Reliability characterization of MRI measurements for analyses of brain networks on a single human

Céline Provins[1], Hélène Lajous[1,2], Elodie Savary[1], Eleonora Fornari[1,2], Benedetta Franceschiello[3,1,4], Yasser Alemán-Gómez[1], William H. Thompson[5, 6], Ileana Jelescu[1], Patric Hagmann[1], Oscar Esteban[1, @]

@ Correspondence:
phd@oscaresteban.es (✉);
@oesteban (🐦); @oesteban (🐙)

Network-based approaches are widely adopted to model functional and structural 'connectivity' of the living brain, extracted noninvasively with magnetic resonance imaging (MRI). However, these analyses —on functional and structural networks— render unreliable at the finer temporal, spatial, and brain-parcellation scales. Consequently, the clinical translation of these analyses has yet to materialize meaningfully, and interpretation of the skyrocketing production of scientific literature requires caution. We will characterize relevant sources of variability and assess the reliability of structural and functional networks extracted from MRI with the repeated acquisition of a single, healthy individual, whom we regard as the 'Human Connectome Phantom'. Two comprehensive MRI protocols will be executed across three different devices (48, 12, and 12 sessions, respectively) while recording a wealth of physiological signals to help model corresponding spurious effects on brain networks. To maximize reuse, e.g., as a benchmark reference, a baseline for machine learning models, or a source of prior knowledge, we will openly share all data and their derivatives. By systematically assessing spurious sources of variability throughout the neuroimaging workflow, we will deliver reliability margins of brain networks that inform future research and contribute to the standardization of 'connectivity measurement'.

Brain networks extracted from magnetic resonance imaging (MRI) offer a noninvasive approach to understanding spatially-distributed correspondences of its structure, function, or both. Intending to enable the identification of useful biomarkers to assess the neurotypical trajectory of the brain, these applications (so-called *MRI connectomics*; ***Hagmann et al., 2010***) first looked at network patterns emerging from structural features of the brain and their links to disease (***Griffa et al., 2013***). Indeed, diffusion MRI (dMRI) allows for brain tractography, which tracks the pathways of major fiber bundles within the brain to interpret them as links between brain regions (structural connectivity, SC; ***Hagmann, 2005***). Functional MRI (fMRI) enables the extraction of neural activity correlations between regions (functional connectivity, FC; ***Damoiseaux et al., 2006***) and permits the investigation of temporal patterns (that is, brain function dynamics; e.g., ***Chang and Glover, 2010***; ***Sadaghiani et al., 2015***; ***Griffa et al., 2017***; ***Preti et al., 2017***). Concurrently, interest has steered toward convergent analyses of SC and FC as a better-founded strategy for understanding the brain (***Bargmann and Marder, 2013***). Lately, connectivity analyses have organically generalized to several brain data sources (e.g., genomics; ***Arnatkeviciute et al., 2021***) that aggregate into a new discipline of 'network neuroscience' (***Bassett and Sporns, 2017***).

SC and FC extracted from MRI have proven sufficient levels of reliability to discriminate between individuals (***Mueller et al., 2013***; ***Finn et al., 2015***, ***2017***), and such reliability is reportedly stable from months to years (***Gratton et al., 2018***; ***Horien et al., 2019***). Moreover, MRI connectomics has displayed strong associations between SC and FC networks at the larger scales of analysis (***Honey et al., 2009***; ***Hermundstad et al., 2013***; ***Rosenthal et al., 2018***; ***Vázquez-Rodríguez et al., 2019***). As reviewed by ***Bijsterbosch et al. (2020)***, imaging and analytical limitations wane such associations out at the finer scales of analysis. Pushing the spatial resolution of dMRI and fMRI to the proximity of 1.0×1.0×1.0 [mm^3] can be done at the expense of the signal-to-noise ratio. Likewise, improving the temporal resolution of fMRI in the sub-second range competes with spatial resolution. In the case of dMRI, increasing the angular resolution or the sensitivity to diffusivity is not free of challenges either (e.g., scan length, device capability, artifacts). Further limitations arise in downstream modeling imposed by the spatial/areal extent of the network nodes (or brain 'units'; ***Bijsterbosch et al., 2020***), which determines the sampling density of edges in the resulting connectivity matrices. The definition of brain units is most often determined by the 'regional' resolution of the parcellation/atlas of reference, a methodological choice (***Ciric et al., 2022***) which has been demonstrated to introduce large variability, e.g., in FC (***Doucet et al., 2019***). These limitations have hindered the discovery of distinct SC and FC associations and links to clinical conditions and behavior at the finer scales of analysis (***Rodrigue et al., 2021***). The most plausible explanation for these limitations is that measurements obtained with MRI are indirect, spatiotemporally uncertain, and confounded by other signal sources (***Esteban et al., 2019***). Particularly in the case of

