A comparison of meta-analysis, mega-analysis, and a hybrid approach

Ezequiel Koile[a], Sho Tsuji[b], & Alejandrina Cristia[c]

[a] ADD

[b] ADD

[c] ADD

Author Note

Correspondence concerning this article should be addressed to Ezequiel Koile, ADD.
E-mail: ADD

Abstract

Laboratory measures of infant speech perception have been central to the development of theories of infant language acquisition, and could be valuable predictors of important individual and group variation. A recent report suggests that these measures' psychometric properties may be limited, based on a meta-analytic analysis. We re-analyze those data using a mega-analytic approach, as well as a variety of hybrid approaches. We find that (a) the results of meta- and mega-analyses diverge significantly, and (b) a mega-analytic approach can be more powerful in detecting stability in performance across days. However, since it is often difficult to recover original data, we also explore a hybrid approach, in which some studies are represented by group statistics, and others by the original data, assessing to what extent biased data sharing may impact overall conclusions.

A comparison of meta-analysis, mega-analysis, and a hybrid approach

Rough paper outline

One paragraph per bullet point

- infant speech perception measures have been central to the development of theories of infant language acquisition, eg phonology & lexicon start developing before 1 year of age

- infant speech perception measures could be valuable predictors of important individual and group variation, eg correlations these & vocabulary or comparisons between groups at risk or not (Cristia et al 2013)

- what are the psychometric properties? Psychometry is crucial – eg test-retest reliability gives upper bound on meaningful variance: "the maximum validity of any measure is the square root of its reliability"

- only two studies published on test-retest of infant speech perception measures, & second contains data on first. Cristia et al use meta-analytic method and find, across 12 studies, weighted median r of zero

- this means any correlation/intervention work using these measures is suspect because they have virtually no good reliability

- correlations within each study make sense, but do not capture same information as correlations collapsing across studies; or considering studies as structured

- current trends in genetics & brain studies pushing for mega-, over meta-, analyses because structured sources of variance can be better accounted for, and analyses may have more power

- here, we reanalyze Cristia's data to revisit the question of reliability, and ask

- in mega-analysis, do you also find basically no prediction of test2 from test1?

- how should structure be accounted for - are studies all different from each other?

- what happens if you only have some data from some studies – picked at random? (assuming original authors do not withhold the data for any reason that is related to the data itself)

- and if you only have data from large studies? (authors who ran more babies are more motivated to share)

- and if you only have data from studies with large main effects? (defined as the average between effect at test1 and effect at test2 – intuition is that authors with strong effects believe their data more)

- and if you only have data from studies with large test-retest correlations? (idea: authors who find reliability more likely to share raw data)

Methods

Very short because we refer to previous paper for full description of experiments

table of experiments: short names, short description, N of children, mean age

We got data from osf, using R, this paper uses Rmd in RStudio & papaja for increased reproducibility.

Results

- how should structure be accounted for - are studies all different from each other?

explain use of AIC to compare models, using also conceptual reasons to group studies – ending up with 5 clusters

- in mega-analysis, do you also find basically no prediction of test2 from test1?

  no, we get something pretty different. Explain why

- use a graph to represent all of the hybrid results (max 4k words!)

Discussion

- under what conditions can we trust infant speech perception measures of individual variation?
- we recommend mega- over meta-analysis
- explain under what conditions this holds, and when mega-analysis provides biased view of data

References