Title

A data integration method for new advances in development cognitive neuroscience.

Authors

Canada, Kelsey
Riggins, Tracy
Ghetti, Simona
et al.

Publication Date

2024-12-01

DOI

10.1016/j.dcn.2024.101475

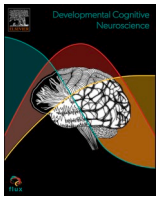
Peer reviewed

Developmental Cognitive Neuroscience 70 (2024) 101475

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn



A data integration method for new advances in development cognitive neuroscience

Kelsey L. Canada [a,*], Tracy Riggins [b], Simona Ghetti [c,d], Noa Ofen [a,e,f], Ana.M. Daugherty [a,g,h,**]

[a] Institute of Gerontology, Wayne State University, Detroit, MI, USA
[b] Department of Psychology, University of Maryland, College Park, MD, USA
[c] Department of Psychology, University of California, Davis, CA, USA
[d] Center for Mind and Brain, University of California, Davis, CA, USA
[e] Center for Vital Longevity, University of Texas at Dallas, Dallas, TX, USA
[f] Department of Psychology, School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA
[g] Department of Psychology, Wayne State University, Detroit, MI, USA
[h] Michigan Alzheimer's Disease Research Center, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:
Integrative data analysis
Secondary data analysis
Neuroimaging
Development
Hippocampal subfields

ABSTRACT

Combining existing datasets to investigate key questions in developmental cognitive neuroscience brings exciting opportunities and unique challenges. However, many data pooling methods require identical or harmonized methodologies that are often not feasible. We propose Integrative Data Analysis (IDA) as a promising framework to advance developmental cognitive neuroscience with secondary data analysis. IDA serves to test hypotheses by combining data of the same construct from commensurate (but not identical) measures. To overcome idiosyncrasies of neuroimaging data, IDA explicitly evaluates if measures across studies assess the same construct. Moreover, IDA allows investigators to examine meaningful individual variability by de-confounding source-specific differences. To demonstrate IDA's potential, we explain foundational concepts, outline necessary steps, and apply IDA to volumetric measures of hippocampal subfields from 443 4- to 17-year-olds across three independent studies. We identified commensurate measures of Cornu Ammonis (CA) 1, dentate gyrus (DG)/CA3, and Subiculum (Sub). Model testing supported use of IDA to create IDA factor scores. We found age-related differences in DG/CA3, not but CA1 and Sub volume in the integrated dataset. By successfully demonstrating IDA, our hope is that future innovations come from the combination of existing neuroimaging data to create representative integrated samples when testing critical developmental questions.

1. Introduction

As multi-site studies produce large-scale data to investigate key questions in developmental cognitive neuroscience, combining existing datasets brings exciting opportunity and unique challenges. Big Data methods have become popular to analyze data from population-based samples, but they have several requirements that limit their application, including identical data collection protocols and analytic approaches across contributing datasets (e.g., Marzi et al., 2024; Tozzi et al., 2021). However, identical assessments are not always feasible for different samples across development. Applied to neuroimaging data, forced harmonization of parameters is not a perfect solution to site-specific and sample-specific variability (Marzi et al., 2024). The high cost associated with large-scale neuroimaging studies compels the research community to leverage existing datasets despite their differences, underscoring the need for alternative approaches that might support analysis of combined data even when data collection protocols are not identical or harmonized from the start. Integrative Data Analysis (IDA) is a promising latent modeling approach to address this need that allows combining data from multiple studies with similar (but not necessarily identical) assessments (Curran et al., 2016; Curran and Hussong, 2009; Hussong et al., 2013). The IDA framework explicitly evaluates pooled measurements to ensure valid construct estimation, which can overcome idiosyncrasies of neuroimaging data from differences in acquisition parameters, scanning environments, and anatomical segmentation protocols.

* Corresponding author.
** Correspondence to: Institute of Gerontology, Wayne State University, 87 E. Ferry St., Detroit, MI 48202, USA.
E-mail addresses: kcanada@wayne.edu (K.L. Canada), ana.daugherty@wayne.edu (Ana.M. Daugherty).

https://doi.org/10.1016/j.dcn.2024.101475
Received 16 June 2024; Received in revised form 13 September 2024; Accepted 4 November 2024
Available online 9 November 2024

In secondary data analysis, the data can be used to replicate or extend previous findings and potentially address questions not tested in the original studies from which the samples were drawn. Developmental cognitive neuroscientists can cover a larger range of development periods and include participants from across the lifespan in a more time- and cost-effective manner than a single-site study can. Existing datasets have the clear benefit of having already been collected but are often viewed with the limitation to innovation as they are restricted to the specifics of a single sample. With IDA, investigators can leverage multiple datasets to create an aggregated large sample, with greater demographic representation, and even with longitudinal measures if constructs of interest (not specific measures) can be identified within the sets and reasonable psychometric assumptions are met. The era of open data and cumulative science approaches provide a way to generate new hypotheses, address innovative research questions in a feasible time frame with minimal infrastructure investment, while setting the foundation for planning and optimization of future studies.

1.1. Approaches to cumulative science

Interest in combining data across studies to address developmental questions is not new (Greenhoot and Dowsett, 2012). At the core of developmental cognitive neuroscience are questions of individual differences in development and their contributing factors. Measured variability is the sum of two parts: true individual differences in the population and measurement error (Schmidt et al., 2003). When combining data from multiple sources, measurement error that is source-specific and correlates with true individual differences cannot be easily filtered from the analysis. The well-known idiom, "Don't throw the baby out with the bath water" (in the original German: 1512, Murner) aptly illustrates removing measurement error at the expense of meaningful variability. This has implications for our ability to draw meaningful inferences from the data; either too much measurement error remains and systematically biases inferences, or we have over-corrected and diminished the individual differences of interest. One logical response is to standardize measurement so to reduce potential source-specific differences at the start—all baths are the same size for all babies, and so we can perfectly filter the same volume of water. In practice this often falls short because even standardized measurement is imperfect and rarely does one-size fit all sub-groups. While addressing one problem, forced standardized methods often reduce sensitivity to true variability and diversity of samples, collectively obscuring the true individual differences in the population and generalization back to the individual. Not all babies are the same so not all baths should be identical, and yet our research still needs to collectively describe the experience of all babies without murky water. In this section, we briefly review different approaches available to analyze data from different sources that were not intentionally collected together so to highlight the value of IDA in the field of development cognitive neuroscience.

1.1.1. Meta-analysis

Rather than standardizing methods in data collection, meta-analytic techniques provide a means of integrating information across studies through the analysis of individual study summary statistics. This approach enables the integration of findings from a large number of studies into one analysis and it is minimally reliant on data sharing (Chan and Arvey, 2012). However, because meta-analysis does not integrate raw data, researchers are limited to addressing questions related to effects estimated and participant data reported in each individual study. The use of summary information also limits researchers' ability to explicitly test equivalence of measurements across studies and account for different sources of variance, which are the foundational assumptions for inference from combined analysis. Combining summary statistics from samples with different diversity representation improves external validity over any one study with limited diversity; however, the individual study estimates are sample-specific and thus limited in that regard. Potential idiosyncrasies of the study protocols or sample representation can be coded and tested as covariates or moderators in the analysis. This highlights the use of meta-analysis to test new hypotheses of conditional effects, but as a solution to correlated measurement error it shares the same limitations as other regression covariate approaches. Moreover, meta-analytic approaches often rely on published findings. While some researchers solicit results from unpublished data, the "file-drawer problem" may pose a practical problem if not considered. If there is bias identified (Lin and Chu, 2018) in the included studies, and null findings are not accounted for, it is difficult to draw meaningful inference from the results of the meta-analysis.

Overall, meta-analysis is particularly effective for summarizing the current state of the science, and to highlight trends and future directions for further investment. However, it has significant limitations as a technique to test new research questions with existing data.

1.1.2. Mega-analysis: regression models with adjustment by covariate or clustered data

Methods that apply regression to combine data from different sources are organized under the umbrella of mega-analysis. In mega-analysis approaches, individual-level data pooled from multiple studies are analyzed simultaneously. Often, the original individual data are shared; when applied to neuroimaging, for example, the raw image files would be shared and processed using the same parameters (Bockholt et al., 2010; De Wit et al., 2014). In mega-analysis, measurements are required to be identical (Hofer and Piccinin, 2009) and are combined into a common dataset (Boedhoe et al., 2019). Regression approaches that are applied (i.e., multiple regression, analysis of covariance, generalized estimating equations) provide parameter estimates and standard errors pooled across studies as a grand, fixed effect estimate (Hao et al., 2023). If all measurement error were identical across studies, this approach would be effective; however, this is a substantial statistical assumption. Pooled standard error typically includes multiple sources of variance, not all identical across data sources or individuals in any given sample, thus entangling it with the meaningful individual differences of interest. Under the umbrella of mega-analysis, mixed effects models allow additional options to account for individual- and study-related factors as either fixed or random effects.

Here we provide a brief summary of three general approaches that could be taken to mega-analysis. As a starting point for discussion, one approach in concept could be that data are pooled and treated as if belonging to a single study and no statistical accounting of source is made. This approach would hold the loftiest assumptions that all measurements were perfectly reliable, and all variances should be treated as true individual differences. However, to our knowledge, this is rarely done in practice and existing literature has highlighted the risks of not accounting for the source of data when combining information into a single data set (Bayer et al., 2022). Second, a covariate that codes for source is included in the regression or in its extension of ANCOVA. The regression weights are independent effects, fully adjusted for the covariate, thus any predictor effects can be interpreted as wholly separate from source. In this approach, the challenge presents with the common scenario of measurement error that is correlated with sample demographics or predictors of interest; that portion of the true variability is then removed along with the measurement error. For example, in a combined dataset if only one site collected data from 5–8-year-olds, age would be confounded with site; thus, when site is entered as a covariate it will reduce the independent estimate of age effects. This challenge becomes more insidious when we consider combining convenience samples, and representation of racial/ethnic diversity, socioeconomic factors, and environmental exposures that are specific to study sites. In addition, the individual data are clustered within source (as in the example, children within site) and this typically will not meet the assumption of independence of residuals that is required for unbiased estimates with ordinary least squares regression.

The alternative third approach treats data as clustered within source and auto-correlations among residuals are adjusted via generalized estimating equations or mixed effects methods (Boedhoe et al., 2019). This approach allows a great deal of flexibility, especially in situations where the study design is unbalanced, requiring a level of data management that may be easier to implement (Curran, 2003). Further, this approach has been applied successfully to neuroimaging data (e.g., Boedhoe et al., 2019). Overall, because covariance structures among individual-level data and higher-level data (e.g., study site) are disaggregated simultaneously, these methods are exceptionally effective as a correction for non-independence. Previous research with data simulation has effectively demonstrated that mixed effects modeling outperforms the other regression approaches we review in reducing bias in the point estimates, appropriately controls for type I error within the model, and improves inferences on the source of variability in IDA (Wilcox and Wang, 2023).