Affiliations: [1]Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [2]CIBM Center for Biomedical Imaging, Switzerland; [3]Institute of Systems Engineering, School of Engineering, HES-SO Valais-Wallis, Sion, Switzerland; [4]The Sense Innovation and Research Center, Lausanne and Sion, Switzerland; [5]Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden; [6]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.

fMRI, a definition of what noise is and what is the signal is itself controverted, partly because it remains unclear how spontaneous neural firing should be modeled (***Uddin, 2020***). If not accounted for, noise and artifacts accumulate downstream of the research workflow, which may also build up further spurious variability (e.g., numerical errors). Consequently, SC and FC contain large false positive and negative ratios in both the structural (***Zalesky et al., 2016***; ***Maier-Hein et al., 2017***) and the functional (***Power et al., 2012***; ***Chen et al., 2020***) networks. Therefore, it is critical to characterize the validity and reliability of MRI data (***Zuo et al., 2019***; ***Milham et al., 2021***), as well as those of SC (***Buchanan et al., 2014***; ***Sarwar et al., 2021***; ***Cai et al., 2021***) and FC (***Noble et al., 2017***, ***2019***, ***2020***; ***Cosgrove et al., 2022***), before biomarkers may be defined and derived from these analytical approaches. Improving the reliability of SC and FC is critical to current network neuroscience approaches and emerging frameworks, such as edge-based functional coactivation (***Zamani Esfahlani et al., 2020***).

Here, we characterize the reliability of SC and FC by adopting a 'dense imaging' approach (***Naselaris et al., 2021***) that involves the repeated acquisition (72 sessions in total) of MRI data on a single human. Data acquisition will be carried out in a short time span to minimize within-subject variations due to aging and to limit the possibility of unexpected life events that may induce changes in the brain (e.g., COVID-19 infection; ***Douaud et al., 2022***). Thus, the single-subject approach is akin to a living phantom, thereby excluding between-subjects variability sources from our analyses. The phantom is chosen to be an individual at a plateau of brain development (i.e., 25–55 years old) to further minimize drifts in SC and FC patterns. We will further minimize transient structural (***Dubol et al., 2021***) and connectivity (***Greenwell et al., 2021***) changes reportedly correlated with the menstrual cycle by choosing a biological male individual. Additional reasons to select a male include the non-negligible differential burden the participant assumes, emerging from closely tracking the menstrual cycle and related factors (e.g., hemoglobin levels and/or abundance of discharges, pain, etc.). Moreover, accepting the publication of their self-reported information about their menstrual cycle would be set as an inclusion criterion to ensure all data can be openly shared. Alternatively, the redaction of such information before sharing would severely degrade the reuse potential of the dataset. Data will be acquired across three 3.0 T scanners (see Table 1). One of the scanners, with part-time research dedication, is suitable for acquiring a wealth of sophisticated MRI sequences with simultaneous recording of several physiological signals and eye tracking, which enable the investigation of various variability sources (***Chen et al., 2020***). In addition to this 'reliability' subset of data (see Figure 1), additional data will be acquired across the three available scanners to (i) characterize the consistency of SC and FC between acquisition devices that are proximal in settings (e.g., all three scanners are 3.0 T systems, Siemens Healthineers); and (ii) generalize the reliability characterization on to more standard, clinical settings and MR acquisition protocols. Rather than seeking a characterization of reliability that generalizes across subjects, the study design is single-subject aiming to serve as preliminary data in the estimation of the number of repetitions per subject that ensure pre-specified levels of statistical power, thereby optimizing the balance between their commitment as healthy volunteers and the research output.

