Lifebrain

# D3.7 Refining and validation of novel statistical tools

| | |
|---|---|
| Project title: | Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts |
| Due date of deliverable: | 31th December, 2021 |
| Submission date of deliverable: | 22nd December, 2021 |
| Leader for this deliverable: | Max Planck Institute for Human Development |

| Contributors to deliverable: | Name | Organisation | Role / Title |
|---|---|---|---|
| Deliverable Leader | Ylva Köhncke<br>Andreas Brandmaier | MPIB | Researcher |
| Contributing authors | Ulman Lindenberger | MPIB | PI |
| | Paolo Ghisletta | UNIGE | PI |
| | Rogier Kievit | CAM | Researcher |
| | Øystein Sørensen | UiO | Researcher |
| | Sezen Cekic | UNIGE | Researcher |
| | Ann-Marie G. de Lange | UOXF | Researcher |
| | Dídac Vidal-Piñeiro | UiO | Researcher |
| Reviewer(s) | Klaus Ebmeier | UOXF | WP3 Leader |
| | William Baaré | REGIONH | WP2 Leader |
| | Christian A. Drevon | Vitas | WP Leader |
| Final review and approval | Barbara B. Friedman | UiO | administrative coordinator |

| Document History | | | | |
|---|---|---|---|---|
| Release | Date | Reason for Change | Status (Draft/In-review/Submitted | Distribution |
| 1.0 | 15.11.2021 | Sent to all Lifebrainers for commenting | Draft | Email |
| 1.1 | 01.12.2021 | Sent to Executive Board | In review | Email |
| 1.2 | 13.12.2021 | Revised version | Revision | Email |
| 1.3. | 22.12.2021 | Final version | Submitted | Participant portal |

| Dissemination level | | |
|---|---|---|
| PU | Public | **x** |
| PP | Restricted to other programme participants (including the Commission Services) | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | |
| CO | Confidential, only for members of the consortium (including the Commission Services) | |

# Table of Contents

## Executive Summary

This deliverable is a report on further refinements of the tools developed according to Task 3.4 ("Development of novel longitudinal statistical tools") in Work Package 3 (WP3 Tool Development). This deliverable is an update of deliverable D3.6.

We have continued and extended our work on statistical tools to exploit fully the potential of the combined and harmonized large cohort studies of Lifebrain. As stated in the project proposal, our plan for Task 3.4 was to "address statistical challenges in three steps: (1) comparative analysis of data sets and research designs; (2) development and application of statistical tools; (3) tool refinement and model selection" (Lifebrain proposal, p. 21), in close interaction with the goals and tasks as reported in WP4 (Demonstration). In the context of the first goal of addressing comparative analyses across the Lifebrain cohorts (1), we finalized and published our meta-analysis on SES, brain structure, and cognition across all Lifebrain cohorts. We also developed and published new meta-analysis software for generalized linear mixed models, which are useful to model non-linear trajectories based on longitudinal data over the lifespan. We continued our work on precision, reliability, and power in longitudinal structural equation models (SEM), and published new tools for assessing statistical power in longitudinal SEM and for longitudinal studies with planned missing data. With respect to the second goal of this deliverable (2), we published our previously described practical guide for variable selection and regularization in SEM, a tutorial for joint longitudinal and survival models, and we were involved in numerous applications of the new tools. Finally, with respect to step (3), we advanced our understanding of methods to reduce the dimensionality of the complex multi-modal cross-sectional and longitudinal data in Lifebrain by statistical modelling tools (SEM/ confirmatory factor analysis, clustering methods, age-prediction models) and offer guidance regarding selecting appropriate models of change. We refined the tools and adapted them for the purposes of Lifebrain and similar endeavours beyond.

# List of acronyms / abbreviations

| | |
|---|---|
| CART | Classification and regression tree |
| CFA | Confirmatory factor analysis |
| COMT | Catechol-O-methyltransferase |
| DAT | Dopamine transporter |
| DTI | Diffusion-tensor imaging |
| DRD2 | Dopamine Receptor D2 |
| DWI | Diffusion-weighted imaging |
| GA(M)M | Generalized additive (mixed) model |
| MD | Mean diffusivity |
| MRI | Magnetic resonance imaging |
| MTR | Magnetization transfer ratio |
| PCA | Principal component analysis |
| PET | Positron Emission Tomography |
| PM | Planned missingness |
| RegSEM | Regularized structural equation modelling |
| ROI | Region of interest |
| SEM | Structural equation modelling |
| SES | Socio-economic status |
| VMAT2 | Vesicular Monoamine Transporter 2 |
| WP | Work package |

# 1. Introduction

## 1.1. Background

The Lifebrain consortium brings together major European neuroimaging studies encompassing a large range of measures from different modalities while spanning different age ranges and follow-up times, covering the entire human life span. The researchers of the consortium have defined a wide range of substantive research questions in WP4. The methodological challenges with answering these questions based on the available complex data are the subject of WP3. Within WP3, basic issues such as optimal data management and streamlined pre-processing pipelines are tackled as well as the selection, development, and evaluation of appropriate statistical questions and tools. In our initial proposal for Lifebrain, we had suggested three lines of research concerning statistical modelling, Task 3.4:

(I) Theoretical frameworks and practical tools for comparative analysis of data sets and research designs with respect to effect size and power;

(II) statistical modelling tools to account for the complex nature of lifespan development, including multivariate and dynamic longitudinal structural equation models (SEM), data-driven approaches related to classification and regression trees (CART), and flexible modelling using generalized additive mixed modelling (GAMM);

(III) tool refinement and optimizing of model selection to tackle the important problem of identifying and selecting those models that best summarize data *across* sites.

In deliverable D3.6 we described our work along the first two lines of research (I & II).
In this deliverable (D3.7), we will update on the status of lines I & II and report on progress on research line III.

## 1.2. Objectives

Our objectives (Task 3.4) were to advance the statistical methods for optimal exploitation of the combined large cohorts in Lifebrain. Challenges included differences in the level of measurement and study design. At the level of measurement, cohorts often used different measures as indicators of the constructs of interest that were in turn common between cohorts. At the level of study design, we faced a variable mix of cross-sectional and longitudinal data in various age ranges, differences in sample sizes, measurement precision, measurement reliability, and statistical power to detect effects. Thus, we needed to model these common constructs as precisely as possible in each cohort while allowing for differences in measurement precision, reliability, and statistical power.