Similar benefits extend to IDA within the structural equation model (SEM) framework to model individual-level and study source-specific heterogeneity. Under certain constraints, mixed effect models and SEM can provide equivalent estimates with nested data, including options for fixed and random effects, and longitudinal analysis (Curran, 2003). The SEM framework to support IDA provides additional opportunities to mixed effect models that will be appealing for some hypothesis tests in developmental cognitive neuroscience. For example, if the data to be integrated is an outcome measure in the hypothesis test (e.g., brain volume predicted by age), mixed effect models and IDA supported by SEM are expected to perform similarly. But, if the integrated factor score is to be used as a predictor of other integrated factor scores (e.g., brain volume predicting cognition), the SEM framework supporting IDA provides more flexibility to develop the factor specifications and the subsequent hypothesis models. Additionally, in SEM, factor loadings of separate indicators need not be identical, nor residual variances equated across indicators, and study-by-covariate interactions can be tested (Curran, 2003). Further, through the use of latent factors with multiple indicators, SEM has the ability to estimate parameters independent of measurement error estimates. In the context of IDA, this is a key strength that provides the means to assess meaningful developmental differences in similar, albeit not identical, measures (Curran, 2003).

As a set of methods, mega-analysis is appealing for using familiar analytic strategies from single-study designs, but an investigator must consider the limitations of likely over-correcting for measurement error and reducing sensitivity to specific effects of interest when assumptions of equivalent factor loadings or residual variances cannot be met.

1.1.3. Principal component scores

Principal component analysis is an alternative to mega-analysis with a single measure, as it creates new individual scores that combine multiple measures of a construct with the opportunity to aggregate data across different sources. This approach reduces the number of variables in the combined dataset into a user-defined number of principal component (PC) scores via linear combination of the original variables (Jolliffe, 2002). The PC scores are intended to maximize the amount of variance retained from the original variables while remaining uncorrelated with each other, and the unit of measurement for each variable becomes standardized via a correlation or covariance matrix. Although all to-be-combined samples must have the same assessments collected, the method does not require the measured variance to be identical across data sources to compute a PC score that can be directly compared in a combined analysis. Despite its advantages, the PC score can be difficult to interpret for at least two reasons. First, the PC procedure is designed for data reduction and so it does not independently estimate measurement error apart from the true individual differences that are assumed for interpretation (Raykov et al., 2017); a PC score will contain correlated error. Second, PC scores are descriptive rather than inferential (Jolliffe and Cadima, 2016). As an example, applied to neuroimaging data, gray-white matter contrast on images is affected by true developmental changes in the brain and noise from imaging parameters; the PC score cannot differentiate these sources when there are site differences in the distribution of age and without independent measures of imaging noise by site. The noted limitations reduce construct validity of the PC score in an effort to mitigate systematic measurement error, subsequently limiting the ability to make meaningful inferences about developmental constructs.

1.1.4. Machine learning

In recent years, machine learning (ML) approaches have gained traction in cognitive neuroscience research (Rosenberg et al., 2018). In general these approaches use neuroimaging data in tandem with prediction algorithms to account for correlations among features of data (Chen et al., 2022; Jollans et al., 2019; Marzi et al., 2024). Often, these approaches aim to harmonize neuroimaging data acquired from different sources by accounting for site-related differences in order to increase accuracy in predicted outcomes (Jollans et al., 2019). See Cohen et al. (2017) and Davatzikos (2019) for commentaries and reviews of the rapid expansion of these methods. A common limitation of these approaches, and most others we have reviewed, is forced harmonization and the requirement of identical scan parameters. For example, ML was applied to data from 36 neuroimaging studies to remove the effect of site from the covariance structure using a PC approach. Although successful in mitigating differences between sites, the authors noted an overcorrection of the data. While ML can provide exciting avenues for researchers when combining data, this study highlights the risk of ML approaches removing meaningful individual-level differences that correlate with site, including demographic and socioeconomic characteristics, and this risk is a warranted consideration of the method (Chen et al., 2022).

1.1.5. Summary discussion of limitations

An exhaustive treatment of each class of these statistical approaches is beyond the purview of this paper. Our intention is to summarize common approaches to cumulative data analysis in developmental research that have shared limitations to addressing the challenges of combining existing, multi-study data that were not intentionally collected together. The greatest limitation of these approaches is diminished sensitivity to detect individual differences in development in the procedure to account for source-specific measurement error. This can reduce statistical power and external validity of the results. The second noted limitation was requiring identical measurements, or harmonization, as a starting condition to combine original source data. Even with substantial amounts of open data sources, this requirement makes many research questions intractable and aggregated large sample sizes less feasible.

In developmental research, the restriction of using identical measures also hampers the ability to investigate constructs across the lifespan, as many tasks are modified for developmental stage. For example, populations that require specialized sequences, such as shorter scan times for very young children, could not be combined with standard scans in older participants. Further, these approaches do not provide straightforward means to evaluate measurement invariance over time in cumulative analysis or valid estimates with partially missing data at the individual-level, both necessary features for longitudinal data modeling, which is central to investigations of developmental change. The IDA framework has been developed to address these limitations (Lambert et al., 2002; McArdle et al., 2002).

1.2. Integrative data analysis

IDA is an applied latent modeling technique that pools different measurements of the same construct across samples into a combined dataset (Curran et al., 2008, 2016; Curran and Hussong, 2009; Hussong et al., 2013). As a latent modeling method, all factor scores include

independent estimation of measurement error apart from the individual score, an advantage to the main critique of PC scores. The benefits of latent factor scores combined with larger integrated developmental samples spanning greater developmental periods, across different geographic regions, and with greater diversity representation collectively improve the validity of the IDA as compared to analysis of any single study (Hussong et al., 2013). IDA has advantages over meta-analysis by testing new hypotheses from aggregated data that go beyond the constraints of any single study, such as questions related to diversity and differential development effects (Hussong et al., 2013).

A unique feature of IDA that sets it apart from the other reviewed methods is that it does not require identical measures as a starting condition. A common construct is required to ensure valid interpretation, but otherwise different assessments and protocols can be applied. This is a particular strength when applied to neuroimaging data. In addition, this approach does not require raw neuroimaging data, but instead can rely on derivative measures that help ensure no privacy issues arise in the sharing of data (White et al., 2022). Given that developmental research typically employs assessments designed to be tailored to the age group of interest, the ability to integrate data across study sites without the restriction of identical measurement creates an exciting opportunity to study wider developmental periods and new questions on individual-level contextual factors.

In this manuscript, we provide a proof of concept demonstrating the feasibility of applying IDA in developmental cognitive neuroscience using neuroimaging studies of hippocampal subfield volumes conducted at multiple sites. Descriptive statistics were calculated using IBM SPSS Statistics Version 28 (Chicago, IL) (Step 2B). All other analyses in our demonstration of IDA were conducted using Mplus 8.10 (Muthén and Muthén, 2017). However, any software suitable for SEM can be used (e.g., lavaan; Rosseel, 2012). To support the application of IDA to additional areas of developmental cognitive neuroscience, we provide a detailed description of the IDA method applied to neuroimaging data and outline the conceptual and statistical modeling steps that other

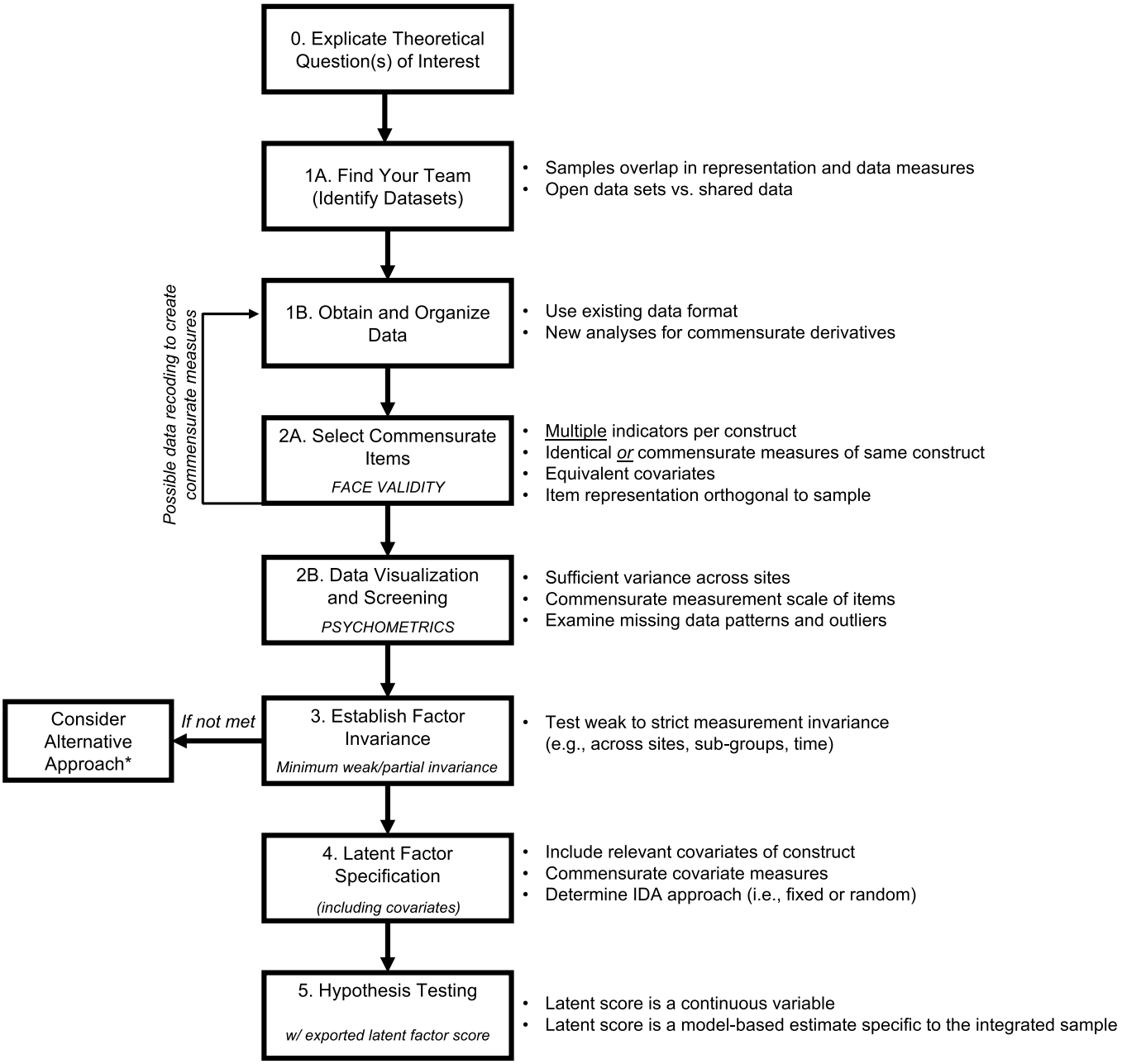


Fig. 1. Schematic figure depicting the process of Integrative Data Analysis (IDA).

scientists can apply to their research questions.

