

1 Estimating age-related change in infants' linguistic and cognitive development using
2 (meta-)meta-analysis

3 Anjie Cao¹ & Michael C. Frank¹

4 ¹ Stanford University

5 ² Konstanz Business School

6 Author Note

7 Add complete departmental affiliations for each author here. Each new line herein
8 must be indented, like this line.

9 Enter author note here.

10 The authors made the following contributions. Anjie Cao: Conceptualization,
11 Writing - Original Draft Preparation, Writing - Review & Editing; Michael C. Frank:
12 Writing - Review & Editing, Supervision.

13 Correspondence concerning this article should be addressed to Anjie Cao, 450 Jane
14 Stanford Way, Stanford, CA 94305. E-mail: anjiecao@stanford.edu

Abstract

Developmental psychology focuses on how psychological phenomena emerge with age. In cognitive development research, however, the specifics of this emergence is often underspecified. Researchers often provisionally assume linear growth by including chronological age as a predictor in regression models. In this work, we aim to evaluate this assumption by examining the functional form of age trajectories across 24 phenomena in early linguistic and cognitive development using (meta-)meta-analysis. Surprisingly, for most meta-analyses, the effect size for the phenomenon was relatively constant throughout development. We investigated four possible hypotheses explaining this pattern: (1) age-related selection bias against younger infants; (2) methodological adaptation for older infants; (3) change in only a subset of conditions; and (4) positive growth only after infancy. None of these explained the lack of age-related growth in most datasets. Our work challenges the assumption of linear growth in early cognitive development and suggests the importance of uniform measurement across children of different ages.

Keywords: keywords

Word count: X

Estimating age-related change in infants' linguistic and cognitive development using
(meta-)meta-analysis

Developmental psychology focuses on how psychological constructs change with age. Throughout the years, many theories have been proposed to characterize and explain how and why developmental changes happen (Bronfenbrenner, 1977; Carey, 2009; Elman, 1996; Flavell, 1994; e.g., Piaget, 1971; Thelen & Smith, 2007). Among these theories, one common assumption is that skills increase with age (positive change assumption): children get better as they get older. Often, researchers treat age as a predictor in linear regression models, and therefore implicitly assume that the constructs of interests follow a linear trajectory (Lindenberger & Pötter, 1998). While both assumptions are widely adopted, especially in early cognitive and language development, their validity is rarely tested.

One common approach to evaluating the functional form of age-related changes is through longitudinal studies. Measurements of psychological constructs, when tracked longitudinally, often reveal the age trajectories that violate the linearity assumption. For instance, a longitudinal study that follows the development of executive function (EF) from 3 to 5 years-old using a battery of EF tasks show that EF follows a non-linear trajectory over age (Johansson, Marciszko, Brocki, & Bohlin, 2016). Similarly, vocabulary in early childhood, measured by MacArthur-Bates Communicative Development Inventories, also follows the exponential trend rather than the linear trend (Frank, Braginsky, Yurovsky, & Marchman, 2021). In many domains with established measurements, longitudinal research has been used to characterize the functional form of the development (Adolph, Robinson, Young, & Gill-Alvarez, 2008; Cole, Loughheed, Chow, & Ram, 2020; Karlberg, Engström, Karlberg, & Fryer, 1987; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; Tilling, Macdonald-Wallis, Lawlor, Hughes, & Howe, 2014). However, longitudinal methods are more rarely applied to experimental studies that identify proposed mechanisms underlying development.

Many important findings in early language and cognitive development are primarily attested in cross-sectional experimental studies. For example, in the language learning domain, many studies have targeted specific mechanisms proposed to underlie how infants acquire specific facets of language. Constructs such as mutual exclusivity (Markman & Wachtel, 1988), statistical learning (Saffran, Aslin, & Newport, 1996), syntactic bootstrapping (Naigles, 1990) and so on, are all attested through decades of experimental evidence acquired through cross-sectional studies. These works are critical to test the causal mechanisms underlying age-related changes, but they are rarely measured in samples with sufficient size and age variation to test the positive change assumption or the assumption of linearity (cf. Frank et al., 2017). In an ideal world, one would run those experiments longitudinally on a large, diverse sample. In practice, this goal is difficult to achieve due to the constraints on both time and financial resources. As a result, the functional forms of age-related changes in critical constructs remain poorly understood.

