regimes	Underreporting	-0.08 (0.04)	[-0.16, -0.00]	[-0.16, -0.00]	[-0.32, 0.15]	13
	Overreporting	0.14 (0.04)	[0.07, 0.21]	[0.05, 0.24]	[-0.14, 0.42]	21
All results	Underreporting	-0.03 (0.01)	[-0.05, -0.01]	[-0.12, 0.05]	[-0.27, 0.20]	196
	Overreporting	0.12 (0.02)	[0.08, 0.15]	[0.03, 0.20]	[-0.13, 0.36]	68

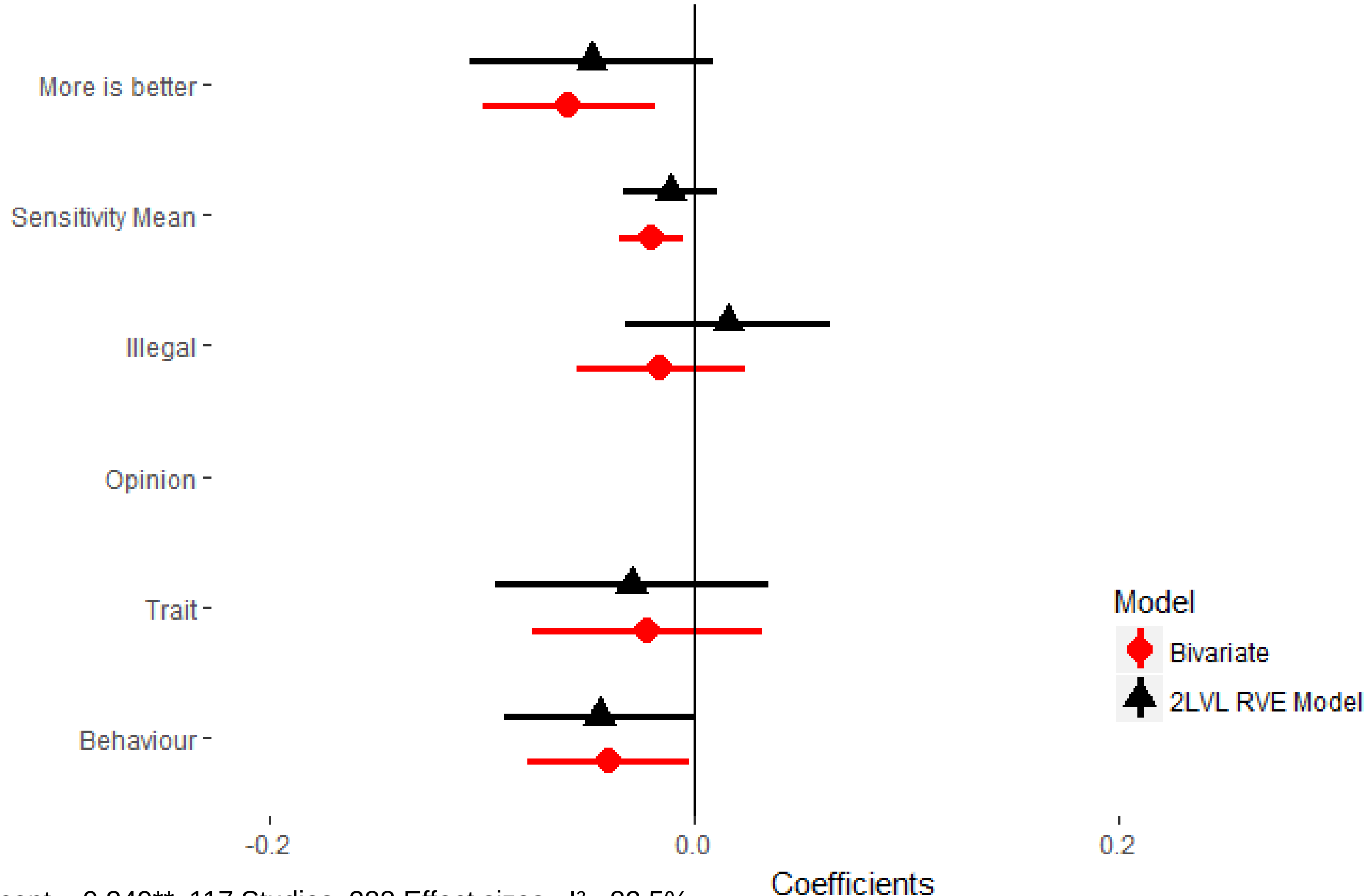
Blair et al. 2019: 22.

Meta Regression – Study Characteristics



Intercept = 0.349**, 117 Studies, 288 Effect sizes , I^2 = 82.5%

Meta Regression – Item Characteristics



Intercept = 0.349**, 117 Studies, 288 Effect sizes , I^2= 82.5%

Literature

- Blair, Graeme/Coppock, Alexander/Moor, Margaret 2019: When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments: Unpublished Manuscript, https://graemeblair.com/papers/sensitivity-bias.pdf.
- Borenstein, Michael et al. 2009: Introduction to Meta-Analysis, Chichester: Wiley.
- Fisher, Zachary/Tipton, Elizabeth 2015: robumeta: An R-package for robust variance estimation in meta-analysis, *arXiv preprint arXiv:1503.02220*.
- Lensvelt-Mulders, Gerty J. L. M. et al. 2005: Meta-Analysis of Randomized Response Research. Thirty-Five Years of Validation, Sociological Methods and Research 33 (3): 319–348.
- Veroniki, A. A./Jackson, D./Viechtbauer, W./Bender, R./Bowden, J./Knapp, G./Kuss, O./Higgins, JP/Langan, D/Salanti, G. (2016): Methods to estimate the between-study variance and its uncertainty in meta-analysis, Research synthesis methods, 7(1): 55–79.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package, Journal of Statistical Software, 36(3): 1–48