1.3. Summary approach and methods

We apply the IDA framework to pool three existing pediatric samples examining hippocampal subfields into a combined dataset that is geographically and demographically diverse, to demonstrate good statistical power and validity of results. As of the time of this writing, the field of research on *in vivo* hippocampal subfields has many variations of segmentation protocols that have many similarities but also substantial anatomical differences. Because there is currently no consensus on definitions on the boundaries of human hippocampal subfields for *in vivo* imaging, and all protocols have undergone validation with different histological reference materials, the various protocols are comparably valid. With variations in neuroimaging parameters, scanning environments, and anatomical segmentation protocols, research questions using existing hippocampal subfield data should not be simply merged for a mega-analysis. We considered this an interesting opportunity to test IDA with measures of hippocampal subfields in which researchers chose imaging parameters and segmentation protocols that best suited their original research questions with their sample. Instead of "throwing out" different protocols, forcing adoption of a single protocol, or re-processing all data with one segmentation protocol that may be sub-optimal for an individual site sample, IDA leverages all data in combination to extract shared variance indicative of true individual differences in hippocampal subfield volumes apart from source-specific idiosyncrasies. Here we will demonstrate the use of this method to create estimates of developmental differences with independent estimation of measurement error from similar, but not identical, measurements of hippocampal subfields. Fig. 1

1.4. Step 0. Explicate theoretical question(s) of interest

The critical starting point of the IDA framework is determining if the research questions to be tested are suitable. As with other approaches to secondary data analysis, researchers are limited to data already collected. Existing datasets will vary in the sample characteristics, constructs assessed, and measures used to assess the constructs. Thus, the first step in evaluating the appropriateness of this approach is to determine whether existing datasets provide a good match to an investigator's research questions.

In the applied example, hippocampal subfield volumes are a key measurement in current investigation of memory development, yet these structures are measured using different definitions and acquisition parameters across studies—an example of a common construct but non-identical assessment. Work in non-human primates and post-mortem human samples suggests different developmental trajectories of hippocampal subfields: earliest maturation of the subiculum (Sub), followed by Cornu Ammonis (CA) 1 and 2, and dentate gyrus (DG) as the most protracted (Lavenex and Banta Lavenex, 2013; Seress, 2001; Seress and Ábrahám., 2008). Most evidence using high-resolution imaging data of the Hippocampus comes from cross-sectional studies and is inconsistent. For example, studies report positive age-volume relations for DG (Canada et al., 2021; Schlichting et al., 2017) and Sub (Canada et al., 2021); negative age-volume relations for DG/CA3 (Daugherty et al., 2016), CA1 (Daugherty et al., 2016), and Sub (Schlichting et al., 2017); and null effects for CA1 (Canada et al., 2021; Riggins et al., 2018; Schlichting et al., 2017) and Sub (Daugherty et al., 2016; Lee et al., 2014; Riggins et al., 2018). Extant findings likely vary because individual studies are limited in representation and sample size, and differ in the age-range examined, methodological approach, and atlases used to define hippocampal volumes. Inconsistencies in findings across studies are not specific to the study of hippocampal subfields, but this example highlights the difficulty researchers can face in drawing inferences related to normative development from this literature and the implications variability in measurement has for integrated interpretation.

Despite variations in specific results, studies provide converging evidence that hippocampal subfields follow distinct non-linear developmental trajectories, with larger CA1 and DG/CA3 volumes early in development followed by volumetric decreases as children enter adolescence and adulthood. Smaller volumes from childhood to adulthood appear to reflect hippocampal subfield maturity: larger CA1 and DG/CA3 volumes correlate with better memory performance in younger children, whereas smaller volumes correlate with better performance in older children, adolescents, and young adults (Canada et al., 2018; Riggins et al., 2018; Schlichting et al., 2017; Tamnes et al., 2014; although there are exceptions, Bouyeure et al., 2021). Yet, due to the limitations within single studies noted above, this hypothesis of distinct developmental differences has not been adequately tested. Knowledge of typical developmental differences in hippocampal subfield volumes is a first step toward building a mechanistic view of neurodevelopmental disorder progression. This study demonstration will test the hypothesis that hippocampal subfield volumes differentially relate to age across development, with age-related differences in DG/CA3 volume, but not CA1 or Sub volume.

1.5. Step 1. Find your team and obtain data

1.5.1. 1A. Identify IDA team and datasets

As noted above, the first step of identifying datasets in IDA is driven by the questions motivating the research. Early leaders in developing IDA threaded the importance of collaborative teams throughout the process (Curran and Hussong, 2009; Hussong et al., 2013). Gathering available data requires building a team that can provide input on the measures used and nuances of each study. The standout strength of IDA is the ability to include measures of the same construct from multiple samples that need not use identical measurement tools for all individuals across samples. However, some overlap in the measures used across samples or representation of key sample features, like age, is needed (Curran and Hussong, 2009; Hussong et al., 2013). Finally, IDA cannot overcome data with poor reliability or poor construct validity. The ability to draw strong inferences from the integrated sample begins with the quality of measurement in the original studies.

1.5.2. 1B. Obtain and organize data

A time intensive and often overlooked step in secondary data analysis is the organization of existing data from different sources. Organizing data for integration requires ensuring the shared data are both complete and variables of interest understood. This means, for example, a data dictionary for each source is provided to ensure understanding of the included variables such as the variable definition and measurement scale. This step might include initial recoding of identical variables to be on the same scale; for example, ensuring reported sex is coded the same across all studies. This step provides an opportunity to identify missing variables and ensure a clear understanding of which variables reflect the intended construct across studies.

1.5.3. Applied demonstration sample

Following steps 1A and 1B, the integrated sample in this demonstration includes high-resolution images of the medial temporal lobe collected from 443 4- to 17-year-olds recruited to three existing independent studies of healthy brain development (Table 1). The full range of the integrated sample extended to age 25 years. However, given the reliance on a single site for estimates beyond age 13 years, we excluded individuals aged 18 years and above due to low coverage (n = 30).[1] Studies differed in measurement methods and geographic diversity, and the overlapping age ranges allowed stitching the samples together to test hypotheses across a 14-year developmental span (Fig. 2).

[1] Results of the example hippocampal subfields integrated sample did not substantially differ from the reported titrated sample.

Table 1
Sample size by site and chronological age in years.

Site	*N*	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Detroit	173	-	13	15	4	11	12	12	11	13	7	6	11	9	11
College Park	148	38	20	35	28	27	-	-	-	-	-	-	-	-	-
Davis	122	-	-	-	8	18	28	36	25	6	1	-	-	-	-
Total	***443***	***38***	***33***	***50***	***40***	***56***	***40***	***48***	***36***	***19***	***8***	***6***	***11***	***9***	***11***

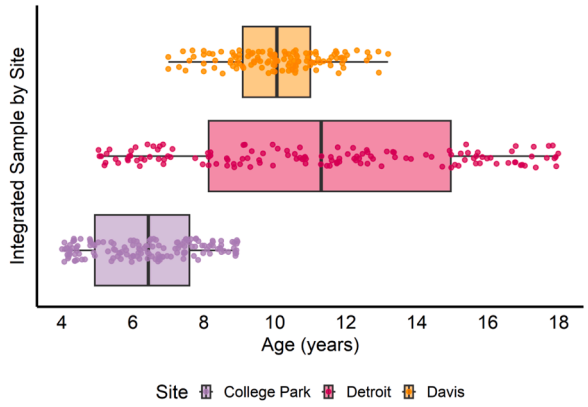


Fig. 2. Distribution of 443 participants in the full integrated sample by site and chronological age in years. Note overlap of the Detroit sample (n= 173; pink) with both the College Park (n = 148; purple), and Davis (n = 122; orange) samples.

1.6. Step 2. Create integrated dataset

1.6.1. 2A. Select commensurate items

Once source data are identified, the next key step is to review variables of the same construct. Although IDA does not require identical measures, all measures to be combined need to be commensurate. Commensurate measures are considered valid assessments of a shared construct and have scale properties that allow aggregation. For example, if two studies surveyed the presence of the same behavior, one survey might ask parents if the behavior is present (1) or absent (0), while the other survey may ask parents if the behavior is always present (2), sometimes present (1), or absent (0). Here, an investigator could recode responses from the second survey to combine always present (2→1) and sometimes present (1) into a single category to align with the first survey. The value of a team for joint expertise to adjudicate items and determine recoding is at the heart of IDA.

In the presented example, we reviewed the boundaries and labels of the three different protocols prior to analysis (Bender et al., 2013; Iglesias et al., 2015; La Joie et al., 2010; Fig. 3). The protocols had a number of regions that were mostly redundant in anatomical representation, although specific boundaries did vary somewhat. The exception was different allocation of the CA2 label across protocols, in addition to the variable use of CA4 as a label across protocols. The team agreed that the commensurate measures reflected constructs of CA1, DG/CA3, and Sub across all three protocols. Another example was the length of the hippocampus measured: some protocols were exclusive to the body of the hippocampus (the majority of the length) whereas others included different extents of anterior regions. Because there is majority consensus from histologists regarding the boundaries of subfields within the hippocampal body, but continued disagreement about definitions within the hippocampal head (Wisse et al., 2020), data were selected for the hippocampal body to create commensurate measures. Note that the data required for this demonstration are not the original MR images (scan data). Instead, IDA leverages derivative measures of MRI, illustrated here with estimated volumes, that can be shared using low resource intensive files, such as spreadsheets or comma-separated value text-based documents. This underscores the utility of IDA for open data sharing with low infrastructure cost. Once this step is completed, the investigator has a data file with commensurate measures aggregated in a

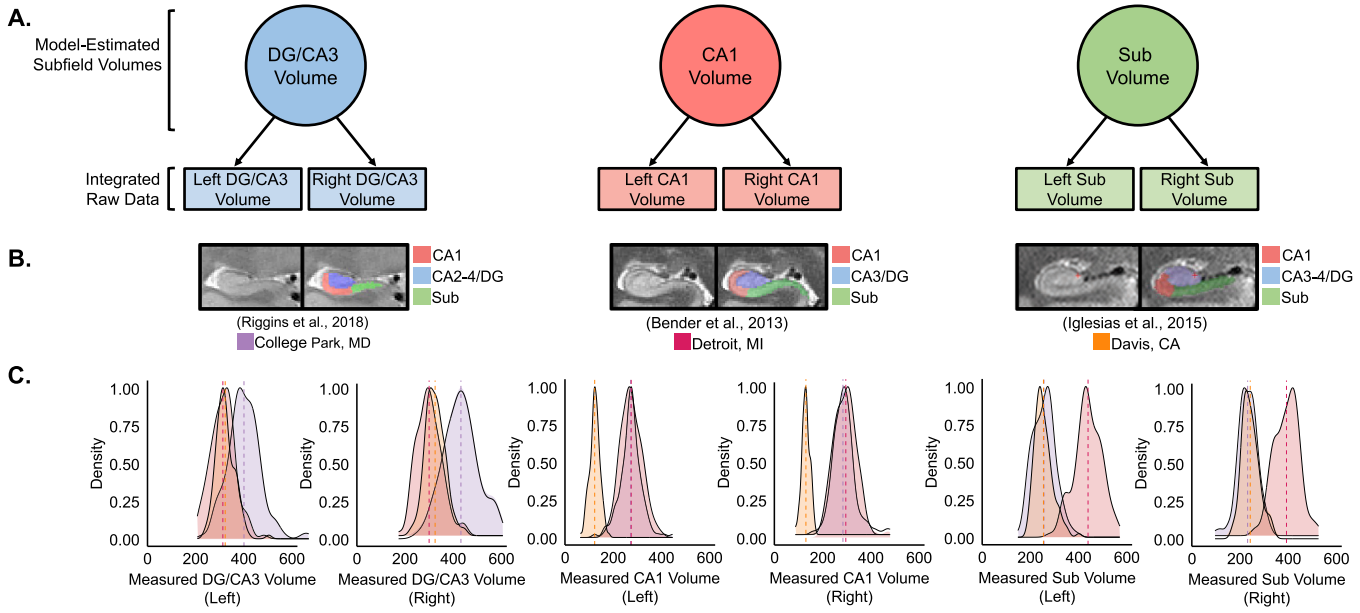


Fig. 3. A) Conceptual depiction of commensurate measures of hippocampal subfield volumes using right and left ROIs integrated across B) three different study sites and protocols: Detroit (Bender et al., 2013, 2018), College Park (Riggins et al., 2018; adapted from La Joie et al., 2010), and Davis (Iglesias et al., 2015). C) Distributions of unadjusted measured right and left measures of DG/CA3, CA1, and Sub subfield volumes from the College Park (purple), Detroit (pink), and Davis (orange) study sites.

column vector.

