Science Bulletin 68 (2023) 795–799

Contents lists available at ScienceDirect

Science Bulletin

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

Perspective

Six cornerstones for translational brain charts

Zi-Xuan Zhou [a,b], Li-Zhen Chen [a,b], Michael P. Milham [c], Xi-Nian Zuo [a,b,d,e,f,*], Lifespan Brain Chart Consortium (LBCC) [1]

[a] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[b] Developmental Population Neuroscience Research Center, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
[c] Child Mind Institute, New York 10022, USA
[d] National Basic Science Data Center, Beijing 100190, China
[e] Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[f] Laboratory of Cognitive Neuroscience and Education, School of Education Science, Nanning Normal University, Nanning 530001, China

It is of great scientific and translational promise to formulate a normative reference for the lifespan development of human brain to precisely quantify individual differences. By aggregating more than 120,000 brain imaging scans across the world, the Lifespan Brain Chart Consortium (LBCC) recently published brain charts for the human lifespan in *Nature* [1]. These charts (two examples showed in Fig. 1a) have revealed previously undocumented neurodevelopmental milestones, marking a research model on team working for the neuroimaging community towards population neuroscience [2]. The LBCC team demonstrated that after decades of advancement and accumulation in technologies, methods, and resources, we now have a tangible opportunity to achieve translational science for brain health. Accordingly, the World Health Organization has articulated the great clinical and public health relevance of lifespan brain charts in its recent position paper [3].

Despite the impressive advances, there is still a non-negligible gap between this seminal paradigm on basic research of brain charts and their translational applications, calling for great community efforts to address translation barriers. To guide gap-filling research on translational brain charts (TBC), we outline 6E (Exploit, Evaluate, Explore, Eliminate, Estimate, and Establish) efforts that we regard as "cornerstones" of TBC research here (Fig. 1b). Among the spectrum of multidisciplinary efforts, the first three cornerstones dissect aspects that require in-depth evaluation (with an emphasis on neuroscience research), the next two cornerstones point to the need for careful modeling with the acquired data (which relies heavily on statistics & algorithms), and the last cornerstone suggests extensive collaborations with open platforms (to facilitate interdisciplinary research).

We expect that the six mutually supporting principles will prospectively inspire future TBC research practices, solidify TBC foundations, and accelerate the realization of clinical research and diagnosis with TBC. We believe that in reaching eventual translations, both basic and clinical research can promote each other, leading to breakthroughs in developmental population neuroscience [4].

Cornerstone 1: exploit optimal sampling strategies to obtain representative samples according to application needs. Missing representative samples is a main factor hampering the construction of TBC. While the LBCC team pooled worldwide available data with a very large sample size, the samples still lack representativeness. Sampling for TBC should follow methodologies from demography and survey research, considering handedness, education, and other sociodemographic and clinical characteristics [5]. Similar sample sizes should be collected for different sexes, and the proportion of samples in an age group to the total sample size should be determined by the distribution patterns and age-related changes of the measurements. For cross-sectional data, studies have shown that generalized additive models for location, scale, and shape (GAMLSS) require tens of thousands of samples to accurately model simple growth reference charts (e.g., reference centiles for weight). Given the more sophisticated distributions and trajectories of brain metrics, and the lower measurement reliability, the required sample size for TBC would be even larger. Acquiring longitudinal measurements can reduce the required sample size, but how much the sample size can be reduced depends on the specific longitudinal design (accelerated or single cohort for a certain age range), including longitudinal time interval, total duration, expected dropout rate, proposed modeling approach, etc. Note that dropout may compromise sample representativeness and countermeasures should be considered accordingly. The design of these conditions still needs to be optimized by exploration in silico and on real datasets.

Practical sampling strategies can be roughly divided into two paths according to application needs. The first path targets regions or ethnic groups with large populations, collecting very large-scale cross-sectional data to adequately sample the population of interest, avoiding dropout issues and reducing logistical challenges. The

* Corresponding author.
E-mail address: xinian.zuo@bnu.edu.cn (X.-N. Zuo).
[1] All members of the Lifespan Brain Chart Consortium can be found at https://github.com/brainchart/lifespan.