We had the goal to "provide a pool of existing and novel longitudinal methods that are specifically suitable to address the challenges of large-scale multi-site longitudinal data and allow testing for various risk and protective factors explaining individual differences in brain and mental health changes over the lifespan" (Deliverable D3.6). Ultimately, we aimed at identifying "models that optimally represent the consolidated findings of the consortium, and hence contribute to theory development and generalizability" (Consortium Proposal, Task 3.4, Page 44).

## 1.3. Collaboration among partners

This task has been led by the MPIB-team in close collaboration with members from UNIGE, UiO, and UCAM. Concerning the role of members involved in statistical modelling, we would like to stress that, since the beginning of Lifebrain, we did not see groups and projects on statistical modelling and tool development as operating in isolation and distant from "substantive" projects, nor as a service unit for statistical counselling. Rather, discussions about (statistical) methods have always been a natural part of all discussions about theory and data, and many more persons engaged in discussions on statistical methods than the contributors of this deliverable. Likewise, the contributors of this deliverable were involved in many of the projects on substantive questions. This is symptomatic of the close entanglement and mutual inspiration of statistical methods development and theoretical and empirical work. This was true for the personal meetings as well as in the virtual meetings during the ongoing Covid-19 pandemic. We aim to make our work in this task sustainable for use for in the Lifebrain community and beyond by publishing practical, open-source tools, practical guidelines, and tutorials, including simulations and use-cases, and related theoretical considerations. The remainder of this document provides details on our contributions.

# 2. Results

## 2.1. Research Line I: Theoretical frameworks and tools for comparative analysis of data sets and designs

### 2.1.1 Meta-analyses in Lifebrain

A meta-analysis combines and synthesizes results from different studies, with the potential to increase robustness, statistical power, and to quantify potential differences between the studies. Estimated effects from the different primary studies are standardized so that their scales are comparable, and they can be derived by different indicators in studies of diverging design (Hedges & Olkin, 2014). A mega-analysis, in contrast, combines the studies not at the level of effects sizes, but at the level of the raw data. In Lifebrain, a meta-analytic approach becomes necessary whenever a mega-analytic approach is not feasible, for instance, if constructs of interest were measured with heterogenous indicators across sites that did not permit harmonization, or if data protection policies do not allow us to share raw data. However, in many cases in which we cannot share the raw data, the collaboration in the consortium still allows us to streamline or harmonize many of the analytical steps leading to the effect size measures. By this, we can minimize non-interesting sources of variation between studies or cohorts much more than a conventional meta-analysis possibly can, when combining effects from different published studies.

***Meta-analysis with parametric models, SEM***

One of the primary research questions in Lifebrain was to investigate links between socio-economic status, brain structure, and general cognitive ability, which is now published in *Cerebral Cortex* (Walhovd et al., 2021). Socio-economic status was indicated by income and education, brain structure by volume, area and thickness of the cerebral cortex, and general cognitive ability was indicated by performance on a range of different tests of fluid intelligence. We had initially planned to take a SEM-approach (see D3.6) but observed that fitting a consistent model across all cohorts produced convergence issues, so that we had to resort to a principal component analysis (PCA) as a more robust, yet similar, approach. As we were primarily interested in general cognitive ability but faced a situation in which this was measured by widely different tasks in the different studies, we extracted the first principal component across cognitive tasks that likely picked up individual differences in general cognitive ability. In our meta-analysis on SES, brain structure, and cognitive ability, we observed remarkable heterogeneity across European cohorts and between European and US cohorts in the degrees to which brain structure and cognitive ability were linked to SES (Walhovd et al., 2021).

***Meta-analysis with nonparametric models, GA(M)M***

Current meta-analytic tools apply meta-analysis to results from parametric statistical models that reduce sample data to a finite set of parameters. Interpretation assumes that this set of parameters captures all information we need to draw inferences for the underlying population and to predict future observations. Non-parametric models do not assume a finite set of parameters; instead, the number of parameters can change when the models are applied to new or more data. They are often regarded as more robust and flexible to account for non-linearities in the data. For the Lifebrain consortium, questions around lifespan development are of core interest. Adequate models of age differences and changes over the lifespan are necessary for more descriptive purposes such as characterization of typical trajectories but are also prerequisite to investigating questions such as whether between-person differences in brain changes are linked to differences in other variables or to changes in these variables over time. In many cases, lifespan trajectories may be best described by a non-parametric model.

To illustrate, we fitted (parametric) linear mixed models with quadratic and cubic polynomials for the age term to hippocampal volume data from the LCBC cohort (4364 observations of 2023 participants, age 4–93 years, 1–8 measurements per participant). As illustrated by Fig. 1, the quadratic fit underestimates the steep increase during adolescence and estimates the hippocampal volume to increase too long into adulthood. The cubic fit captures the volume growth during adolescence better than the quadratic fit but fails to capture the decline that occurs after the age of around 70. For a discussion on polynomials in growth modelling, see also Research Line III. The generalized additive mixed model (GAMM) fit, on the other hand, is flexible enough to capture both the steep increase during adolescence, a period of moderate decline during adulthood, and a steeper decline at old age. GAMMs are extensions of generalized additive models (GAMs; (Hastie & Tibshirani, 1987)) including random effects; thus, applicable to longitudinal or other forms of clustered data.

In our recently published paper “Meta-analysis of generalized additive models in neuroimaging studies” (Sørensen et al., 2021), we describe a method of meta-analysing effects from GAMs or GAMMs, which we developed and implemented in an R package named *metagam* (https://cran.r-project.org/web/packages/metagam/index.html). The method requires a model that relates an outcome of interest to a set of explanatory variables. This model is fitted to data from each cohort. Model estimates are then shared across cohorts such that the expected response and their standard errors at new values of the explanatory variables can be computed.