IDA can still leverage the strengths of mixed methods approaches that we reviewed. When pooling data across studies, there are two approaches to consider and evaluate heterogeneity in the combined sample due to study source-specific characteristics (Curran and Hussong, 2009). One approach is to treat each of the study samples as random draws from a large population of studies, called "random-effects IDA." This approach allows investigators to assess the data using random effects modeling and to treat study as a nesting variable to account for between-study variability. The "random-effects IDA" approach assumes the ability to theoretically establish that each included study is sampled from a single population of studies (Curran and Hussong, 2009). If the studies identified for IDA differ in theory or design, they should not be treated as random draws from a single population of studies, and a "fixed-effects IDA" instead treats each study as part of an available fixed set of studies (Curran and Hussong, 2009). In data simulations with mixed linear modeling, random effects models appeared to provide the most robust control of type I error across individual- and source-specific levels of the analysis; however, these models are difficult to converge when integrating data from only a few studies, in which case a fixed effects approach is recommended (Wilcox and Wang, 2023). In the current demonstration, we adopted a fixed-effects IDA and study site becomes a covariate in the model to account for between-study variability in the measurement.

1.6.2. Step 2B. Data visualization and screening

After selecting and organizing data, the combined dataset should be examined with common data screening practices and to ensure variables are correctly coded by site.

In the presented example, data were examined for univariate and multivatiate outliers within each individual study to ensure no errors in data entry or processesing occurred (e.g., negative volumes are impossible). Univariate and multivariate outliers were then identified in the combined sample to be flagged for subsequent analyses. Univariate outliers were determined by examining Z value of |3.3| reflecting more than 3 standard deviations from the mean, and multivariate outliers using Mahalanobis distance. Nine multivariate outliers were identified in this demonstration: 3 from the Detroit sample, 2 from the College Park sample, and 4 from the Davis sample. Univariate and multivariate outliers were identified in the combined dataset and model results with and without outliers were compared to ensure effects were not obscured or inflated by outlying data points.

We also examined differences in predictors and covariates of interest (i.e., Age, Sex, intracranial volume; ICV) across study sites using ANOVA. Age of participants differed by study site (see Fig. 2). The result of this screening step supports a need to disengtagle the effects of site and age in subsequent analyses in order to draw valid inferences. The ability to integrate across samples to examine larger developmental spans is a strength of IDA, but, researchers must be cognizant of how to best analyze the combined dataset.

1.7. Step 3. Establish factor invariance

Up to this step, the previous procedures have aggregated data similar to mega-analysis. In step 3, we now confirm that the combined commensurate measures operate comparably between sites. We established factor invariance by testing the necessary assumptions that allow commensurate (not identical) measures to be combined, with independent estimation of measurement error. This is the linchpin step that distinguishes IDA from all the other approaches we reviewed. In IDA, integrating data across studies requires establishing that the same construct is assessed across studies, which is represented in the latent score (Curran and Hussong, 2009). This is an implicit assumption to all other methods we reviewed, but an explicit test in IDA that determines the feasibility of the approach for the selected measures and samples. In our example, different protocols for defining hippocampal subfield volumes are used for measurement of each regional construct. In other aspects of developmental cognitive neuroscience, it is likely that differences will exist in measurements of cognition or psycho-social factors across developmental groups and other subgroups. To draw valid conclusions from integrated data across groups or over time, formal inferential tests of measurement invariance for identified commensurate items are conducted using confirmatory factor analysis (e.g., Meade and Lautenschlager, 2004).

Latent factor scores have no inherent scale by definition, and so must be identified with equivalent measurement scale across groups or time to allow for valid inferences. This is established with measurement invariance. Different levels of measurement invariance can be met: configural, weak (i.e., metric), strong (i.e., scalar), and strict invariance (Meredith, 1993). Configural invariance tests equivalence of the factor structure across groups; weak invariance tests the equivalence of factor loadings across groups; strong invariance tests the equivalence of indicator intercepts across groups; and strict invariance tests the equivalence of indicator residuals across groups (see Fig. 4 for illustrated depiction). Much of the literature on measurement invariance is framed in comparing groups in cross-sectional study. However, the same principles apply for variance over time in longitudinal study. Strict measurement invariance is not required for IDA; however, it provides the strongest evidence that commensurate measures reflect the constructs of interest similarly across studies.

At a minimum, weak invariance is required for IDA, otherwise no common construct for the aggregated items can exist (Davoudzadeh et al., 2020). Specifically, the factor structure and factor loadings should be invariant across studies. This allows the metric of the latent factor to be estimated consistently and compared across individuals in a combined dataset of included studies. This example of partial invariance, wherein a subset of the parameters can be held equal over groups or time, is sufficient for accurate cross-group comparisons (Byrne et al., 1989; Van De Schoot et al., 2012; Yoon and Millsap, 2007). The inability to establish, at minimum, weak invariance of the latent factor across studies indicates a need for investigators to reassess the commensurate items selected, ensure measures are comparably coded and re-code if necessary, or consider an alternative approach to secondary data analysis.

Invariance is tested by imposing model constraints to hold parameters equal across sources, or time for longitudinal study. For between-

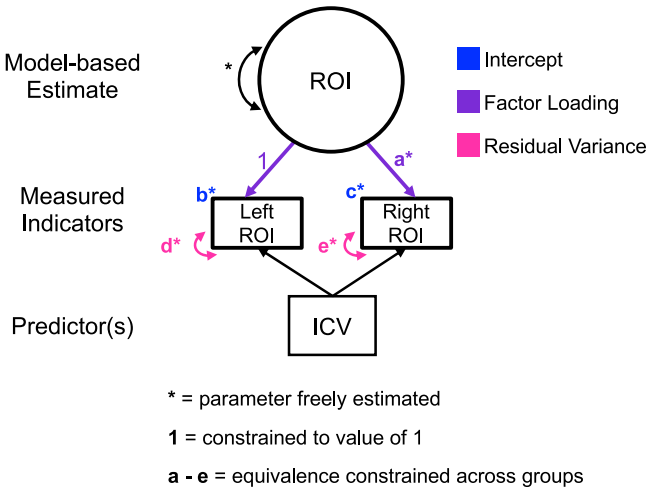


Fig. 4. Illustrated depiction of model parameters and constraints needed to meet different levels of measurement invariance. Factor loadings (purple) must be equal across groups to establish weak invariance (IDA minimum requirement). Factor loadings (purple) and intercepts (blue) must be equal across groups to establish strong invariance. Factor loadings (purple), intercepts (blue), and residual variances (pink) must be equal across groups to establish strict invariance.

group comparisons (i.e., different sources, study protocols, sites), multigroup models are useful to test the invariance of indicators (Meade and Lautenschlager, 2004; Meredith, 1993; Reise et al., 1993; Widaman and Reise, 1997). When combining samples across developmental groups or other subgroups, different grouping variables can be included to test invariance, making IDA a flexible tool to accommodate the different idiosyncrasies across designs.

In the current example, invariance was tested using change in fit indices with nested model comparisons and set constraints across study sites. This begins with an initial model that is evaluated for objectively good fit by a set of acceptable fit indices: root mean square error of approximation (RMSEA $\leq$ 0.08 supports good fit; Browne and Cudeck, 1992), comparative fit index (CFI $\geq$ 0.95 indicates good fit; Hu and Bentler, 1999), and standardized root mean residual (SRMR $\leq$ 0.08 supports good fit; Hu and Bentler, 1999). A constrained model was rejected if the loss in CFI value was .02 or greater (Putnick and Bornstein, 2016; Rutkowski and Svetina, 2014). Acceptability of models was determined using multiple fit indicators, as reliance on a single index to assess model fit is often insufficient.

As is the case in most applications of IDA to neuroimaging data, and in our example, only two indicator measures are available: right and left ROIs. Latent factors are reflective of their indicators, and identification typically is with three or more indicators per latent factor (Mueller and Hancock., 2018). In applications of IDA with a greater number of indicators, differential item functioning of commensurate measures can be assessed using moderated nonlinear factor analysis (MNLFA; Curran et al., 2014), which is beyond the scope of the current example. The "problem of 2" indicators limits the options for constraint to identify a latent factor, before even beginning tests of the additional equivalence constraints for measurement invariance. Although the "problem of 2" is rare in studies of cognitive, psychological, and social constructs, it is a common occurrence for neuroimaging data. MRI derivative measures typically include two hemisphere measures that represent the same region of interest, or common construct, and so can be used as indicators of a factor. The advantage of using the sum volume by hemisphere opposed to several individual slice measures is to have higher reliability of the measure to start with. Because IDA is beholden to the quality of data that goes into this step, measures with high reliability will provide better opportunity to isolate true variability in the factor score apart from measurement error. There are straightforward solutions to the "problem of 2" indicators, whereas there is little to be done with poor reliability of starting measures.

Latent factor identification is a balance of degrees of freedom and constraints that are plausible in the population. To provide the latent factor with scale, factor loadings, measure residuals and intercepts, or latent variances can be constrained, and thus not costing a degree of freedom to estimate. In a scenario with two indicators, at least two parameters must be constrained to identify the latent factor. When applied to neuroimaging data, in which both hemisphere measures are partially dependent and equally relevant to the regional construct, it is logical to fix the factor loadings for left and right hemisphere measures each to 1. In applications to other types of data, typically one indicator would be fixed to 1 and the other non-dependent measure would have a factor loading estimated but constrained to be equal across groups to meet weak invariance. In our application, we identify the latent factor and begin with weak invariance by constraining the factor loadings of both hemisphere indicators to 1. This factor score is conceptually similar to a bilateral sum volume, with consistent calculation across study sites.

There is the added benefit of allowing the residuals and intercepts for each left and right hemisphere indicators to be estimated and available to test for constraints to evaluate what degree of measurement invariance is supported. The additional constraints, if supported, improve the quality of subsequent hypothesis testing, including sensitivity of the analyses. This is the bedrock of the extracted "IDA factor score." However, a perfect solution with strict invariance is not required for valid inferences in the aggregated data. Note that in applications with multiple regions of interest, the same level of measurement invariance does not need to be met across regions. Instead, consider a minimum standard of weak invariance for each factor to move forward, and note all other tests as information gathering for limitations of the subsequent analysis. Moreover, the information on the scale properties of the measures across samples is highly valuable to the field at large; failures in measurement invariance point to future directions for instrument and protocol development, not abandonment of the line of research.

In the presented example, partial measurement invariance was supported for all hippocampal subfields. As we described, the specified factor loadings each at 1 begins with weak invariance, and all additional constraints were imposed sequentially to evaluate if the model fit was objectively acceptable, and no meaningful loss in fit by change in CFI. When indicator intercepts were constrained to be equal across sites, model fit was acceptable using at least two indices of model fit. CA1: CFI = 1.00, RMSEA = 0.000 (.000,.083), SRMR = 0.008; DG/CA3: CFI =0.992, RMSEA = 0.120 (.000,.237), SRMR =.056; Sub: CFI =.981, RMSEA =.132 (.031,.248), SRMR =.042. The models were specified as depicted in Fig. 5, with Site used as the grouping variable.