https://doi.org/10.1016/j.scib.2023.03.047

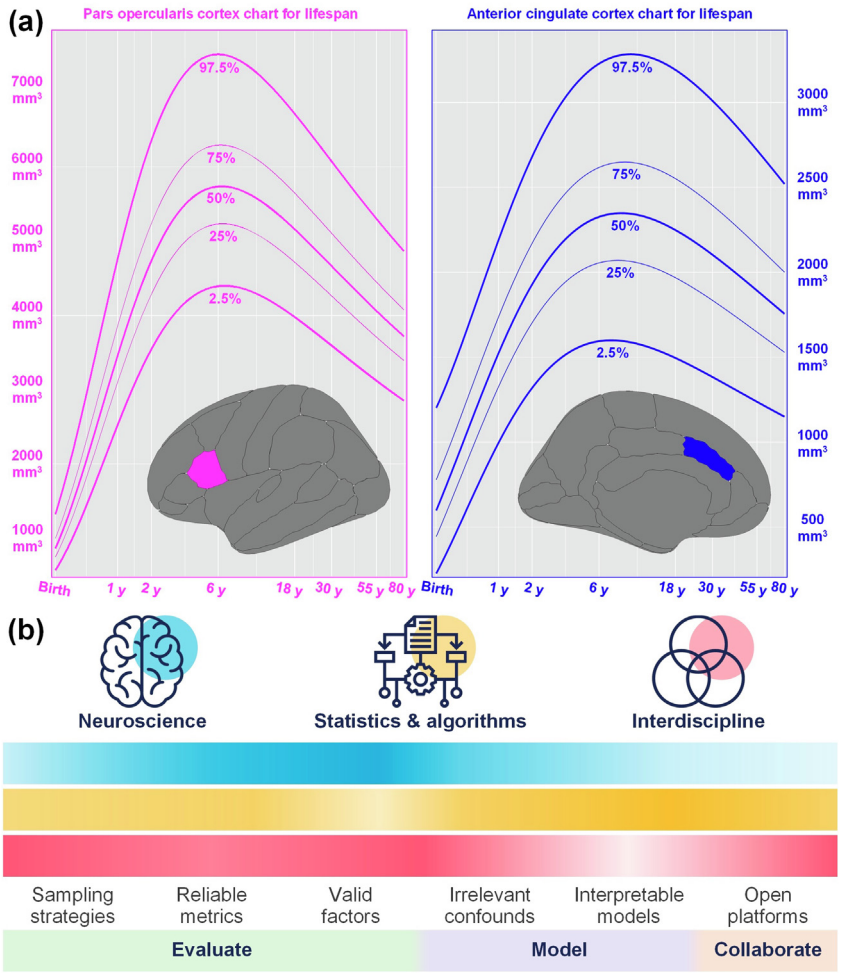

Fig. 1. Brain charts and their cornerstones. (a) The lifespan brain charts are constructed and depicted with the five centile curves for the pars opercularis cortex (pink, female) and the caudal anterior cingulate cortex (blue, male); (b) The spectrum of six cornerstones for translational brain charts. y: years.

brain charts modeled in this path will be of reference significance to a wide range of populations. Furthermore, very large sample sizes increase the feasibility of modeling multivariate features (see Cornerstone 5), allowing the exploration of latent multivariate associations over the lifespan. The second path targets specific populations in specific age ranges, such as young professional athletes or indigenous adolescents living at high altitudes, with longitudinal measurements of moderate sample sizes to meet specific application needs.

Cornerstone 2: evaluate the reliability of metrics not only between but also within individuals or states. Concerning reliability, test–retest reliability is generally considered [6,7], where high reliability requires low intra/within-individual variability and high inter/between-individual variability. For TBC, it may be particularly necessary to control covariates such as age and sex whenever possible when evaluating inter-individual variability. Here, we note that inter-state (e.g., between ages or between clinical characteristics) variability should also be evaluated according to application needs of TBC, based on within-individual (i.e., longitudinal) or between-group comparisons. Intra-class correlation is a commonly used index to quantify reliability, which is defined as the ratio of the variance between classes (e.g., between individuals or between states) to the total variance. A more recently proposed index, discriminability, can be used as an alternative to better deal with non-Gaussian distributions [8]. An immediate benefit of increasing measurement reliability is that the required sample size can be reduced for a given effect size and statistical power. Reliability is also a necessity for validity (see Cornerstone 3), and thus the neuroscience community has increased its focus on reliability assessment in the last decade [7]. Morphological magnetic resonance imaging (MRI) measurements have almost perfect measurement reliability for core metrics, which is a key reason why the LBCC team focused on morphological development by leveraging brain chart models [1].