To address this issue, we turned to meta-analysis. Meta-analysis is a statistical method to aggregate evidence across studies quantitatively. This approach has been widely adopted in many disciplines and subfields, including developmental psychology (Doebel & Zelazo, 2015; e.g. Hyde, 1984; Letourneau, Duffett-Leger, Levac, Watson, & Young-Morris, 2013). Compared to the single study approach, meta-analysis has several advantages. First, it allows us to examine the robustness of the phenomena documented in the literature. By combining results from multiple studies, meta-analysis enhances the statistical power to detect effects that might be too small to identify in individual studies. Second, meta-analysis provides a framework for assessing the consistency of research findings across different contexts (Borenstein, Hedges, Higgins, & Rothstein, 2021; Egger, Smith, & Phillips, 1997). Further, pooling across developmental studies with different cross-sectional samples may yield sufficient variation to explore the functional form of age-related change with greater precision than individual studies.

In this work, we aim to leverage meta-analysis to examine the shape of the

developmental trajectory in key constructs in infant language and cognitive development. Specifically, we use existing meta-analyses from Metalab (<https://langcog.github.io/metalab/>), a platform that hosts community-augmented meta-analyses. Metalab was established to provide dynamic databases publicly available to all researchers (Bergmann et al., 2018). Researchers can deposit their meta-analysis dataset in the platform, and they can also use the dataset for custom analyses (e.g. Cao, Lewis, & Frank, 2023; Lewis et al., 2016). To this date, Metalab contains #FIXME effect sizes from #FIXME different meta-analysis, spanning different areas of developmental psychology. This resource allows us to examine the suitability of meta-analysis as a tool to characterize developmental trajectory – and if suitable, provides insights into how these key constructs develop across the early months of childhood.

We acknowledge at the outset that meta-analysis has significant limitations. The quality of a meta-analysis is necessarily constrained by the quality of the existing studies (Simonsohn, Simmons, & Nelson, 2022). If the studies being aggregated are flawed, the conclusions drawn from the meta-analysis will also be questionable. Moreover, one significant issue in interpreting meta-analysis is the heterogeneity among studies. Heterogeneity refers to the variability in study participants, interventions, outcomes, and methodologies. This diversity can make it challenging to aggregate results meaningfully, because differences between studies may reflect true variation in effects rather than a singular underlying effect size (Fletcher, 2007; Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006; Thompson & Sharp, 1999). Critically, understanding the source of heterogeneity often requires detailed coding of the potential moderators; this process is frequently hampered by the inadequate reporting standards prevalent in psychological literature, which often leaves essential information for coding these moderators absent (Nicholson, Deboeck, & Howard, 2017; Publications, Journal, & Standards, 2008). In other words, whether meta-analysis can provide insights into the nature of age-related change is dependent upon the quality of the existing literature.

This paper is organized as follows. In the first section, we provide an overview on the estimated general shape of age-related change across the datasets in Metalab. To preview our findings, we found that most datasets showed relatively constant effect size across age. This finding challenges the commonly held linearity assumption and the positive increase assumption. In the second section, we test four hypotheses on why the current meta-analyses failed to reveal age-related changes: (1) age-related selection bias against younger infants; (2) methodological adaptation for older infants; (3) change in only a subset of conditions; and (4) positive growth only after infancy. We found that none of the four explanations provided a satisfying explanation for the lack of age-related change in most meta-analyses.

Datasets

Datasets were retrieved from Metalab. As of February 2024, Metalab hosted 32 datasets in total, with research areas ranging from language learning to cognitive development. All datasets included effect size estimates converted to standardized mean difference (SMD; also known as Cohen's d) as well as estimates of effect size variance and a variety of other moderators (e.g., average age of participants) provided by the contributors. There were 2 desiderata for the datasets to be included in the final analysis:

1. The dataset must describe an experimental (non-correlational) effect that uses behavioral measures, and
2. For a dataset that has already been published, the meta-analytic effect reported in the published form must not be null (i.e., must be significantly different than zero).

Five datasets did not meet the first desideratum (*Pointing and vocabulary (concurrent)*; *Pointing and vocabulary (longitudinal)*; *Video deficit*; *Symbolic play*; *Word segmentation (neuro)*), and one dataset did not meet the second desideratum (*Phonotactic learning*). These datasets were not included in the analysis.