Strict measurement invariance was not met across Site groups for CA1, DG/CA3, or Sub. When residual variances were constrained to be equal across sites, model fit decreased beyond the acceptable amount (i.e., >.02 for CFI). CA1: CFI = 0.769, RMSEA = 0.268 (.211,.329), SRMR =.337; DG/CA3: CFI =.767, RMSEA = 0.367 (.311,.427), SRMR =.180; Sub: CFI =.819, RMSEA =.237 (.181,.299), SRMR =.205. In the applied example, we considered the minimum requirement of partial measurement invariance was met for continued analysis.

1.8. Step 4. Latent factor specification

In step 4, we fine tune the integrated factor scores. The tests of measurement invariance and the constraints across data sources that are supported in step 3 serve to ensure commensurate (rather than identical) measures can be integrated and begins to address source-specific idiosyncrasies in the data. We build upon this by adding additional covariates at the measurement and latent factor levels to refine statistical estimation of measurement error apart from true individual variability. There is a rich literature on the use of covariates and demographic features to improve the estimation of factor scores as valid representations of the construct in the population (Curran et al., 2016; Curran and Hussong, 2009; Davoudzadeh et al., 2020). In IDA, covariates that are added predicting the indicator are acting to adjust the individual measurement separate from the factor score, whereas covariates of the factor are accounting for sources of individual variability that may be meaningfully related to the construct. As an example, regional brain volumes are typically adjusted for ICV to account for sexual dimorphism in head size; a common adjustment is by residualization (i.e., ANCOVA approach Jack et al., 1989), which can be implemented by including ICV as a predictor of the hemisphere indicators in the measurement model (see Fig. 5). Age is a demographic factor of interest, and so can be included as a covariate of the factor score in the latent model; inclusion at this level of the model improves the estimation of the measurement residuals so that true age-related variability is at the factor score and mitigate lost to correlated measurement error.

In the present example, after establishing partial invariance of measures for each hippocampal subfield independently, the latent models for each region were combined into the full integrated model from which model-based estimates of each factor score were extracted for each individual. All constraints from prior steps were carried forward: indicator loadings were constrained to 1 for right and left measures of CA1, DG/CA3, and Sub; indicator intercepts were estimated but constrained to be equal across groups; and indicator residual variances were freely estimated. To account for correlations between hippocampal subfields, latent constructs of CA1, DG/CA3, and Sub were correlated. To account for potential hemispheric differences in indicator measures due to imaging protocol (e.g., protocols that were aligned to one

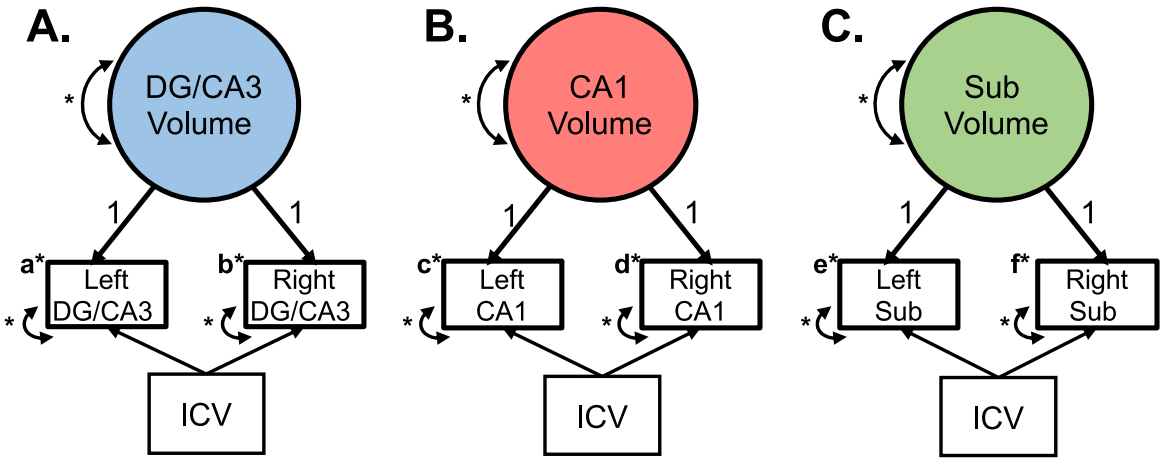


Fig. 5. Supported partial invariance structures across Site for each subfield. Loadings were constrained to 1 for right and left measures of DG/CA3, CA1, and Sub, indicator intercepts were estimated but constrained to be equal across groups, and residual variances were freely estimated across groups.

hemisphere for acquisition), right and left indicator measures were correlated.

Age and Sex predicted latent constructs of CA1, DG/CA3, and Sub, as well as measured ICV. Biological sex was included in the model to account for potential differences in volume between reported females and males. As in the prior steps' models, all individual hippocampal subfield volume measures were regressed on ICV. Age and Sex were correlated by virtue of the convenience samples in the source data. Models were fit and estimated with robust full information maximum likelihood, which is not an imputation approach and instead leverages all available data to produce unbiased estimates with data missing at random and including auxiliary variables to account for patterns of missing data (e.g., site effects in missing data patterns; Little et al., 2014; McNeish, 2017). Fit for the full integrated model was good, CFI = 0.986, RMSEA = 0.074 (.040, 105), SRMR =.060. Model fit provides evidence that the specified model reproduces the observed data well and provides support for the validity of the specified factor scores.

From this model, IDA factor scores were estimated per individual and exported as a new variable to be used in subsequent hypothesis testing. Extraction of the IDA factor score is not often explicit in the literature. Instead, there is often reference to "creating harmonized scores for subsequent hypothesis testing" (p. 1033, Hussong et al., 2021). This is a compromise, in part, out of practical necessity due to the complexity of the model and number of parameters to estimate in proportion to feasible sample sizes (Hussong et al., 2020). Fig. 6 illustrates the IDA factor score estimation and histogram distributions of the scores. There remain site differences in the distributions, which is to be expected because the distribution of age and other meaningful factors differ by site. In the practice of combined secondary data analysis, the purpose is not to obliterate all possible site-related differences, but to precisely remove the bias of source-specific error in the measurement of interest while leaving true individual differences intact.

Although we demonstrate the usefulness of extracted IDA factor score in making population-level inferences, it is important to note that the extracted model based-estimates are specific to the integrated sample used and as a latent factor score, and thus its metric cannot be directly compared to other external measures of the construct (Curran et al., 2014). Moreover, the factor score scale is defined within its native model, which does not include additional variables planned for hypothesis testing. These limitations should be weighed against the strengths of the approach including the ability to estimate IDA scores from non-identical measurements, with thorough testing of measurement invariance, which can still be used to thoughtfully test hypotheses in other models.

1.9. Step 5. Hypothesis testing

Finally, having completed steps 1–4, the extracted IDA factor score

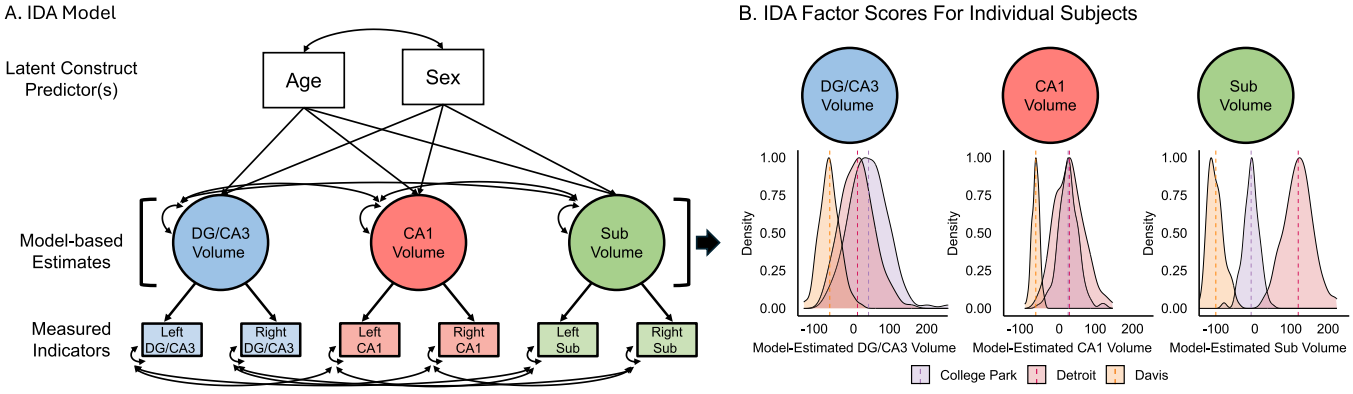


Fig. 6. A) IDA model resulting in factor scores for all 443 individual subjects in the combined dataset. Loadings were constrained to 1 for right and left measures of CA1, DG/CA3, and Sub, indicator intercepts were estimated but constrained to be equal across groups, and residual variances were freely estimated across groups. Although omitted for simplicity, all subfield indicators were predicted by measured ICV, and measured ICV was predicted by both Age and Sex. This model provides measures of hippocampal subfield with equivalent meaning across study sites. B) Distributions of model-estimated values of hippocampal subfields for DG/CA3 (blue), CA1 (red), and Sub (green) are shown for Detroit (orange), College Park (purple), Davis (pink). The IDA factor scores here reflect the disattenuation of study site from ROI measures.

(s) can then be used in subsequent analysis of the hypothesis originally identified in step 0, and subject to all typical data screening and modeling procedures. In this example, the hypothesis test was done using a path model that included the extracted IDA factor scores. This final model allows us to test developmental differences in hippocampal subfield volumes that are de-confounded from site-related error (Fig. 7A). Sex was tested as a covariate and found to be not significant, and therefore omitted from further hypothesis testing for parsimony (its exclusion did not change the pattern of results). Larger DG/CA3 volume significantly related to older age (standardized $\beta = 0.107$ $p = 0.018$; Fig. 7B). Similar direction of effects was observed in CA1 ($\beta = 0.052$, $p = 0.155$; Fig. 7C) and Sub volume ($\beta = 0.037$, $p = 0.051$, Fig. 7D), but these did not reach statistical significance. Estimated age effects statistically significantly differed between DG/CA3 and CA1 ($p = .01$), but not between DG/CA3 and Sub ($p = .25$) nor CA1 and Sub ($p = .62$). Notably, Site was still included as a potential covariate of regional volumes in order to disentangle the effects of age and study site identified during initial data screening (all $\beta = -1.08$–$.01$, $p = .00$–$.86$). Remaining site-related differences likely reflect differences in meaningful demographic factors of interest that vary by geographic regions across studies, such as race/ethnicity and sociodemographic status, that can be confidently tested in further study with the benefits of the extracted IDA factor score. The promise of IDA for testing new questions is on display, as other commensurate variables can be included in the model based on the hypothesis testing planned.