It is imperative to develop more reliable neuroimaging metrics for TBC research (e.g., reliability >0.8 to serve the clinical standard). More anatomical MRI metrics seem reliable and worth further investigation. Functional MRI (fMRI) studies have produced massive metrics but were bottlenecked by the convergence into reliable and uniform measurements of the human brain function (see Ref. [9] for recent advances and exceptions). The gradient metrics of functional connectivity have been shown to well reflect the intrinsic brain hierarchy [10], which are relevant to cognition, psychopathology, development, and evolution (see reviews in Ref. [10–12]). There is also potential to develop reliable metrics related to lifespan development from other brain imaging techniques (e.g., magnetoencephalography or functional near-infrared spectroscopy). These methodologies should leverage novel interdisciplinary frameworks for psychometric assessment to improve reliability, posing challenges and opportunities for optimizing the diverse algorithms and hyperparameters involved.

In addition to the derivation and evaluation of the metrics themselves, attention should be paid to improving the data collection, preprocessing, and standardization strategies before constructing the metrics, which will generally improve validity but may occasionally lead to trade-offs between different factors of reliability (e.g., removal of reliable individual differences in contaminants can lead to reduced inter-individual variability but increased inter-state stability for certain states of interest). Reducing head motion and increasing scan duration generally improves reliability while states of eyes can have distinct impacts on different forms of variability of resting-state fMRI measurements. In preprocessing and standardization, better denoising and registration algorithms would be helpful (see Cornerstone 4).

Cornerstone 3: explore valid factors that are sensitive to cross-sectional and longitudinal variations of interest. Since it is particularly difficult to determine validity in behavioral and social sciences, various indirect forms of validity have been proposed, which can also be found in neuroscience [7]. For example, criterion validity is the agreement with the gold standard, predictive validity is the ability to predict (e.g., predict behavioral scores with brain metrics), and face validity is the agreement with common sense. Rather than diving into definitions of various forms of validity and attempting to tease out their implications in the context of TBC research, we will analyze the exploration of valid biological factors relevant to TBC research from two aspects. The first will discuss ways to bridge the gap between reliability and validity. The second will address the specific requirements of TBC for validity and how TBC can contribute to valid findings in neuroscience.

First, bridging the gap between reliability and validity is desirable for various translational applications. Recent advances have led to an emphasis on the compromises that sometimes exist between reliability and validity [13]. For instance, head motion may be associated with certain individual traits, and thus excluding samples with severe head motion can improve reliability but may reduce external validity (i.e., the generalizability). In fMRI, restricting head motion may lead to atypical states of functional activity and reduce validity. It would be better not to overly restrict participants' head motion, but to introduce prospective motion correction methods, tracking motion during acquisition to maximize the quality of raw data. Head motion is also a typical factor that varies with age, and such factors would be significant confounds in brain charts and should therefore be carefully eliminated (see Cornerstone 4). However, the observed trade-offs between reliability and validity are in many cases due to the neglect of other forms of variability beyond inter-individual variability. As mentioned in Cornerstone 2, within-individual and between-group comparisons can also be performed to evaluate variability between states such as between behavioral traits. In this way, what really exists is a trade-off between different forms of variability rather than a fundamental trade-off between reliability and validity. We can always construct tailored forms of variability according to the specific context of TBC research to narrow the gap between reliability and validity, and increase the potential for extracting biological factors associated with the variations of interest (e.g., variations in genetic and environmental factors).

Second, there is an upward spiral between TBC research and the exploration of valid factors. Going beyond simple indices for reliability evaluation, valid factors can be identified by combining cross-sectional (both inter-individual and between-group) and longitudinal studies to comprehensively examine the sensitivity of reliable metrics (e.g., with predictive models for brain-wide associations). In this framework, we can conduct and synthesize multiple studies to construct valid metrics for TBC with corresponding application needs. Furthermore, we can derive centiles of individuals with constructed brain charts, and investigate the relationships between longitudinal centile score changes and variations of interest to assess, screen, and optimize metrics for TBC. This means that the research on the validity of TBC will not be one-size-fits-all and will require continuous iteration to improve research and application practices. Finally, TBC integrate the most basic and robust development-related variations, offering novel opportunities to harness differences in age and explore dynamic brain-behavior relationships. Specifically, TBC can serve as "microscopes" to precisely quantify relative scores of metrics for individuals by interpreting and removing age-related population trends, allowing more subtle effects to be observed, leading to more valid cross-sectional and longitudinal brain-wide association studies (BWAS), and promoting basic and translational neuroscience.

