

Meta-analysis on the effects of interview supportiveness on the accuracy of children's reports

Christine Wells, Ph.D.,¹ Karen Saywitz, Ph.D.,¹
Rakel Larson, Ph.D.,² and Sue Hobbs, MA²

¹University of California, Los Angeles

²University of California, Davis

July 30, 2015

In preparing for a national survey of children in foster care, the Swedish government commissioned unbiased researchers to do a systematic review of the experimental literature. They wanted recommendations about how to better obtain reliable information from the children in foster care to assess their well-being and other outcomes.

- Baseline accuracy: Are children more accurate in a supportive v. a non-supportive context?
- Baseline accuracy: Do children make more errors in a supportive v. a non-supportive context?
- Are children more suggestible in a supportive v. non-supportive interview context?
- Does support give children resistance to suggestive questions?

We did the following to find relevant research:

- Searched six electronic databases: PsychInfo, PubMed, Sociological Abstracts, Social Services Abstracts, Web of Knowledge and Cochrane Central
- Hand-searched the reference lists from 30 authoritative reviews
- Contacted leading scholars in the field

We decided that all articles found would be included until we found a reason to exclude the article. We started with nine exclusion criteria, but we had a total of 12 by the time that we were finished. This required that we re-examine some articles. The number of exclusion criteria increased because we had underestimated the variability in the literature on this topic.

- 2,794 potentially relevant articles identified
- Exclusion criteria applied with 99% agreement
- After applying 12 exclusion criteria, 15 articles were left
- 10% of the excluded studies were reviewed to ensure accuracy (100% agreement)

- Checklist for Measuring Quality (Downs and Black, 1998)
- 97% agreement between two raters
- Differences resolved by a third judge
- The range of scores was 18 - 22 out of 27 (modified last item); higher scores are better
- Downs and Black mean = 20.14; SD = 1.49
- Quality of study design v. quality of reporting
 - reporting that Ss were randomly assigned
- Issues regarding age of study
 - p-values
 - power analysis
 - ANOVA with count variable as DV

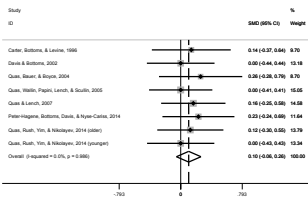
- 15 studies identified for the systematic review
- Because of missing data, only eight studies in the meta-analysis
- One study had independent groups of younger and older children, and these groups became two rows in the dataset

Random effects meta-analysis with Hedges' correction

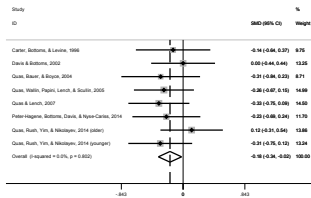
```
metan spec_supp_N_corr spec_supp_MEAN_corr spec_supp_SD_corr ///
spec_nonsupp_N_corr spec_nonsupp_MEAN_corr spec_nonsupp_SD_corr, ///
hedges random name("forest_spec_corr", replace) ///
title("Correct Responses" "to Neutral Questions", ///
color(black)) label(namevar = Authors) sortby(year) texts(100) ///
astext(60) graphregion(fcolor(white) ifcolor(white)) ///
olineopt(lcolor(black) lpattern(dash)) diamopt(lcolor(black)) ///
nowarning notable
```

Forest plots

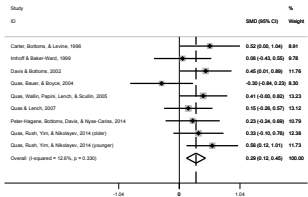
Correct Responses to Neutral Questions



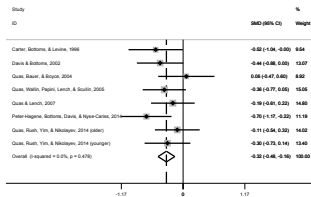
Incorrect Responses to Neutral Questions