2. Discussion

Here, we review IDA as a promising method to advance developmental cognitive neuroscience with secondary data analysis. IDA allows investigators to leverage meaningful variability across individuals while de-confounding source-specific differences in neuroimaging measures; that is, IDA can help ensure researchers are not "throwing out the baby with the bath water." We found age-related differences in DG/CA3 volume, but not CA1 and Sub, in an integrated sample of 443 individuals. Our demonstration illustrates the potential of this method to enable and facilitate progress in the study of brain development by leveraging existing efforts to generate robust insights based on large,

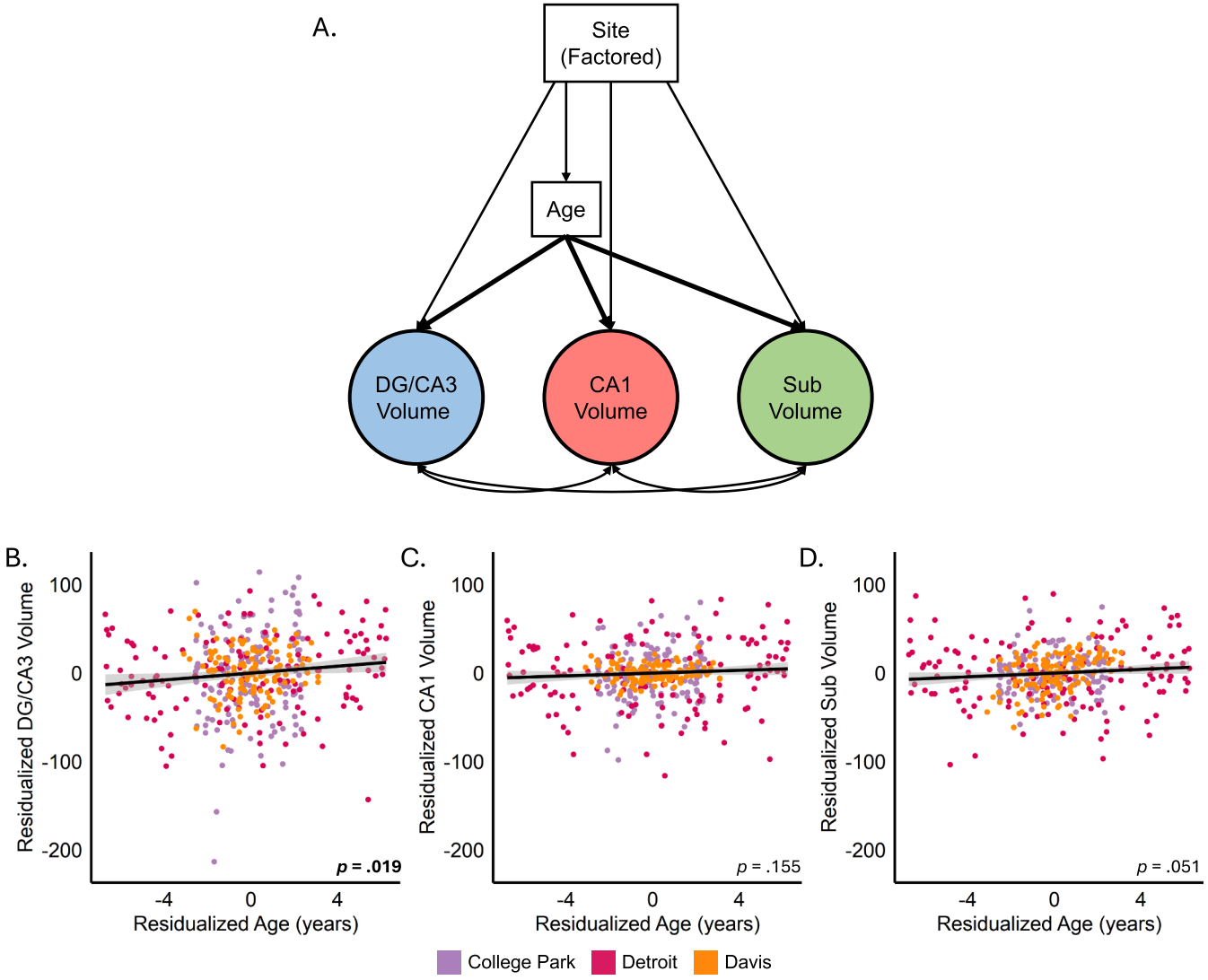


Fig. 7. A) Model testing hypothesis of age-related differences in hippocampal subfield volumes using extracted IDA factor scores accounting for Site differences not related to the measurement. Results of the bolded paths are depicted in scatterplots B, C, and D. In the scatterplots of relations between hippocampal subfield volume and age in the full integrated sample, age and each measure of subfield volume have been residualized. These plots demonstrate the strength of the approach, as the relation between measures is disentangled from study site differences. Importantly, the distributions still overlap on the y-axis but rank order differences between site are no longer present. B) Significant relation between age and DG/CA3 volume in the full integrated sample; C) Non-significant relation between age and CA1 volume in the full integrated sample; D) Non-significant relation between age and Sub volume in the full integrated sample.

representative samples.

We have outlined the necessary steps involved in using IDA and included a complete demonstration using volumetric measures of hippocampal subfields. The existing IDA methods literature has provided a strong starting point in foundational concepts (Shrout, 2009), and we have built on it to provide a procedural guide for developmental cognitive neuroscience researchers to implement on their data. In particular, we outline and demonstrate special considerations when working with neuroimaging data. As a flexible analytic tool for secondary data, IDA has the potential to set new frontiers in developmental cognitive neuroscience, addressing some of the challenges of using other cumulative science methods. Beyond providing insights into age-related differences in hippocampal subfield volumes across the period of 4-to-17-years, our demonstrated successful application and integration of commensurate volume measures opens the possibility to test new questions of individual differences in brain development that can go beyond the limitations of any one study. Given the susceptibility of the brain during development to both positive and negative influences, a critical goal of many developmental cognitive neuroscientists is to identify factors that modify development. However, investigations are often limited by sample size and the relatively homogenous demographic of individual samples. The IDA approach allows an increased sample size and representation of different sociodemographic backgrounds for analysis. In the current demonstration, samples included individuals from three different regions of the United States of America: the Midwest (Detroit, MI), East Coast (College Park, MD), and West Coast (Davis, CA). In addition to differences in the protocols used to assess hippocampal subfield volumes, individuals across each of these sites likely vary in their racial and socioeconomic makeup. Thus, a next step for the combined dataset used in our demonstration is to examine how early life SES impacts age-related differences in brain and cognition.

Applied to other brain regions and hypotheses, the IDA approach outlined here can allow researchers to build understanding of neurocognitive development based on large, representative samples that do not require identical neuroimaging or cognitive measures (Curran and Hussong, 2009; Hussong et al., 2013). The value of IDA to our field is further underscored by the limited feasibility of conducting new large-scale MRI studies due to time, access, and potential issues of sample diversity. In this manner, integrated secondary data analysis is a complementary approach to new data collection: we can progress our understanding of how the brain develops and supports improvements in cognition across development by working collaboratively. Moreover, this framework allows researchers to test questions of developmental change while explicitly testing the equivalence of constructs across both sources of data and time.

Taken together, IDA of existing developmental data can provide standalone hypothesis testing of new research questions and provide critical information for future data collection efforts to optimize for the time and financing available. For example, selection of assessments that have compatible psychometric properties but can be customized to samples without forced harmonization. Additionally, sensitive developmental periods to prioritize for new data collection can be identified. In our future work we will be applying this to test individual differences in longitudinal change of hippocampal subfield volumes to identify not only average trajectories, but also sensitive periods to sociodemographic factors across the integrated diverse sample.

While the focus of our demonstration is the novel application of IDA to neuroimaging data, IDA is appealing to developmental research more broadly. This framework provides a means for assessing the same cognitive and psychosocial constructs across development using different stimuli that are age- and population-appropriate. Examples include verbal and non-verbal stimuli; computerized vs. paper-pencil administration; lab vs. community-based data collection; child vs. parent or teacher reporting. Further, it facilitates study of diverse populations, offering the opportunity to use measures that differ in cultural content or administration language. Because assumptions of comparable constructs and commensurate measures are explicitly tested in the IDA framework, results showing a lack of equivalence in measures provides important insights into *how* we study development across different contexts.

A major strength of IDA supported by the SEM framework is the ability to estimate parameters independent of measurement error estimates. However, it is important to recognize that in practice we cannot completely account for all error. Hypothesis testing in IDA relies on the use of extracted factor scores. Because factor scores are defined in the context of the model they are estimated in, they require careful consideration of interpretation when moved outside of that context. Said differently, IDA factor scores derived from one combined sample are valid only within the context of that sample and the constraints of its native model. While it is not uncommon to use factor scores outside of their native models for subsequent hypothesis testing, there is still the risk of bias (Hoshino and Bentler, 2011) and others have discussed additional considerations for mitigating such issues (Hayes and Usami, 2020; Skrondal and Laake, 2001).

Moreover, the quality of the IDA factor score begins with the reliability of the individual measures in each sample. Investing time and best practices to ensure high quality data may reduce user-related error in the measures of interest and improve the likelihood of successful integration. There are a growing number of resources available to researchers to support data quality procedures. For example, the hippocampal subfield volumes used in this demonstration were reviewed for quality using different procedures (e.g., Canada et al., 2024; Homayouni et al., 2021), and reviews on structural (e.g., Backhausen et al., 2016, 2021) and functional (e.g, Teves et al., 2023) data quality are also available.

As developmental cognitive neuroscientists continue to leverage existing data, we think it is worthwhile to reflect upon longstanding conversations surrounding family-wise error (Ranganathan et al., 2016). In relation to IDA, investigators may publish on their study samples using the same variables included in the IDA as a combined sample. One can consider if the conceptual repetition of tests in samples included for IDA constitutes a family of tests with implications for type I error control. In practice, adjustments for type I error based on the number of previous analyses in mega-analysis or meta-analysis, or even in the same program of research is, to our knowledge, rare in the cognitive neuroscience field. While there is the possibility for increased type I error rate in analysis of combined data for which hypotheses in contributing samples have been independently tested, and results should be considered in the context of the literature, the true strength of IDA is the ability to open up hypothesis testing that is not otherwise possible in any single study. Introducing new hypothesis tests in a combined sample with different demographic representation creates an interesting philosophical question if IDA would fall within the same family of tests as anything done in the originating samples. While it is unclear how the field might proceed for applications of IDA, there has been some discussion of sequential analysis by different investigators on the same open data source (Thompson et al., 2020). We hope future work addressing these questions from other investigators will offer guidance on best practices, as open data will only become more prolific in the future.

Conclusion. Our successful application and demonstration of IDA for neuroimaging data sets the stage for other investigators to pursue fundamental questions related to promoting healthy brain development, to identify factors that modify development, and to promote early detection of and intervention for neurodevelopmental disorders. Our application of IDA to hippocampal subfield measures serves as a blueprint for a feasible alternative to redundant study design, forced harmonization, and new large-scale multi-site studies. Our hope is that future innovations in cognitive neuroscience will come from collaboration among scientists to combine existing data and create representative integrated samples when testing critical developmental questions.

Grants and acknowledgements

We thank Patrick J. Curran for helpful conversations surrounding our application of Integrative Data Analysis to neuroimaging data. This work was supported by funding by the NIH/NICHD F32-HD108960 (K.L. Canada), NIH/NIA P30AG072931 and NIH/NIA 5R01-AG011230 (A.M. Daugherty), NIH/NIMH R01-MH107512 (N. Ofen), NIH/NIMH R01-MH091109 (S. Ghetti), and NIH/NICHD R01-HD079518 (T. Riggins).