Cornerstone 4: eliminate confounds caused by suboptimal sampling and irrelevant changes over the lifespan. Even with reliable and valid data collected from representative samples, further investigation and elimination of multiple sources of confounds before or during modeling can still substantially improve the quality of the constructed brain charts (see Table 1 for an overview of major confounds in TBC research). Here, we divide the confounds into two types. One is related to suboptimal sampling, which underscores the importance of optimizing the sampling design, calibrating image acquisition conditions, and harmonizing random effects. The other is largely related to irrelevant changes over the lifespan, which confounds the detected age effects in TBC and can often be mitigated by careful refinement of the data processing protocols.

For the collected samples, the representativeness of the samples should be reassessed after quality control, and sub-selecting or supplementing samples may be considered if necessary. Moreover, cohort effects can lead to tricky confounds. For cross-sectional or accelerated longitudinal designs, cohort effects refer to the differences in birth years of individuals, leading to systematic differences in their growth environments and experiences, which can be evaluated and modeled (and thus eliminated) by leveraging longitudinal measurements. Confounds may also arise in long-term follow-up, i.e., due to scanner hardware and software changes as well as operator inconsistencies, which can be mitigated by standardizing experimental protocols and with statistical harmonization. Another confound, which is also usually mitigated with statistical harmonization, is the scanner effects, arising from differences between scanners. However, statistical harmonization is unlikely to completely remove scanner effects, but may instead remove meaningful differences between populations. In this regard, it would be helpful to coordinate hardware and software across scanners, as well as employ consistent data acquisition parameters. Furthermore, advanced quantitative MRI techniques, such as scanner calibration with "phantoms", can be introduced to correct inter-scanner variations at the level of imaging systems and algorithms.

For the data processing protocols, it is necessary to consider whether the pipelines can accommodate systematic differences between age groups. The brains of fetuses, children, and the elderly are morphologically distinct from those of young adults. The algorithms used, if not validated and optimized for different age groups, may result in systematic biases that can seriously compromise the validity of brain charts. For instance, it has been shown that developing different templates for populations of different ages and cultural backgrounds is necessary [14], which will improve the quality of brain image processing and TBC modeling. Also, given the differences in head motion and various physiological variables across ages [15], the algorithms used to reduce these confounds must be systematically evaluated with samples over the lifespan.

A final remark is that brain charts provide a unique opportunity to parse controversial confounds from a lifespan perspective. One example is whether individual differences in intracranial volume or whole brain volume should be standardized. Studies have

Table 1
Major confounds in TBC research.

Confounds	Descriptions	Elimination/mitigation approaches
Non-representative samples	Even if the raw samples are representative, the samples may be non-representative after quality control	Sub-selecting or supplementing samples
Cohort effects	Differences between age groups related to unique experience/exposure, but not age	Evaluating and modeling with longitudinal measurements
Inconsistent image acquisition conditions	Changes in scanner hardware and software, differences in instructions of operators and placement of individuals, etc.	Standardizing experimental protocols, calibrating images across dates and scanners, and harmonizing pooled datasets
Pipelines not generalizable to different ages	For instance, templates based on specific age groups, and the related denoising, segmentation, and registration algorithms	Optimizing validity of pipelines across ages and populations
Head motion	Head motion varies significantly across age groups, thus distorting the detected age effects	Mitigating head motion effects with prospective and retrospective motion correction methods
Physiological variables	Changes in variables such as breathing rate/depth, heart rate, blood pressure, and arterial carbon dioxide concentration over the lifespan	Measuring various physiological variables and developing algorithms to effectively compensate the effects of these variables

shown allometry in brain development, and thus linearly adjusting brain structure metrics may impair the validity. A better approach would be to model the joint distribution of whole brain volume and a specific structural metric with brain chart models, so as to better understand the dynamic relationships between whole brain volume and various structural metrics.