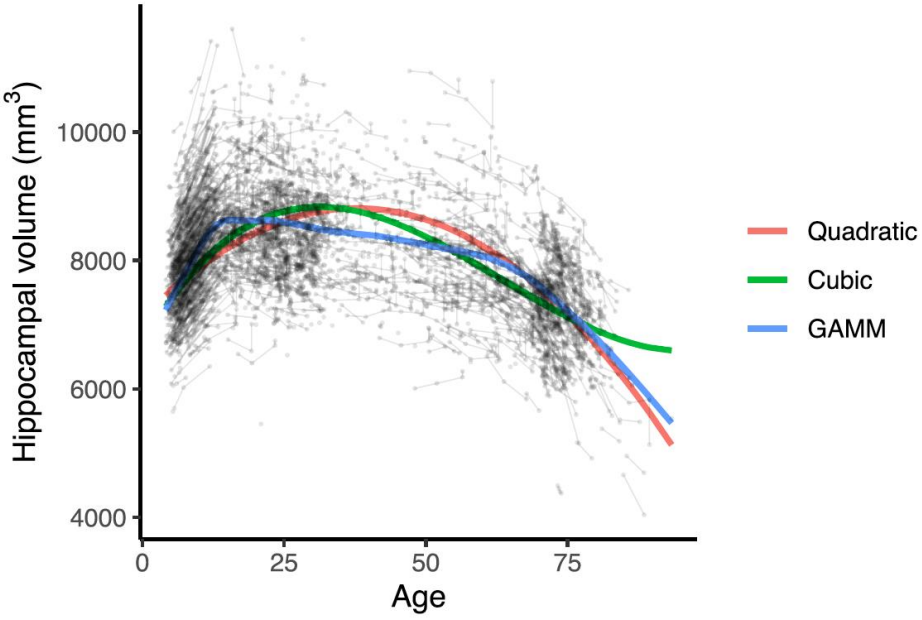


*Figure 1. Example of modelling lifespan trajectories in hippocampal volume with longitudinal data using linear mixed models with quadratic and cubic terms for age, as well as a GAMM. The black dots show individual observations, and the black lines connect subsequent observations from the same individual. The GAMM was fitted with 20 cubic regression splines and a random intercept term for each individual, and the optimal smoothing parameter estimated with restricted maximum likelihood. Reproduced from Sørensen et al., 2021 (CC BY 4.0).*

In simulations of some common scenarios of lifespan data and lifespan research questions, we showed that our proposed methods perform well. We demonstrated an application of *metagam* in a real data analysis on hippocampal volume and self-reported sleep quality data from the Lifebrain consortium. Because we had the full data from the consortium at one place, we were able to compare the results from *metagam* to those from the ideal case, a mega-analysis on all data combined. The results were very similar. For example, the estimated effects of age on hippocampal volume did not differ much between the two approaches. The meta-analytically fitted curve lies somewhat above the mega-analytically fitted curve below age 60 years and has somewhat narrower confidence bands at low ages and wider confidence bands at high ages (Figure 2). A possible reason for the narrow confidence bands of the meta-analytic estimate in the age range 30 to 55 years is that two cohorts dominate this age range as data sources (LCBC and Cam-CAN) and that these two have very similar functional forms.

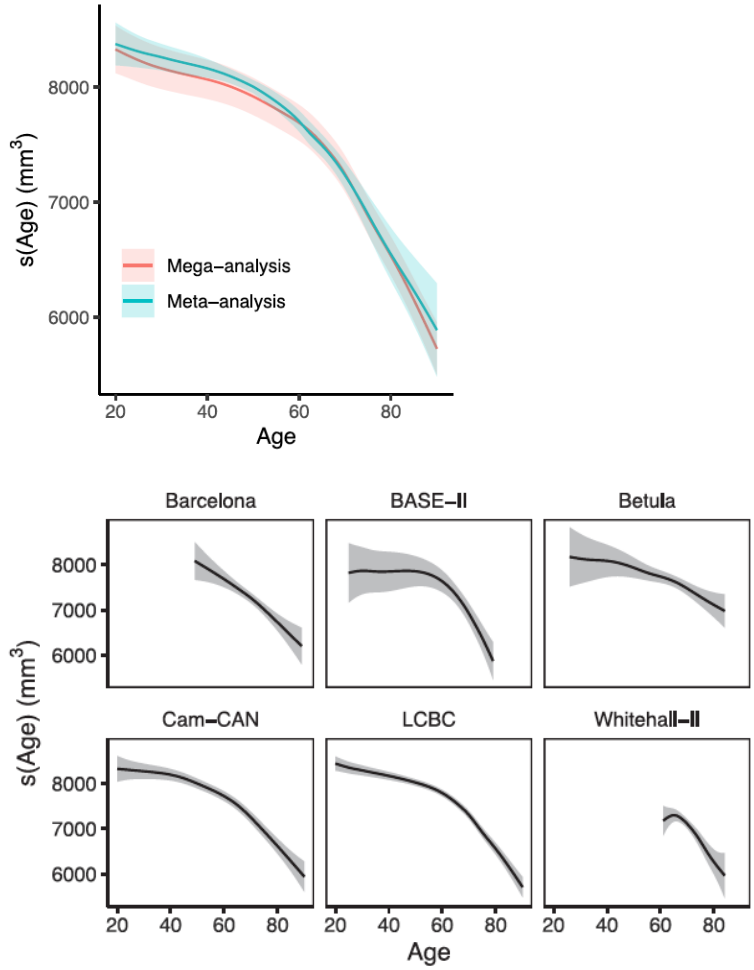


*Figure 2. Top panel: Meta-analytic fits obtained by metagam, compared to the corresponding fit obtained applying GAMM to full data (Mega-analysis). Effect of age on hippocampal volume, including the overall intercept. Bottom panel: Age trajectories of hippocampal volume for each cohort. Shaded areas are 95 % confidence intervals. Reproduced from Sørensen et al., 2021 (CC BY 4.0).*

We introduced two novel plots for inspecting meta-analytic GAMM results (Figure 3). Dominance plots show which age ranges are potentially dominated by individual studies (e.g., if a given study contributes a large fraction of the available data in a given age range). Heterogeneity plots illustrate the heterogeneity across sites for different age ranges. Taken together, these plots help researchers to identify interesting and robust patterns in their data.