CRediT authorship contribution statement

Ana M Daugherty: Writing – review & editing, Supervision, Resources, Funding acquisition, Formal analysis, Conceptualization. **Noa Ofen:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Simona Ghetti:** Writing – review & editing, Resources. **Tracy Riggins:** Writing – review & editing, Resources. **Kelsey L Canada:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Example Mplus Code for IDA Step 4 – Latent Factor Specification and Extraction

```
TITLE:
IDA Three Site Example Syntax for IDA Factor Score Extraction
DATA:
FILE IS IDA_Mplus_Import_Site.inp;
VARIABLE:
  NAMES ARE MPLUSID Age Sex
  left_sub left_CA1 left_DG right_sub right_CA1 right_DG
  ICV Site;
  IDVAR IS MPLUSID;
USEVARIABLES ARE
  Age
  Sex
  left_sub right_sub
  left_CA1 right_CA1
  left_DG right_DG
  ICV_Scaled;
  GROUPING IS Site(-1=Site1, 0=Site2, 1=Site3);
  MISSING IS ALL (-999);
DEFINE:
  ICV_Scaled = ICV/10000;
ANALYSIS:
MODEL:
  CA1 BY left_CA1@1 right_CA1@1;
  left_CA1 right_CA1;
  left_CA1 ON ICV_Scaled;
  right_CA1 ON ICV_Scaled;
  CA1 WITH ICV_Scaled@0;
  Sub BY left_sub@1 right_sub@1;
  left_sub right_sub;
  left_sub ON ICV_Scaled;
  right_sub ON ICV_Scaled;
  Sub WITH ICV_Scaled@0;
  DG BY left_DG@1 right_DG@1;
  left_DG right_DG;
  left_DG ON ICV_Scaled;
  right_DG ON ICV_Scaled;
  DG WITH ICV_Scaled@0;
  left_DG ON Sex@0 Age@0;
  right_DG ON Sex@0 Age@0;
  DG ON Sex Age;
  left_CA1 ON Sex@0 Age@0;
  right_CA1 ON Sex@0 Age@0;
  CA1 ON Sex Age;
  left_sub ON Sex@0 Age@0;
  right_sub ON Sex@0 Age@0;
  SUB ON Sex Age;
  ICV_Scaled ON Sex;
  ICV_Scaled ON AGE;
  left_sub left_CA1 WITH left_DG left_sub;
  right_sub right_CA1 WITH right_DG right_sub;
  SUB CA1 WITH DG SUB;
MODEL Site1:
  left_sub right_sub;
  left_CA1 right_CA1;
  left_DG right_DG;
MODEL Site2:
  left_sub right_sub;
  left_CA1 right_CA1;
  left_DG right_DG;
MODEL Site3:
  left_sub right_sub;
  left_CA1 right_CA1;
  left_DG right_DG;
OUTPUT:
  SAMPSTAT STANDARDIZED RESIDUAL MODINDICES (4)
SAVEDATA:
  FILE IS HcExtractedIDAScores.csv;
  SAVE IS FSCORES;
```

Appendix B

Example Mplus Code for IDA Step 5 – Hypothesis Testing

```
TITLE:
IDA Three Site Path Model Example Syntax with Extracted IDA Factor Score
DATA:
FILE IS HcExtractedIDAScores.inp;
VARIABLE:
NAMES ARE LEFT_SUB RIGHT_SU LEFT_CA1 RIGHT_CA LEFT_DG RIGHT_DG ICV_SCAL AGE SEX CA1 CA1_SE SUB SUB_SE DG DG_SE MPLUSID SITE;
  IDVAR IS MPLUSID;
  USEVARIABLES ARE
  Age
  Sex
  CA1
  DG
  Sub
  Site;
ANALYSIS:
MODEL:
  CA1 ON Age Sex Site;
  DG ON Age Sex Site;
  Sub ON Age Sex Site;
  CA1 SUB WITH DG SUB;
  Age WITH Sex@0;
  Age WITH Site;
  Sex WITH Site@0;
OUTPUT:
  SAMPSTAT STANDARDIZED RESIDUAL MODINDICES (4)
```

Data availability

Individual and integrated data available upon request.

References

Backhausen, L.L., Herting, M.M., Buse, J., Roessner, V., Smolka, M.N., Vetter, N.C., 2016. Quality control of structural MRI images applied using freesurfer—a hands-on workflow to rate motion artifacts. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00558.

Backhausen, L.L., Herting, M.M., Tamnes, C.K., Vetter, N.C., 2021. Best practices in structural neuroimaging of neurodevelopmental disorders. Neuropsychol. Rev. https://doi.org/10.1007/s11065-021-09496-2.

Bayer, J.M.M., Thompson, P.M., Ching, C.R.K., Liu, M., Chen, A., Panzenhagen, A.C., Jahanshad, N., Marquand, A., Schmaal, L., Sämann, P.G., 2022. Site effects how-to and when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. Front. Neurol. 13, 923988. https://doi.org/10.3389/fneur.2022.923988.

Bender, A.R., Daugherty, A.M., Raz, N., 2013. Vascular risk moderates associations between hippocampal subfield volumes and memory. J. Cogn. Neurosci. 25 (11), 1851–1862. https://doi.org/10.1162/jocn_a_00435.

Bender, A.R., Keresztes, A., Bodammer, N.C., Shing, Y.L., Werkle-Bergner, M., Daugherty, A.M., Yu, Q., Kühn, S., Lindenberger, U., Raz, N., 2018. Optimization and validation of automated hippocampal subfield segmentation across the lifespan. Hum. Brain Mapp. 39 (2), 916–931. https://doi.org/10.1002/hbm.23891.

Bockholt, H.J., Scully, M., Courtney, W., Rachakonda, S., Scott, A., Caprihan, A., Fries, J., Kalyanam, R., Segall, J.M., de la Garza, R., Lane, S., Calhoun, V.D., 2010. Mining the mind research network: a novel framework for exploring large scale, heterogeneous translational neuroscience research data sources. Front. Neuroinform. 3, 36. https://doi.org/10.3389/neuro.11.036.2009.

Boedhoe, P.S.W., Heymans, M.W., Schmaal, L., Abe, Y., Alonso, P., Ameis, S.H., Anticevic, A., Arnold, P.D., Batistuzzo, M.C., Benedetti, F., Beucke, J.C., Bollettini, I., Bose, A., Brem, S., Calvo, A., Calvo, R., Cheng, Y., Cho, K.I.K., Ciullo, V., Twisk, J.W. R., 2019. An empirical comparison of meta- and mega-analysis with data from the ENIGMA obsessive-compulsive disorder working group. Front. Neuroinform. 12, 102. https://doi.org/10.3389/fninf.2018.00102.

Bouyeure, A., Patil, S., Mauconduit, F., Poiret, C., Isai, D., Noulhiane, M., 2021. Hippocampal subfield volumes and memory discrimination in the developing brain. Hippocampus 31 (11), 1202–1214. https://doi.org/10.1002/hipo.23385.

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. Sociol. Methods Res. 21 (2), 230–258. https://doi.org/10.1177/0049124192021002005.

Byrne, B.M., Shavelson, R.J., Muthén, B., 1989. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. Psychol. Bull. 105 (3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456.

Canada, K.L., Hancock, G.R., Riggins, T., 2021. Modeling longitudinal changes in hippocampal subfields and relations with memory from early- to mid-childhood. Dev. Cogn. Neurosci. 48, 100947. https://doi.org/10.1016/j.dcn.2021.100947.

Canada, K.L., Mazloum-Farzaghi, N., Rådman, G., Adams, J.N., Bakker, A., Baumeister, H., Berron, D., Bocchetta, M., Carr, V., Dalton, M.A., de Flores, R., Keresztes, A., La Joie, R., Mueller, S.G., Raz, N., Santini, T., Shaw, T., Stark, C.E.L., Tran, T.T., Wang, L., Wisse, L.E.M., Wustefeld, A., Yushkevich, P.A., Olsen, R.K., Daugherty, A.M., on behalf of the Hippocampal Subfields Group, 2024. A (Sub)field guide to quality control in hippocampal segmentation on high-resolution T_2-weighted MRI. Hum. Brain Mapp. https://doi.org/10.1002/hbm.70004.

Canada, K.L., Ngo, C.T., Newcombe, N.S., Geng, F., Riggins, T., 2018. It's All in the details: relations between young children's developing pattern separation abilities and hippocampal subfield volumes. Cereb. Cortex 1–7. https://doi.org/10.1093/cercor/bhy211.

Chan, M.E., Arvey, R.D., 2012. Meta-analysis and the development of knowledge. Perspect. Psychol. Sci. 7 (1), 79–92. https://doi.org/10.1177/1745691611429355.

Chen, A.A., Beer, J.C., Tustison, N.J., Cook, P.A., Shinohara, R.T., Shou, H., The Alzheimer's disease neuroimaging initiative, 2022. Mitigating site effects in covariance for machine learning in neuroimaging data. Hum. Brain Mapp. 43 (4), 1179–1195. https://doi.org/10.1002/hbm.25688.

Cohen, J.D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K.A., Pillow, J., Ramadge, P.J., Turk-Browne, N.B., Willke, T.L., 2017. Computational approaches to fMRI analysis. Nat. Neurosci. 20 (3), 304–313. https://doi.org/10.1038/nn.4499.

Curran, P.J., 2003. Have multilevel models been structural equation models all along? Multivar. Behav. Res. 38 (4), 529–569. https://doi.org/10.1207/s15327906mbr3804_5.

Curran, P.J., Cole, V., Bauer, D.J., Hussong, A.M., Gottfredson, N., 2016. Improving factor score estimation through the use of observed background characteristics. Struct. Equ. Model.: A Multidiscip. J. 23 (6), 827–844. https://doi.org/10.1080/10705511.2016.1220839.

Curran, P.J., Hussong, A.M., 2009. Integrative data analysis: the simultaneous analysis of multiple data sets. Article 2. Psychol. Methods 14 (2). https://doi.org/10.1037/a0015914.

Curran, P.J., Hussong, A.M., Cai, L., Huang, W., Chassin, L., Sher, K.J., Zucker, R.A., 2008. Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. Article 2. Dev. Psychol. 44 (2). https://doi.org/10.1037/0012-1649.44.2.365.

Curran, P.J., McGinley, J.S., Bauer, D.J., Hussong, A.M., Burns, A., Chassin, L., Sher, K., Zucker, R., 2014. A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. Multivar. Behav. Res. 49 (3), 214–231. https://doi.org/10.1080/00273171.2014.889594.

Daugherty, A.M., Bender, A.R., Raz, N., Ofen, N., 2016. Age differences in hippocampal subfield volumes from childhood to late adulthood. Article 2. Hippocampus 26 (2). https://doi.org/10.1002/hipo.22517.

Davatzikos, C., 2019. Machine learning in neuroimaging: progress and challenges. NeuroImage 197, 652–656. https://doi.org/10.1016/j.neuroimage.2018.10.003.

Davoudzadeh, P., Grimm, K.J., Widaman, K.F., Desmarais, S.L., Tueller, S., Rodgers, D., Van Dorn, R.A., 2020. Estimation of latent variable scores with multiple group item response models: implications for integrative data analysis. Article 00 Struct. Equ. Model. 00 (00). https://doi.org/10.1080/10705511.2020.1724113.

De Wit, S.J., Alonso, P., Schweren, L., Mataix-Cols, D., Lochner, C., Menchón, J.M., Stein, D.J., Fouche, J.-P., Soriano-Mas, C., Sato, J.R., Hoexter, M.Q., Denys, D., Nakamae, T., Nishida, S., Kwon, J.S., Jang, J.H., Busatto, G.F., Cardoner, N., Cath, D. C., van den Heuvel, O.A., 2014. Multicenter voxel-based morphometry mega-analysis of structural brain scans in obsessive-compulsive disorder. Am. J. Psychiatry 171 (3), 340–349. https://doi.org/10.1176/appi.ajp.2013.13040574.