Cornerstone 5: estimate data distributions and derive centiles with flexible yet intrinsically interpretable models. We prioritize the intrinsic interpretability of the models used to estimate data distributions and derive centiles over other requirements, such as flexibility, to ensure that the models allow for accurate troubleshooting and refinement, thereby improving their external validity. Distributional regression approaches, represented by GAMLSS, are intrinsically interpretable and highly flexible in capturing nonlinear trajectories and estimating higher-order statistics of data distributions with centile curves, as demonstrated in Fig. 1a. GAMLSS has been utilized in an early TBC effort to chart morphological development from 6 to 85 years old (https://github.com/zuoxinian/CCS), and is also leveraged by LBCC to uncover the neurodevelopmental milestones across the human lifespan (0–100 years old) [1]. The flexibility of the distributional regression framework allows it to combine advances in statistical learning, thereby better accommodating sophisticated data structures and random effects (e.g., scanner effects) with good interpretability [16].

For future TBC modeling research, there are at least three important aspects worth considering. The first is to improve the use of longitudinal measurements. Although there are over 20,000 longitudinal measurements in the dataset, the LBCC team [1] ignored these within-individual dependencies to avoid convergence issues. However, these within-individual dependencies can provide unique information that can be used to mitigate cohort effects. Therefore, even when the total sample size is very large and longitudinal measurements account for only a small proportion, it is still relevant to fully exploit longitudinal dependencies. The second is to scale towards multivariate distributions, which is an inevitable direction for future TBC research to reveal multivariate characteristics of normative brain development by simultaneously accounting for multiple predictor and response variables. To avoid the curse of dimensionality, it may be necessary to impose additional structural assumptions for the multivariate distributions. The third consideration is to develop Bayesian methods for specific populations with moderate sample sizes to reduce the sampling cost of TBC. The priors could be brain charts derived from very large sample sizes, or previously constructed brain charts for the same specific population (to update the charts). The methods, although similar to estimating statistical offsets of brain charts to attenuate scanner effects for a new study (see Ref. [1] for a more detailed discussion), allow the charts to be fundamentally updated and are therefore different in spirit. The key issues to be investigated are the construction of a parameter to determine the proportion of new data contributing to the updated brain chart, and the algorithms to achieve optimal hyperparameters under specific criteria.

With brain charts constructed, one can assess the predictive and longitudinal validity of the charts by evaluating within-individual longitudinal variability of centile scores. Nevertheless, it should be noted that high variability of centile scores within healthy individuals does not necessarily indicate issues with data quality, reliability, or modeling approaches, but may also be with the validity of the metric in terms of clinical applications. As suggested in Cornerstone 3, by linking variability in centile scores with genetic and environmental factors as well as behavioral traits, we will gain new perspectives to explore associations that were previously obscured by age effects and thus usher in a paradigm shift in BWAS.

Cornerstone 6: establish a whole chain of open platforms to facilitate interdisciplinary research and translation. Open and easy-to-use platforms can promote collaborations among multiple parties to facilitate TBC research. In this regard, the LBCC team developed an interactive open resource (http://www.brainchart.io) to visualize and update brain charts, providing a paradigm for such efforts. One can expect that even if TBC have been constructed, making them easily accessible and understandable to clinicians and other users remains a challenge. To this end, online platforms will help in understanding, using, and disseminating TBC resources. Here, the point to emphasize is that TBC research entails collaborations from the very beginning. For instance, innovative methodological research will support TBC in terms of optimizing data acquisition and processing, enhancing the reliability and validity of extracted metrics, identifying and removing confounds, and improving statistical models. Interdisciplinary research with physics, computer science, and statistics, as well as the open science practices in sharing data and codes, will accelerate this process, leading to an urgent need to develop a whole chain of open platforms for interdisciplinary research and translation.

The collection of high-quality and representative data according to the sampling strategies especially requires extensive collaborations. To achieve this, in addition to the optimization of sampling strategies, it is essential to obtain policy and funding support and, most importantly, to gain public understanding and trust in data acquisition, as well as a wide range of willingness to participate in the project. For example, neuroimaging data acquisition on fetuses, infants, and toddlers remains sparse due to safety concerns, which not only hinders the community from assessing and improving the reliability and validity of methodologies for this crit-

ical period of neurodevelopment, but also poses a significant challenge to acquiring large representative samples in these rapidly developing age groups. In-depth cooperation among relevant parties and brain science popularization are necessary foundations to overcome the above challenges.