For the remaining 26 datasets, we made the following modifications. Following the organization in the original meta-analysis (Gasparini, Langus, Tsuji, & Boll-Avetisyan, 2021), we separated the Language discrimination and preference dataset into two datasets, one for discrimination and one for preference. We also combined two pairs of datasets because they were testing the same experimental effects: *Gaze following (live)* and *Gaze following (video)* was combined into *Gaze following (combined)*; *Function word segmentation* and *Word segmentation (behavioral)* was combined into *Word segmentation (combined)*. We also replaced the *Infant directed speech preference* dataset with a more up-to-date version reported in Zettersten et al. (2023).

To make the comparison more equivalent to each other, we would run models with the same random effect structure specifications across all datasets. To achieve this goal, we recoded the relevant grouping variables in the datasets with missing grouping variables.

Since we were mostly interested in the age trajectory of these constructs in early childhood, we further trimmed the datasets to include only effect sizes from participants under 36 months of age. This decision did not qualitatively affect our findings as most datasets did not include data above age 36 months. The final analysis included 25 datasets in total. Table 1 presented the names of all the datasets, along with the number of effect sizes and participants included for each dataset.

Methods

All of the statistical analyses were conducted in R. Meta-analytic models were fit using the metafor package (Viechtbauer, 2010). This was an exploratory study in which no hypotheses were pre-registered.

For each dataset, we considered four functional forms as possible candidates for the shape of the developmental trajectory: linear, logarithmic, quadratic, and constant. A linear form is the most common assumption in the literature, whereas logarithmic and

quadratic were chosen to represent sublinear growth and superlinear growth, respectively. The constant form served as a baseline null hypothesis for the other alternative growth patterns. Although other, more complex growth patterns are of course possible, we opted to compare these forms as a first pass. Note that the constant model includes one parameter (an intercept), linear and logarithmic models include two parameters (an intercept and a slope), and the quadratic model includes three parameters (intercept, slope, and quadratic growth term).

For all analyses, we fit multilevel random-effects meta-regression models using nested random intercepts to account for both the testing of individual samples in multiple conditions (e.g., in a between-participants design) and multiple studies within a single paper. Meta-regression models predicted effect sizes (standardized mean difference / Cohen's d) with mean age in months in different functional forms. We fit four meta-regression models in total for each dataset.

Results

Model comparison. Our initial goal was to compare the fit of models with different functional forms for each meta-analysis. Because models differed in their complexity (number of parameters), we extracted the corrected AIC (AICc) for each model. The model with the lowest AICc was considered the baseline model, and all the remaining models were compared against the baseline. The remaining model each received a Δ_{AIC} , which was the difference between the AIC of the model and the AIC of the baseline model. Following standard convention, we treated $\Delta_{AIC} > 4$ as the statistical significance threshold (Burnham & Anderson, 2004). A baseline model was significantly better than an alternative model if and only if the alternative model had $\Delta_{AIC} > 4$.

Surprisingly, the four functional forms could not be meaningfully distinguished in 19 out of 25 datasets.. (This situation typically arises because the data are constant and

hence more complex models with zero parameters fit the data equally well ¹). The remaining 6 datasets yielded meaningful contrasts between different functional forms, but the linear form was not the best-fitting form for any dataset. Table 2 shows the model comparison results for each dataset. Figure 1 shows the prediction of each functional form.

Linearity and Positive Increase Assumption. One limitation of the model comparison approach is that it does not quantify growth over time. To further examine the positive increase assumption, we estimated linear meta-regression models and examined the estimates on the age predictor. We found that the slope estimate for age was not significantly different from zero in the majority of the datasets (17/25; Fig 2).

Discussion

We conducted model comparisons to assess the functional forms of age-related change across 25 datasets. Four functional forms—Logarithmic, Linear, Quadratic, and Constant—were largely indistinguishable within most datasets. Notably, in datasets where contrasts were meaningful, linear models received no support, challenging the prevalent linearity assumption for early linguistic and cognitive development. Further, we only detected any positive growth in 8/25 meta-analyses. Past work has successfully revealed age-related changes using meta-analysis (e.g. Best & Charness, 2015; McCartney, Harris, & Bernieri, 1990; Sugden & Marquis, 2017). But in most datasets that we have considered, effect size does not increase with age. Why?

¹ In the situation of a completely constant pattern of effects across age, the maximal difference in model fit would be an AICc of exactly 4 between the constant and quadratic model, reflecting a two-parameter difference.