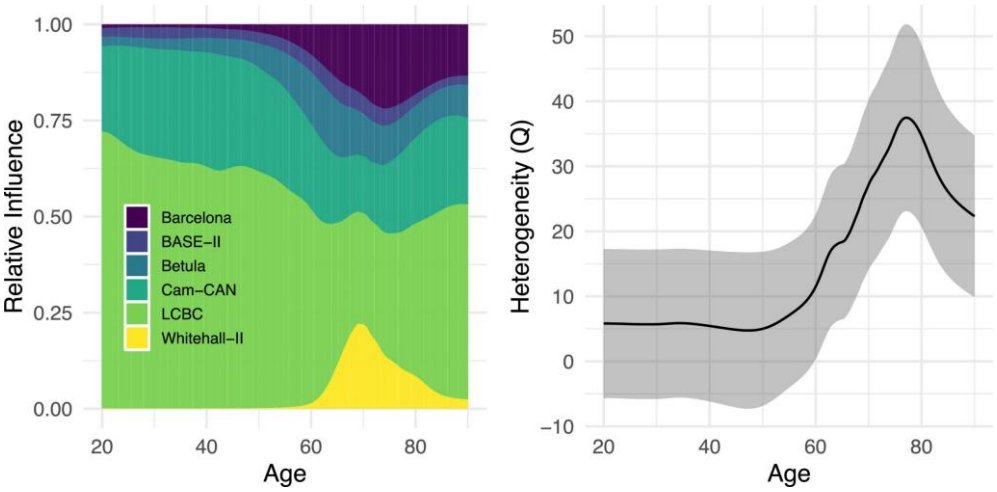


*Figure 3. Dominance and heterogeneity plots. Left panel: the relative contribution from each study to the meta-analytic fit over age. Right panel: Cochran's Q statistic for heterogeneity over age. Shaded areas represent 95 % confidence intervals. Reproduced from Sørensen et al., 2021 (CC BY 4.0).*

We expect that our method of meta-analysing GAMMs has great potential to be applied in more projects within Lifebrain, but also other multi-cohort and consortium studies using neuroimaging data. The method we propose is particularly useful when the studies contribute data in different age ranges, or patient distributions across a clinical variable of interest, because we offer an alternative to the straightforward method of enforcing the same knot placements, which might lead to nonidentified models. Instead, we propose a point-wise meta-analysis, which alleviates these issues and works well in our simulated data and case study on hippocampal volume and sleep quality within Lifebrain. In addition, we expect our method to be especially useful in neuroimaging genetics research because it can handle situations where classic approaches are pushed to their limits, e.g., when genetic effects are studied in interaction with other variables, such as age. In sum, we propose *metagam* will be especially beneficial in lifespan neuroscience and imaging genetics (Sørensen et al., 2021).

### 2.1.2 Structural equation modelling, power, effect size, and reliability

A central methodological aspect of all meta- or mega-analyses is the question of how precise our measurements are. Only when we consider the precision with which each study measures a construct of interest or a relationship between two constructs of interest, can we draw meaningful conclusions about the presence and magnitude of effects. We reported on some of our contributions in this regard in D3.6.

We had theoretically laid out the relationships among the concepts of precision, power, and reliability of a (longitudinal) study design and concluded that "effective error, reliability, and statistical power are different measures that each quantify a study's ability to detect change under different restrictive assumptions (Brandmaier, von Oertzen, Ghisletta, Lindenberger, & Hertzog, 2018)." Based on these theoretical developments, we implemented a tool for evaluating statistical power in latent variable models with focus on longitudinal designs, the R package *semper* (Brandmaier, 2018).

As also reported in D3.6, we have developed a formal general framework for assessing reliability in neuroimaging studies with repeated measurements named intra-class effect decomposition (ICED). The paper was then submitted and published shortly afterwards (Brandmaier et al., 2018).

Building on our previous work on the efficiency of longitudinal study designs to detect change, we assessed efficiency of longitudinal study designs with planned missingness (PM). We expect this topic to be of great interest for researchers planning new longitudinal neuroimaging studies, but also for researchers working with existing data sets with complex missing data patterns. In general, PM designs are a way of optimizing the efficiency of a longitudinal study by deliberately omitting some of the measurements for some of the participants. Collecting longitudinal data is much more time-consuming and cost-intensive than collecting cross-sectional data, especially in cases of costly data collection as is often the case in neuroscience. However, when investigating changes over the lifespan, longitudinal data are crucial (Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011; Raz & Lindenberger, 2011). We had previously introduced the concept of effective error (D3.6) and how we used it to find optimal longitudinal study designs for complete data; here, we extend the approach to PM designs.

The Lifebrain consortium set out to investigate statistical power and optimal research design for future longitudinal studies of similar kind. We have developed an analytical approach to estimate the statistical precision (that is, precision of measurement, reliability, and statistical power) of longitudinal study designs with planned missing data (Brandmaier, Ghisletta, & von Oertzen, 2020). Longitudinal planned missing data designs prescribe patterns of measurement occasions that vary across participants; that is, not all participants are measured at every time point. Our formal solution allows gauging the efficiency of different planned missing data designs against each other and against a complete data design in terms of both statistical precision and cost efficiency/feasibility.

Theoretical contributions with potential benefit for the Lifebrain consortium and involving single members include, for example, Anvari, Kievit, and colleagues' (2021) work on the practical relevance of effect sizes in psychological research, which has the potential to help researchers to make their claims about generalizability of findings more transparent, and to assess whether psychological findings can be applied in daily life. This work is currently at the preprint stage (Anvari et al., 2021).
Another theoretical contribution, published in Perspectives on Psychological Science (Borsboom, van der Maas, Dalege, Kievit, & Haig, 2021), offers psychological researchers practical guidance in theory formation.