Overall, we anticipate a whole chain of open platforms based on inclusive protocols, consisting of data collection, sharing, and management systems (for Cornerstone 1), server clusters for both exploratory and standardized data processing (for Cornerstones 2 to 4), pipelines for building and updating models (for Cornerstones 4 and 5), and modules for deriving centile scores of individual brain metrics and generating personalized reports (translating TBC into applications), among other components. In this chain for TBC research, collecting and sharing data is the prerequisite, refining data processing algorithms and statistical modeling methods is the foundation, and translational research and validation is the key. With the platforms established and continuously upgraded [1,4,17], TBC research can be accelerated, and translational applications of TBC can further facilitate the collection of representative samples, foster developmental population neuroscience, and lead to better and more powerful TBC.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the STI 2030 – the major project of the Brain Science and Brain-Inspired Intelligence Technology (2021ZD0200500). We thank the Research Program on Discipline Direction Prediction and Technology Roadmap of China Association for Science and Technology for bibliometric resources, and the Chinese Data-sharing Warehouse for In-vivo Imaging Brain at National Basic Science Data Center for informatics resources. We thank Dr. Yin-Shan Wang for helping to visualize the brain charts in Fig. 1a. This work is the blueprint results of the research project, namely the Chinese Color Nest Project (http://deepneuro.bnu.edu.cn/?p=163).

References

[1] Bethlehem RAI, Seidlitz J, White SR, et al. Brain charts for the human lifespan. Nature 2022;604:525–33.
[2] Paus T. Population neuroscience: why and how. Hum Brain Mapp 2010;31:891–903.
[3] World Health Organization. Optimizing brain health across the life course: WHO position paper, 2022. https://www.who.int/publications/i/item/9789240054561.
[4] Zuo XN, He Y, Su X, et al. Developmental population neuroscience: emerging from ICHBD. Sci Bull 2018;63:331–2.
[5] Falk EB, Hyde LW, Mitchell C, et al. What is a representative brain? Neuroscience meets population science. Proc Natl Acad Sci USA 2013;110:17615–22.
[6] Xing XX, Zuo XN. The anatomy of reliability: a must read for future human brain mapping. Sci Bull 2018;63:1606–7.
[7] Zuo XN, Xu T, Milham MP. Harnessing reliability for neuroscience research. Nat Hum Behav 2019;3:768–71.
[8] Milham MP, Vogelstein J, Xu T. Removing the reliability bottleneck in functional magnetic resonance imaging research to achieve clinical utility. JAMA Psychiat 2021;78:587–8.
[9] Bolt T, Nomi JS, Bzdok D, et al. A parsimonious description of global functional brain organization in three spatiotemporal patterns. Nat Neurosci 2022;25:1093–103.
[10] Huntenburg JM, Bazin PL, Margulies DS. Large-scale gradients in human cortical organization. Trends Cogn Sci 2018;22:21–31.
[11] Sydnor VJ, Larsen B, Bassett DS, et al. Neurodevelopment of the association cortices: patterns, mechanisms, and implications for psychopathology. Neuron 2021;109:2820–46.
[12] Keller AS, Sydnor VJ, Pines A, et al. Hierarchical functional system development supports executive function. Trends Cogn Sci 2023;27:160–74.
[13] Noble S, Scheinost D, Constable RT. A guide to the measurement and interpretation of fMRI test-retest reliability. Curr Opin Behav Sci 2021;40:27–32.
[14] Dong HM, Castellanos FX, Yang N, et al. Charting brain growth in tandem with brain templates at school age. Sci Bull 2020;65:1924–34.
[15] Zuo XN, He Y, Betzel RF, et al. Human connectomics across the life span. Trends Cogn Sci 2017;21:32–45.
[16] Marmolejo-Ramos F, Tejo M, Brabec M, et al. Distributional regression modeling via generalized additive models for location, scale, and shape: an overview through a data set from learning analytics. Wiley Interdiscip Rev: Data Min Knowl Discov 2023;13:e1479.
[17] Xing XX, Xu T, Jiang C, et al. Connectome Computation System: 2015–2021 updates. Sci Bull 2022;67:448–51.

Zi-Xuan Zhou is a postdoctoral researcher at State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. He received his Ph.D. degree (Communication and Information Systems) from University of Chinese Academy of Sciences in 2022. His research interest includes combining physics, network science, and novel information system models to derive brain imaging representations from an interdisciplinary perspective, and building a methodological framework for distilling knowledge from population neuroimaging data.

Xi-Nian Zuo is a full professor at State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. He leads the Chinese Color Nest Project (CCNP) and the Consortium for Reliability and Reproducibility (CoRR). He devotes his research to reveal the cortical gradient shifts from childhood to adolescence as well as the related differences in brain growth charts between China and the United States, and proposes developmental population neuroscience.