References

- Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, 115(3), 527.
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996–2009.
- Best, R., & Charness, N. (2015). Age differences in the effect of framing on risky choice: A meta-analysis. *Psychology and Aging*, 30(3), 688.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.
- Cao, A., Lewis, M., & Frank, M. C. (2023). A synthesis of early cognitive and language development using (meta-) meta-analysis. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45.
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Cole, P. M., Loughheed, J. P., Chow, S.-M., & Ram, N. (2020). Development of emotion regulation dynamics across early childhood: A multiple time-scale approach. *Affective Science*, 1, 28–41.
- Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the dimensional change card sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, 38, 241–268.
- Egger, M., Smith, G. D., & Phillips, A. N. (1997). Meta-analysis: Principles and procedures. *Bmj*, 315(7121), 1533–1537.

- Elman, J. L. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT press.
- Flavell, J. H. (1994). *Cognitive development: Past, present, and future*.
- Fletcher, J. (2007). What is heterogeneity and is it important? *Bmj*, 334(7584), 94–96.
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., et al.others. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, 22(4), 421–435.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.
- Gasparini, L., Langus, A., Tsuji, S., & Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants’ language discrimination abilities: A meta-analysis. *Cognition*, 213, 104757.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or i^2 index? *Psychological Methods*, 11(2), 193.
- Hyde, J. S. (1984). How large are gender differences in aggression? A developmental meta-analysis. *Developmental Psychology*, 20(4), 722.
- Johansson, M., Marciszko, C., Brocki, K., & Bohlin, G. (2016). Individual differences in early executive functions: A longitudinal study from 12 to 36 months. *Infant and Child Development*, 25(6), 533–549.
- Karlberg, J., Engström, I., Karlberg, P., & Fryer, J. G. (1987). Analysis of linear growth using a mathematical model: I. From birth to three years. *Acta Paediatrica*, 76(3), 478–488.
- Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. *Journal of Emotional*

and *Behavioral Disorders*, 21(3), 211–224.

Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., et al.

(2016). *A quantitative synthesis of early language acquisition using meta-analysis*.

Lindenberger, U., & Pötter, U. (1998). The complex nature of unique and shared effects in

hierarchical linear regression: Implications for developmental psychology. *Psychological*

Methods, 3(2), 218.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to

constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009).

Modeling life-span growth curves of cognition using longitudinal data with multiple

samples and changing scales of measurement. *Psychological Methods*, 14(2), 126.

McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A

developmental meta-analysis of twin studies. *Psychological Bulletin*, 107(2), 226.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*,

17(2), 357–374.

Nicholson, J. S., Deboeck, P. R., & Howard, W. (2017). Attrition in developmental

psychology: A review of modern missing data reporting and practices. *International*

Journal of Behavioral Development, 41(1), 143–153.

Piaget, J. (1971). *The theory of stages in cognitive development*.

Publications, A., Journal, C. B. W. G. on, & Standards, A. R. (2008). Reporting standards

for research in psychology: Why do we need them? What might they be? *The*

American Psychologist, 63(9), 839.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old

infants. *Science*, 274(5294), 1926–1928.

Simonsohn, U., Simmons, J., & Nelson, L. D. (2022). Above averaging in literature

reviews. *Nature Reviews Psychology*, 1(10), 551–552.

Sugden, N. A., & Marquis, A. R. (2017). Meta-analytic review of the development of face

discrimination in infancy: Face race, face gender, infant age, and methodology
moderate face discrimination. *Psychological Bulletin*, 143(11), 1201.

Thelen, E., & Smith, L. B. (2007). Dynamic systems theories. *Handbook of Child
Psychology*, 1.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A
comparison of methods. *Statistics in Medicine*, 18(20), 2693–2708.

Tilling, K., Macdonald-Wallis, C., Lawlor, D. A., Hughes, R. A., & Howe, L. D. (2014).
Modelling childhood growth using fractional polynomials and linear splines. *Annals of
Nutrition and Metabolism*, 65(2-3), 129–138.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal
of Statistical Software*, 36(3), 1–48.

Zettersten, M., Cox, C. M. M., Bergmann, C., Tsui, A., Soderstrom, M., Mayor, J., et
al.others. (2023). *Evidence for infant-directed speech preference is consistent across
large-scale, multi-site replication and meta-analysis.*