Greenhoot, A.F., Dowsett, C.J., 2012. Secondary data analysis: an important tool for addressing developmental questions. Article 1. J. Cogn. Dev. 13 (1). https://doi.org/10.1080/15248372.2012.646613.

Hao, Y., Xu, H., Xia, M., Yan, C., Zhang, Y., Zhou, D., Kärkkäinen, T., Nickerson, L.D., Li, H., Cong, F., 2023. Removal of site effects and enhancement of signal using dual projection independent component analysis for pooling multi-site MRI data. Eur. J. Neurosci. 58 (6), 3466–3487. https://doi.org/10.1111/ejn.16120.

Hayes, T., Usami, S., 2020. Factor score regression in the presence of correlated unique factors. Educ. Psychol. Meas. 80 (1), 5–40. https://doi.org/10.1177/0013164419854492.

Hofer, S.M., Piccinin, A.M., 2009. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Article 2. Psychol. Methods 14 (2). https://doi.org/10.1037/a0015566.

Homayouni, R., Yu, Q., Ramesh, S., Tang, L., Daugherty, A.M., Ofen, N., 2021. Test–retest reliability of hippocampal subfield volumes in a developmental sample: Implications for longitudinal developmental studies. J. Neurosci. Res. 24831. https://doi.org/10.1002/jnr.24831.

Hoshino, T., & Bentler, P.M. (2011). Bias in factor score regression and a simple solution. eScholarship, University of California.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. A Multidiscip. J. 6 (1), 1–55. https://doi.org/10.1080/10705519909540118.

Hussong, A.M., Bauer, D.J., Giordano, M.L., Curran, P.J., 2021. Harmonizing altered measures in integrative data analysis: a methods analogue study. Behav. Res. Methods 53 (3), 1031–1045. https://doi.org/10.3758/s13428-020-01472-7.

Hussong, A.M., Cole, V.T., Curran, P.J., Bauer, D.J., Gottfredson, N.C., 2020. Integrative Data Analysis and the Study of Global Health. In: Chen, X., Ding-Geng Chen, (Din) (Eds.), Statistical Methods for Global Health and Epidemiology: Principles, Methods and Applications. Springer International Publishing, pp. 121–158. https://doi.org/10.1007/978-3-030-35260-8_5.

Hussong, A.M., Curran, P.J., Bauer, D.J., 2013. Integrative data analysis in clinical psychology research. Annu. Rev. Clin. Psychol. 9, 61–89. https://doi.org/10.1146/annurev-clinpsy-050212-185522.

Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., Fischl, B., Van Leemput, K., 2015. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. NeuroImage 115, 117–137. https://doi.org/10.1016/j.neuroimage.2015.04.042.

Jack, C.R., Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., Cascino, G.D., 1989. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology 172 (2), 549–554. https://doi.org/10.1148/radiology.172.2.2748838.

Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, S., Grigis, A., Martinot, J.-L., Paus, T., Smolka, M.N., Walter, H., Schumann, G., Garavan, H., Whelan, R., 2019. Quantifying performance of machine learning methods for neuroimaging data. NeuroImage 199, 351–365. https://doi.org/10.1016/j.neuroimage.2019.05.082.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer-Verlag. https://doi.org/10.1007/b98835.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 374 (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.

La Joie, R., Fouquet, M., Mézenge, F., Landeau, B., Villain, N., Mevel, K., Pélerin, A., Eustache, F., Desgranges, B., Chételat, G., 2010. Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence. NeuroImage 53 (2), 506–514. https://doi.org/10.1016/j.neuroimage.2010.06.024.

Lambert, P.C., Sutton, A.J., Abrams, K.R., Jones, D.R., 2002. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J. Clin. Epidemiol. 55 (1), 86–94. https://doi.org/10.1016/S0895-4356(01)00414-0.

Lavenex, P., Banta Lavenex, P., 2013. Building hippocampal circuits to learn and remember: insights into the development of human memory. Behav. Brain Res. 254, 8–21. https://doi.org/10.1016/j.bbr.2013.02.007.

Lee, J.K., Ekstrom, A.D., Ghetti, S., 2014. Volume of hippocampal subfields and episodic memory in childhood and adolescence. NeuroImage 94, 162–171. https://doi.org/10.1016/j.neuroimage.2014.03.019.

Lin, L., Chu, H., 2018. Quantifying publication bias in meta-analysis. Biometrics 74 (3), 785–794. https://doi.org/10.1111/biom.12817.

Little, T.D., Jorgensen, T.D., Lang, K.M., Moore, E.W.G., 2014. On the joys of missing data. J. Pediatr. Psychol. 39 (2), 151–162. https://doi.org/10.1093/jpepsy/jst048.
Marzi, C., Giannelli, M., Barucci, A., Tessa, C., Mascalchi, M., Diciotti, S., 2024. Efficacy of MRI data harmonization in the age of machine learning: a multicenter study across 36 datasets. Sci. Data 11 (1), 115. https://doi.org/10.1038/s41597-023-02421-7.
McArdle, J.J., Ferrer-Caja, E., Hamagami, F., Woodcock, R.W., 2002. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. Article 1. Dev. Psychol. 38 (1). https://doi.org/10.1037/0012-1649.38.1.115.
McNeish, D., 2017. Missing data methods for arbitrary missingness with small samples. J. Appl. Stat. 44 (1), 24–39. https://doi.org/10.1080/02664763.2016.1158246.
Meade, A.W., Lautenschlager, G.J., 2004. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organ. Res. Methods 7 (4), 361–388. https://doi.org/10.1177/1094428104268027.
Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58 (4), 525–543.
Mueller, R.O., Hancock, G.R., 2018. Structural equation modeling. The reviewer's guide to quantitative methods in the social sciences. Routledge, pp. 445–456.
Muthén, L.K., & Muthén, B.O. (2017). Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8). Authors.
Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.
Ranganathan, P., Pramesh, C., Buyse, M., 2016. Common pitfalls in statistical analysis: The perils of multiple testing. Perspect. Clin. Res. 7 (2), 106. https://doi.org/10.4103/2229-3485.179436.
Raykov, T., Marcoulides, G.A., Li, T., 2017. On the fallibility of principal components in research. Educ. Psychol. Meas. 77 (1), 165–178. https://doi.org/10.1177/0013164416629714.
Reise, S.P., Widaman, K.F., Pugh, R.H., 1993. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. Psychol. Bull. 114 (3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552.
Riggins, T., Geng, F., Botdorf, M., Canada, K., Cox, L., Hancock, G.R., 2018. Protracted hippocampal development is associated with age-related improvements in memory during early childhood (Article March.). NeuroImage 174 (March). https://doi.org/10.1016/j.neuroimage.2018.03.009.
Rosenberg, M.D., Casey, B.J., Holmes, A.J., 2018. Prediction complements explanation in understanding the developing brain. Nat. Commun. 9 (1), 589. https://doi.org/10.1038/s41467-018-02887-9.
Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. J. Stat. Softw. 48 (2), 1–36. https://doi.org/10.18637/jss.v048.i02.
Rutkowski, L., Svetina, D., 2014. Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. Educ. Psychol. Meas. 74 (1), 31–57. https://doi.org/10.1177/0013164413498257.
Schlichting, M.L., Guarino, K.F., Schapiro, A.C., Turk-Browne, N.B., Preston, A.R., 2017. Hippocampal structure predicts statistical learning and associative inference abilities during development. Article 1 J. Cogn. Neurosci. 29 (1). https://doi.org/10.1162/jocn.
Schmidt, F.L., Le, H., Ilies, R., 2003. Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. Psychol. Methods 8 (2), 206–224. https://doi.org/10.1037/1082-989X.8.2.206.
Seress, L., 2001. Morphological changes of the human hippocampal formation from midgestation to early childhood. Handb. Dev. Cogn. Neurosci. 45–58.
Seress, L., Ábrahám, H., 2008. Pre- and postnatal morphological development of the human hippocampal formation (2nd Ed.). Handb. Dev. Cogn. Neurosci. 187–211. https://doi.org/10.1039/b610774e.
Shrout, P.E., 2009. Short and long views of integrative data analysis: comments on contributions to the special issue. Psychol. Methods 14 (2), 177–181. https://doi.org/10.1037/a0015953.
Skrondal, A., Laake, P., 2001. Regression among factor scores. Psychometrika 66 (4), 563–575. https://doi.org/10.1007/BF02296196.
Tamnes, C.K., Walhovd, K.B., Engvig, A., Grydeland, H., Krogsrud, S.K., Østby, Y., Holland, D., Dale, A.M., Fjell, A.M., 2014. Regional hippocampal volumes and development predict learning and memory. Article 3–4. Dev. Neurosci. 36 (3–4). https://doi.org/10.1159/000362445.
Teves, J.B., Gonzalez-Castillo, J., Holness, M., Spurney, M., Bandettini, P.A., Handwerker, D.A., 2023. The art and science of using quality control to understand and improve fMRI data. Front. Neurosci. 17, 1100544. https://doi.org/10.3389/fnins.2023.1100544.
Thompson, W.H., Wright, J., Bissett, P.G., Poldrack, R.A., 2020. Dataset decay and the problem of sequential analyses on open datasets. eLife 9, e53498. https://doi.org/10.7554/eLife.53498.
Tozzi, L., Anene, E.T., Gotlib, I.H., Wintermark, M., Kerr, A.B., Wu, H., Seok, D., Narr, K. L., Sheline, Y.I., Whitfield-Gabrieli, S., Williams, L.M., 2021. Convergence, preliminary findings and future directions across the four human connectome projects investigating mood and anxiety disorders. NeuroImage 245, 118694. https://doi.org/10.1016/j.neuroimage.2021.118694.
Van De Schoot, R., Lugtig, P., Hox, J., 2012. A checklist for testing measurement invariance. Eur. J. Dev. Psychol. 9 (4), 486–492. https://doi.org/10.1080/17405629.2012.686740.
White, T., Blok, E., Calhoun, V.D., 2022. Data sharing and privacy issues in neuroimaging research: opportunities, obstacles, challenges, and monsters under the bed. Hum. Brain Mapp. 43 (1), 278–291. https://doi.org/10.1002/hbm.25120.
Widaman, K.F., & Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain.
Wilcox, K.T., Wang, L., 2023. Modeling approaches for cross-sectional integrative data analysis: evaluations and recommendations. Psychol. Methods 28 (1), 242–261. https://doi.org/10.1037/met0000397.
Wisse, L.E.M., Chételat, G., Daugherty, A.M., Flores, R., Joie, R., Mueller, S.G., Stark, C.E. L., Wang, L., Yushkevich, P.A., Berron, D., Raz, N., Bakker, A., Olsen, R.K., Carr, V. A., 2020. Hippocampal subfield volumetry from structural isotropic 1 mm 3 MRI scans: a note of caution (Article May). Hum. Brain Mapp.. https://doi.org/10.1002/hbm.25234.
Yoon, M., Millsap, R.E., 2007. Detecting violations of factorial invariance using data-based specification searches: a Monte Carlo study. Struct. Equ. Model. A Multidiscip. J. 14 (3), 435–463. https://doi.org/10.1080/10705510701301677.