## 2.2. Research Line II: Statistical modelling tools using longitudinal SEM, CART, and joint longitudinal and survival models

### 2.2.1 The Variable Selection Problem: Novel approaches using SEM Trees, SEM Forests, and Regularized SEM

We reported in D3.6 our contributions to the development of new approaches combining theory-based SEM with data-driven techniques; in this case, classification and regression trees (CART), random forests, and regularized regression. Model-based trees, such as SEM Trees (Brandmaier, von Oertzen, McArdle, and Lindenberger, 2013), recursively split observed data into homogeneous groups sharing similar parameters of a model. These model-based trees can be thought of as adaptive multiple group models, in which the group structure and its predictors are learnt from the data. This allows us to explore those variables (and interactions between variables) that best predict differences in multivariate outcomes. Particularly, this includes the possibility to explore predictors of individual differences in models of (correlated) change over time.

SEM Forests (Brandmaier, Prindle, McArdle, & Lindenberger, 2016) are a recent extension of SEM Trees. They are large ensembles of SEM Trees, comprising hundreds to thousands of individual trees, each based on a random sample of the original data. By aggregating the predictive information in a forest, one obtains a measure of variable importance that is more robust than corresponding measures from single trees. Another way of identifying the most important predictor variables from a larger set is through regularization. The cross-site methodological collaboration on the topic of regularization in Lifebrain joined forces with Ross Jacobucci (University of Notre Dame), which led to a methodological contribution that included the well-established technique of regularized regression with SEM (Jacobucci, Grimm, & McArdle, 2016). Regularization is a regression model with many predictors that penalizes small regression coefficients and thus retains only the most important predictors. In regularized SEM, a structural equation model can be reduced to the most important paths in a data-driven manner.

Our work on regularized SEM has resulted in a publication of a tutorial, or practical guide, with one use-case in which we applied regularized SEM to data from the Cam-CAN cohort to predict visual short-term memory performance with microstructural properties in a number of white matter tracts (Jacobucci, Brandmaier, & Kievit, 2019). We reported on this in D3.6, at the time it was still a preprint.

These new tools that combine data-driven ways of variable selection with theory-based SEM have since been applied in several published papers within Lifebrain. To start with applications of SEM trees, our first project with SEM trees within Lifebrain (D3.6) investigated age differentiation. Specifically, we investigated the question of whether intercorrelated neural and cognitive variables will be less correlated with advancing age. In this study, published in the *Journal of Neuroscience* in 2018 (de Mooij, Henson, Waldorp, & Kievit, 2018), the SEM tree approach helped to explore the age range in search for meaningful differences.
These research activities influenced two further recently published papers (which are not part of Lifebrain, but involve a Lifebrain member), where SEM trees were used to explore age differences in the associations between white matter microstructure and cognitive abilities (working memory and processing speed as predictors of fluid intelligence). Based on data from two cohorts of children and adolescents between 6 and 18 years of age, of which one cohort is part of Lifebrain (CALM), our SEM tree analyses suggested that many of the associations between the microstructure of single white matter tracts and single cognitive abilities were weaker during prepuberty and early puberty and stronger before and after puberty, in line with the notion of a reorganization of the neurocognitive architecture in these age ranges (Fuhrmann, Simpson-Kent, Bathelt, Rogier, & Kievit, 2020; Simpson-Kent et al., 2020).
Furthermore, we have investigated how regularization can help to select from a large set of candidate genes to find a model that best predicts differences in working memory (Karalija et al., 2021). This allowed us to identify four genes (DRD2, DAT, COMT, and VMAT2) that jointly formed a prediction model that performed well in the BASE-II data, which is part of the Lifebrain consortium, whereas we found a simple effect of one of the four genes (DAT) in a replication sample (the Cognition, Brain, and Aging study, (Nevalainen et al., 2015)). Using PET imaging, we found that this gene also predicted differences in *in vivo* dopamine, thus showing that DA function contributes to differences in working memory performance in old age, presumably by regulating dopamine availability.

### 2.2.2 Joint longitudinal-survival models

In this Lifebrain research project, we focussed on combining flexible models that describe intraindividual trajectories over time with discrete event data. This statistical approach is called a joint longitudinal-survival model, and it allows for studying how parameters that describe individuals' trajectories are best modelled jointly with a survival model describing the hazard associated with a specific outcome event, e.g., disease onset or death. We have reported on the details of the approach in D3.6, and how we solved difficulties of its practical implementation, including questions of how to decide for a suitable longitudinal model, how to relate the longitudinal model to the survival model, as well as interpretation issues.

We have recently published a tutorial on this method (Cekic, Aichele, Brandmaier, Köhncke, & Ghisletta, 2021) that helps researchers understand and use the tool when investigating questions about longitudinal, time-varying factors that predict a discrete event to happen, such as what changes in brain structure or behaviour that predict institutionalisation, or death. We expect this to be useful for the Lifebrain community and researchers interested in modelling lifespan and ageing data.

## 2.3. Research Line III- tool refinement and optimizing of model selection

In this research line, we worked on the refinement of various tools. We will first report about our advances on model selection – which models are best suited to summarize data across sites. We have touched upon the differences between mega-analysis and meta-analysis and have reported on our work on meta-analytic techniques to model changes using meta-GAM in Research Line I. Here, we discuss the broader topic of dimensionality reduction as a way of "summarizing data," and the applicability of a range of approaches of dimensionality reduction to cross-sectional and longitudinal data, in meta- and mega-analysis in Lifebrain. We will give some examples of how we refined different concrete tools related to these questions. Last, we discuss the utility of growth modelling for questions of lifespan development.

### 2.3.1 Refinement and model selection: multivariate modelling and dimensionality reduction

Neuroimaging data are often high-dimensional, which means that there are many measured variables (typically, thousands to tens of thousands of voxels), which also are correlated with each other to different degrees. At the same time, the space of concepts of interest is of much lower dimensionality; that is, we are interested in a smaller set of more general characteristics (specific regional functional activation patterns or notions of functional and structural integrity). In these cases, techniques of dimensionality reduction are required. An advantage with having recorded many measures of the same general characteristic instead of only one is robustness; individual measures are often biased and noisy. By aggregating across them, we expect to reduce the influence of measurement-specific bias and noise, and extract the commonality representing the general characteristic of interest. Dimensionality reduction techniques can be generally categorized into the broad classes of supervised/ theory-based/ confirmatory approaches, or unsupervised/ data-driven/ exploratory approaches. In theory-based dimensionality reduction, we have an *a priori* theory that specifies which measures relate to which theoretical construct.
In data-driven dimensionality reduction, we only expect or assume that there is shared information across the measures, and we interpret the meaning of the commonness or lower-dimensional structure after we apply dimensionality reduction.