Correct Responses to Misleading Questions



Incorrect Responses to Misleading Questions



Potential sources of non-independence

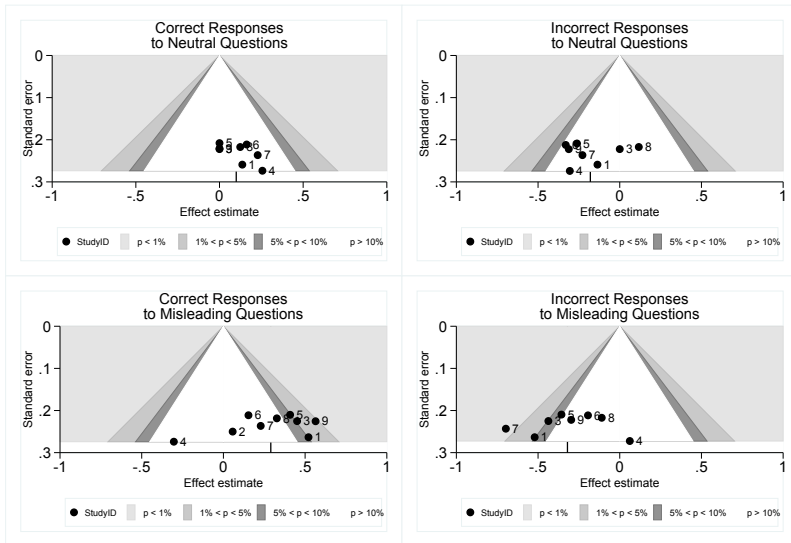
- Two rows of data from the same study
- All studies include one of three authors
- We calculated design effects to assess each possible source of non-independence

- Several possible techniques for assessing outliers
- We used the method described in Viechtbaruer and Cheung (2004)
- t-value cutoff of ± 1.96
- Assessing outliers is difficult in such a small dataset
- One possible outlier in each of two analyses
- We did a sensitivity analysis for both, and the results did not change much
- We left both studies in the analyses because we couldn't find a good reason to exclude them

- Both -metabias- and -metatrim- were used
- We did not find evidence of publication or small sample bias
- Not enough variability to warrant meta-regressions
- All I-squares less than 13% and non-significant

```
confunnel ESSpecSuppCorr seESSpecSuppCorr, ///  
graphregion(fcolor(white) ifcolor(white)) xtitle(, size(small)) ///  
ytitle(, size(small)) name(spec_supp_corr_contour, replace) ///  
legendopts(rows(1) size(vsmall) region(lstyle(none))) ///  
symxsize(*.2) scatteropts(mlabel(StudyID) mlabcolor(black) ///  
title("Correct Responses" "to Neutral Questions", color(black)) ///  
xline(.10, lcolor(black))) studylab(StudyID)
```

Contour plots of bias



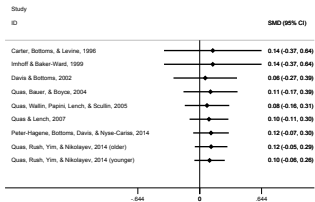
Cumulative meta-analysis

- Order the studies chronologically
- Often used to see when an answer has been reached and the research can move on
- Support is helpful with misleading questions, but it is not yet clear if support is helpful with neutral questions

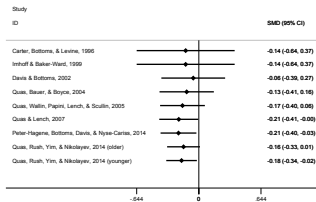

```
metacum spec_supp_N_corr spec_supp_MEAN_corr spec_supp_SD_corr ///  
spec_nonsupp_N_corr spec_nonsupp_MEAN_corr spec_nonsupp_SD_corr, ///  
hedges random sortby(year) name(cum_spec_corr_random, replace) ///  
label(namevar = Authors) textsize(100) astext(60) ///  
title("Correct Responses" "to Neutral Questions", color(black))
```

Cumulative meta-analyses

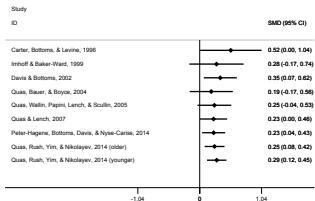
Correct Responses to Neutral Questions



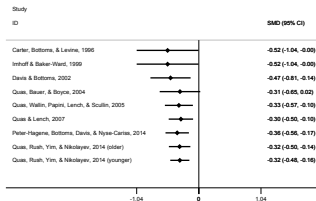
Incorrect Responses to Neutral Questions



Correct Responses to Misleading Questions



Incorrect Responses to Misleading Questions



- There is so much variability in the literature that creating the dataset was difficult.
- Meta-analysis is a rapidly developing area of statistics.
- Tons of quality checklists and guidelines - often rather different for medicine and social sciences.
- Despite the advances in statistical methodology, the most challenging questions have to do with getting the effect size.
- Researchers are perhaps more attune to issues of data quality because of the data collection method (and how long it takes!).