### 2.3.2 Theory-based dimensionality reduction

The theory-based approach has a long tradition in psychometrics for measuring not directly observable, latent, constructs such as “extraversion” or “intelligence.” In this tradition, the flexible framework of structural equation models is often used to build measurement models (which are confirmatory factor models) to capture a latent construct and then relate it to other constructs of interest. A measurement model is also a theory-based way to reduce the dimensionality of a set of indicators of a theoretical construct. The theoretical construct is represented by a latent (unobserved) variable, measured by the indicator variables, or manifest (observed) variables. The latent variable captures the shared variance of the manifest variables, apart and independent from the specific variance each manifest variable has. Thus, the dimensionality of the data is reduced from the number of observed variables to the number of theoretical constructs that we are interested in. In such a theory-based application (Köhncke et al., 2021), we built a model of grey matter integrity based on measures from three different structural MR imaging modalities; that is, grey matter intensity from T1-weighted MRI, mean diffusivity (MD) from diffusion-weighted MRI (DWI), and magnetization-transfer ratio (MTR) from magnetization transfer imaging.

These measures are brought about by very different physical processes, and capture different tissue properties, but have in common that they are all interpreted as indicators of the integrity of the tissue. Applying SEM to data from the BASE-II study, we observed that there is enough common variance across these three modalities for each of several brain regions of interest (ROI) to be captured in a modality-general factor of (structural) integrity. In other words, the variance-covariance structure across these three imaging modalities in several episodic memory-related ROIs is well represented by ROI-wise integrity factors. We interpret these factors as a general but regional characteristic that can be interpreted as *integrity*. The fact that we can reduce the set of modality-specific indicators to a modality-general factor shows that individuals who are better off in one of the indicators tend to be better off in the others as well. We can only speculate about the biological processes that may be responsible for this pattern to emerge. We are currently working on longitudinal extensions of this model. In doing so, we are also thinking about ways to generalize the applicability of such a longitudinal model to be able to include data from other Lifebrain cohorts.

In a different project, in collaboration of several Lifebrain members from MPIB and UmU with Elliot Tucker-Drob from the University of Texas, Austin, we investigated the dimensionality of cognitive changes in late adulthood using longitudinal data from a Lifebrain cohort (Betula, UmU) and a US-cohort, VCAP (Salthouse, 2017). Theories of adult cognitive development classically distinguish between fluid abilities, which require effortful processing at the time of assessment, and crystallized abilities, which require the retrieval and application of previously acquired knowledge. On average, fluid abilities decline throughout adulthood, whereas crystallized abilities show stability, or even gains until later adulthood. These diverging age trends, along with marked individual differences in rates of change, have led to the proposition that individuals might compensate for fluid declines with crystallized gains. Here, we show that rates of change are strongly correlated across fluid and crystallized abilities. Hence, individuals showing greater losses in fluid abilities tend to show smaller gains, or even losses, in crystallized abilities). This observed commonality between fluid and crystallized changes places constraints on theories about compensation and directs attention towards domain-general drivers of adult cognitive decline and maintenance. The manuscript is submitted to Science Advances and is currently under (minor) revision.

Theory-implied dimensionality is also a central issue of another recently added Lifebrain research project. In this project, we plan to examine comprehensively the dimensionality underlying individual differences in longitudinal changes in cognitive performance during childhood.

Furthermore, we investigate the question of whether the theory of differentiation of cognitive abilities during childhood - a general factor of intelligence becoming less influential, and cognitive abilities becoming more differentiated during childhood and early adolescence - is supported by longitudinal measurements of cognitive abilities from Lifebrain cohorts. We will apply SEM to simultaneously estimate differences between dimensionality of cognitive ability in early and late childhood both with regard to levels and changes.

### 2.3.3 Data-driven dimensionality reduction

In many neuroimaging studies, exploratory, data-driven techniques are used to reduce dimensionality. Techniques such as Exploratory Factor Analysis (EFA) reduce measurements from a large number of brain regions to a tractable number of factors without *a priori* assumptions on the number of factors or the mapping of brain regions to factors. However, dimensionality reduction often ignores relevant *a priori* knowledge about the structure of the data. For example, it is well established that the brain is highly symmetric. In a recent paper we (a) show the adverse consequences of ignoring *a priori* structure in factor analysis; (b) propose a technique to accommodate structure in EFA using structured residuals (EFAST), and (c) apply this technique to three large and varied brain imaging network datasets, demonstrating the superior fit and interpretability of our approach (Van Kesteren & Kievit, 2021). We provide an R software package (https://rdrr.io/github/vankesteren/efast/) to enable researchers to apply EFAST to other suitable datasets.

As an example of tool refinement, we advanced our understanding of yet another type of data-driven dimensionality-reducing statistical model in neuroscience, brain age prediction models (Ann-Marie G De Lange et al., 2020; Vidal-Pineiro et al., 2021). In these models, machine learning algorithms are trained on a wide range of magnetic resonance imaging (MRI) scans to build a normative trajectory of age-related brain differences and project a broad variety of correlated brain characteristics onto a single quantity per individual. Prediction models can then be applied to new data, providing an estimate of brain-predicted age for each individual in the new dataset. The difference between an individual's brain-predicted and chronological age (brain age delta) has become an influential index for brain health and has been associated with clinical risk factors as well as neurological and neuropsychiatric conditions. It is often interpreted as a proxy for deviations from expected age, assumed partially to reflect the rate of brain aging. Vidal-Piñeiro et al. (2021) explicitly tested this assumption in UK Biobank data and a combined dataset from five Lifebrain cohorts (mega-analysis). Counter to most common interpretations of this index, we found no association between cross-sectional brain age and steeper brain decline. Rather, brain age in adulthood was associated with early-life influences indexed by birth weight and polygenic scores.

The results call for much more nuanced interpretations of cross-sectional indices of the aging brain such as brain age (delta). Whereas this is important as a substantive result, we would also argue that our focus on comparing cross-sectional and longitudinal evidence from brain-age prediction models is a crucial step to further refinement of this kind of dimensionality reduction methods in neuroimaging. It will be of importance and interest for the lifespan neuroimaging community, because brain age prediction gains popularity in neuroimaging as more big data are available, and such prediction models offer a way to explore them with respect to an outcome of interest. This work has just been published in *eLife* (Vidal-Pineiro et al., 2021).

A recently established research project within Lifebrain aims to model multimodal longitudinal imaging as a set of different patterns of brain aging (independent components of longitudinal brain change). We aim to pair these different patterns of brain aging with cognitive data (memory tests) and genetic data by computing pathway-specific polygenic risk scores (PGS) indexing genetic liability to neurodegeneration.
Whereas a variety of machine-learning algorithms can provide accurate predictions of age based on brain characteristics, there is significant variation in model accuracy reported across studies, and researchers need valid metrics to assess the performance of their models at hand. Thus, we investigated the comparability of common performance metrics for brain-age prediction models (Ann-Marie G. de Lange et al., 2021).
We predicted age based on neuroimaging data in Cam-CAN and UK Biobank, and assessed the effects of age range, sample size, and age-bias correction on the model performance metrics r, R2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). We observed that these metrics vary considerably depending on the cohort age range. Across subsets with different age ranges, performance metrics improve with increasing sample size. Performance metrics further vary depending on prediction variance as well as mean age difference between training and test sets, and age-bias corrected metrics indicate high accuracy - also for models showing poor initial performance. In conclusion, performance metrics used for evaluating age prediction models depend on cohort and study-specific data characteristics and cannot be directly compared across different studies.

A different exploratory approach to data is to identify modes of variation in the high-dimensional data (Smith et al., 2020), not related to any specific outcome (such as age in brain age prediction). We plan to adapt this approach to Lifebrain (mega-analysis) with the major advantage of being able to compare cross-sectionally derived modes of variation to such derived from longitudinal data.

### 2.3.4 Model selection: Theoretical considerations on the use of growth modelling, longitudinal clustering, and Granger-causality

Growth models (GM), such as mixed-effects models and latent growth curve models, have become popular methodological tools in lifespan research. One of the major advantages of GM is their flexibility in studying individual differences in change. We scrutinized the change functions of GM used in five years of publications on cognitive aging (Ghisletta et al., 2020). A large majority of studies applied polynomial decomposition, and only a small fraction considered exponential or yet another non-linear function of change. However, it is conceptually and empirically plausible to assume exponential decline from adulthood to old age. This calls for exploring what conclusions about individual differences in change are likely to be drawn if one applies linear or quadratic GMs to data simulated under a model of exponential cognitive decline. Hence, we set up a simulation that manipulated the rate of exponential decline, measurement reliability, number of occasions, interval width, and sample size. True rate of decline and interval width influenced results strongly, number of occasions and measurement reliability exerted a moderate effect, and the effects of sample size appeared relatively minor. We encourage researchers to also consider plausible nonlinear change functions when studying behavioural development over the lifespan (Ghisletta et al., 2020).

As another contribution to model selection, we have compared clustering methods for longitudinal data (Taushanov & Ghisletta, 2020). Specifically, we compare a Hidden Mixture Transition Distribution (HMTD) and its clustering performance to the popular Growth Mixture Model (GMM), as well as to the recently introduced GMM based on individual case residuals (ICR-GMM). In conclusion, we see slight advantages of GMM in cases when users have strong theoretical reasons in favour of an explicit mathematical change function to describe the individual sequences. The HMTD approach comes with parametric parsimony (mean, variance, and auto-regressive parameter), is quite easily estimated and adapts well to different types of trajectories even in small samples (Taushanov & Ghisletta 2020).
In neuroscience, there is an increased interest in finding causal links between signals recorded at different brain locations during an experiment. We proposed a systematic methodological review and an objective criticism of existing methods enabling the derivation of time, frequency, and time-varying Granger-causality statistics in neuroscience. In this article, we first present a general framework of Granger-causality statistics in the time domain. Then, the spectral and the time-varying extensions are discussed, with their estimation and distributional properties. Links to partial and conditional Granger causality, dynamic causal modelling, directed transfer function, directed coherence, partial directed coherence, and their variant, are also mentioned (Cekic, Grandjean, & Renaud, 2018).

We expect this contribution to be useful for any neuroscientist interested in the causal interpretation of time-series data, where one time series may precede and potentially influence another time series.

## 3. Conclusion

We have refined and extended the set of tools needed to answer the substantive questions as proposed in work package 4. We have engaged in the development and comparison of statistical models that allow for aggregating data across sites and answering more general, substantive questions that reach far beyond the Lifebrain consortium and hopefully will serve to establish best practices in other fields as well. As a particular challenge, we actively advanced our understanding of longitudinal methods as compared to the more established cross-sectional approaches. Methodological discussions have always been intertwined with discussions of substantive questions and theoretical perspectives. We are gratified to report that this led to several newly proposed research projects within Lifebrain beyond the initial agenda of the original grant proposal. To the greatest extent possible we have made our tools and innovations freely available to the public through preprints, Open Access publications, and Open Source software packages. For the remaining lifetime of the consortium, and beyond, we will continue to dedicate our time and effort to meet the needs of the continuing projects requiring data-analysis.

# 4. References

Anvari, F., Kievit, R., Lakens, D., Przybylski, A. K., Tiokhin, L., Wiernik, B. M., & Orben, A. (2021). Evaluating the practical relevance of observed effect sizes in psychological research.

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 1745691620969647.

Brandmaier, A. M. (2018). semper. R package version 0.3. Retrieved from https://github.com/brandmaier/semper

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods, 21*(4), 566–582. doi:10.1037/met0000090

Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Lindenberger, U., & Hertzog, C. (2018). Precision, Reliability, and Effect Size of Slope Variance in Latent Growth Curve Models: Implications for Statistical Power Analysis. *Frontiers in Psychology*.

Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychol Methods, 18*(1), 71-86. doi:10.1037/a0030001

Brandmaier, A. M., Wenger, E., Bodammer, N. C., Kühn, S., Raz, N., & Lindenberger, U. (2018). Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). *eLife, 7*, e35718.

Cekic, S., Aichele, S., Brandmaier, A. M., Köhncke, Y., & Ghisletta, P. (2021). A tutorial for joint modeling of longitudinal and time-to-event data in R. *Quantitative and Computational Methods in Behavioral Sciences, 1*.

Cekic, S., Grandjean, D. M., & Renaud, O. (2018). Time, frequency, and time-varying Granger-causality measures in neuroscience. *Statistics in Medicine, 37*(11), 1910-1931.

de Lange, A.-M. G., Anatürk, M., Rokicki, J., Han, L. K. M., Franke, K., Alnæs, D., . . . Cole, J. H. (2021). Mind the gap: performance metric evaluation in brain-age prediction. *BioRxiv*, 2021.2005.2016.444349. doi:10.1101/2021.05.16.444349

De Lange, A.-M. G., Anatürk, M., Suri, S., Kaufmann, T., Cole, J. H., Griffanti, L., . . . Singh-Manoux, A. (2020). Multimodal brain-age prediction and cardiovascular risk: The Whitehall II MRI sub-study. *Neuroimage, 222*, 117292.

de Mooij, S. M., Henson, R. N., Waldorp, L. J., & Kievit, R. A. (2018). Age differentiation within gray matter, white matter, and between memory and white matter in an adult life span cohort. *Journal of Neuroscience, 38*(25), 5826-5836.

Fuhrmann, D., Simpson-Kent, I. L., Bathelt, J., Rogier, C. T. H. J. G. S. A. D. M. T. K., & Kievit, R. A. (2020). A hierarchical watershed model of fluid intelligence in childhood and adolescence. *Cerebral Cortex, 30*(1), 339-352.

Ghisletta, P., Mason, F., von Oertzen, T., Hertzog, C., Nilsson, L.-G., & Lindenberger, U. (2020). On the use of growth models to study normal cognitive aging. *International Journal of Behavioral Development, 44*(1), 88-96.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association, 82*(398), 371-386.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*: Academic press.

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A Practical Guide to Variable Selection in Structural Equation Models with Regularized MIMIC Models. *Adv Methods Pract Psychol Sci, 2*(1), 55-76. doi:10.1177/2515245919826527

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Struct Equ Modeling, 23*(4), 555-566. doi:10.1080/10705511.2016.1154793

Karalija, N., Köhncke, Y., Düzel, S., Bertram, L., Papenberg, G., Demuth, I., . . . Brandmaier, A. M. (2021). A common polymorphism in the dopamine transporter gene predicts working memory performance and in vivo dopamine integrity in aging. *Neuroimage, 245*, 118707. doi:https://doi.org/10.1016/j.neuroimage.2021.118707

Köhncke, Y., Düzel, S., Sander, M. C., Lindenberger, U., Kühn, S., & Brandmaier, A. M. (2021). Hippocampal and parahippocampal gray matter structural integrity assessed by multimodal imaging is associated with episodic memory in old age. *Cerebral Cortex, 31*(3), 1464–1477. doi:10.1093/cercor/bhaa287

Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging, 26*(1), 34–47. doi:10.1037/a0020525

Nevalainen, N., Riklund, K., Andersson, M., Axelsson, J., Ögren, M., Lövdén, M., . . . Nyberg, L. (2015). COBRA: A prospective multimodal imaging study of dopamine, brain structure and function, and cognition. *Brain Research, 1612*, 83-103. doi:http://dx.doi.org/10.1016/j.brainres.2014.09.010

Raz, N., & Lindenberger, U. (2011). Only time will tell: Cross-sectional studies offer no solution to the age–brain–cognition triangle: Comment on Salthouse (2011). *Psychological Bulletin, 137*(5), 790–795. doi:10.1037/a0024503

Salthouse, T. A. (2017). Contributions of the individual differences approach to cognitive aging. *The Journals of Gerontology: Series B, 72*(1), 7-15.

Simpson-Kent, I. L., Fuhrmann, D., Bathelt, J., Achterberg, J., Borgeest, G. S., & Kievit, R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Developmental cognitive neuroscience, 41*, 100743.

Smith, S. M., Elliott, L. T., Alfaro-Almagro, F., McCarthy, P., Nichols, T. E., Douaud, G., & Miller, K. L. (2020). Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife, 9*, e52677.

Sørensen, Ø., Brandmaier, A. M., Macià, D., Ebmeier, K., Ghisletta, P., Kievit, R. A., . . . Fjell, A. (2021). Meta-analysis of generalized additive models in neuroimaging studies. *Neuroimage, 224*, 117416. https://doi.org/10.1016/j.neuroimage.2020.117416

Taushanov, Z., & Ghisletta, P. (2020). The Use of a Hidden Mixture Transition Distribution Model in Clustering Few but Long Continuous Sequences: An Illustration with Cognitive Skills Data. *Symmetry, 12*(10), 1618.

Tucker-Drob, E.M., de la Fuente, J., Köhncke, Y., Brandmaier, A.M., Nyberg, L., Lindenberger, U., (in press): A strong dependency between changes in fluid and crystallized abilities in human cognitive aging. *Science Advances*.

Van Kesteren, E.-J., & Kievit, R. A. (2021). Exploratory factor analysis with structured residuals for brain network data. *Network Neuroscience, 5*(1), 1-27.

Vidal-Pineiro, D., Wang, Y., Krogsrud, S. K., Amlien, I. K., Baaré, W. F. C., Bartres-Faz, D., . . . Fjell, A. (2021). Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change. *eLife, 10*, e69995. doi:10.7554/eLife.69995

Walhovd, K. B., Fjell, A. M., Wang, Y., Amlien, I. K., Mowinckel, A. M., Lindenberger, U., . . . Brandmaier, A. M. (2021). Education and Income Show Heterogeneous Relationships to Lifespan Brain and Cognitive Differences Across European and US Cohorts. *Cerebral Cortex*. doi:10.1093/cercor/bhab248