The design of the study (Figure 1 provides a graphical overview) observes the latest recommendations for neuroimaging studies (***Niso et al., 2022***; ***Rosenberg and Finn, 2022***) regarding sample size, pre-registration of the analyses before acquiring the data, quality assessment and control, and transparency (e.g., comprehensive reporting of methods and results of all the analyses, as well as openly sharing data and software). The new, open dataset will evenly cover dMRI and resting-state fMRI (RSfMRI), in contrast to precedent efforts focused on either SC or FC. We overview available alternative datasets in Table 2 and discuss how this undertaking complements the existing wealth of data. For a broader context in data-sharing of MR connectomics and its ethical support, refer to ***Laird*** (***2021***). Correspondingly, preprocessed data derivatives will also be openly released to alleviate the carbon footprint of reprocessing the original dataset and to ease the full replication of this study. The study design also aims to minimize analytical variability sources, in particular arising from the neuroimaging pipeline implementation (***Li et al., 2021***), by using 'standardized' preprocessing tools from our *NiPreps* (*NeuroImaging PREProcessing toolS*; www.nipreps.org) ecosystem, such as *fMRIPrep* (***Esteban et al., 2019***). Moreover, this registered report includes the plan for confirmatory analyses in complete detail as an additional measure protecting against methodological variability (***Botvinik-Nezer et al., 2020***; ***Sarwar et al., 2021***).

We will model the reliability of edges in SC networks with a Bayesian analysis framework (***Hinne et al., 2013***). Such a framework will generate edge-wise posterior distributions on which we will determine the most reliable

Table 1. A 'dense imaging' approach across three different systems of a single vendor. A total of 72 sessions will be carried out across three different scanners, however all built by Siemens Healthineers and of 3.0 T magnetic field strength (Figure 1). One scanner will repeat a 'reliability' protocol with 36 sessions. In addition, a 'standard' protocol will be executed 12 times in each of three scanners (including the scanner of the 'reliability' protocol), which will be utilized in the study to assess between-scanner variability. In order to keep between-scanner variability limited, only single-vendor, 3.0 T scanners will be considered.

Identifier	Manufacturer	Model	Software Version	Field	Clinical use hours	Sessions	MRI Protocol
BH07060	Siemens	Magnetom Prisma^{Fit}	XA30	3T	7h-13h	36	Reliability
						12	Standard
BH07067	Siemens	Magnetom Vida^{Fit}	XA20	3T	7h-22h	12	Standard
BH07035	Siemens	Magnetom Vida	XA30	3T	7h-22h	12	Standard

edges and characterize credible margins of the presence or absence of connections. The resource will also enable exceptional opportunities for researchers concerned with the confounding effects of physiology (***Chang et al., 2009***) owing to the collection of fMRI data for cardiovascular reactivity mapping (***Bright et al., 2020***; ***Pinto et al., 2021***). Such mapping is critical for isolating BOLD (blood-oxygen-level-dependent, the most widely used fMRI contrast) components from other sources of confounds to characterize the hemodynamic activity and inform downstream denoising and modeling. Further, separating and quantifying the variability sourcing from spontaneous neural firing (***Uddin, 2020***) and non-repeatable activation elicited by behavior (***Rosenberg et al., 2020***; ***Song and Rosenberg, 2021***) are of paramount interest in the case of fMRI. Indeed, ***Finn and Rosenberg*** (***2021***) have raised awareness that optimizing FC analyses for reliability (e.g., 'fingerprinting'; ***Finn et al., 2015***) effectively sets bounds to the prediction of behavior. To boost behavior prediction in the long term, we set out to investigate the reliability of the FC extraction process and outcomes. Therefore, we will minimize transient behavioral effects during the RSfMRI experiments by reproducing a naturalistic movie (***Finn and Bandettini, 2021***) in all sessions. Naturalistic movie watching has also been confirmed to reduce head-motion and to help keep higher levels of arousal as compared to more conventional RSfMRI with eyes-open and a fixation point (***Vanderwal et al., 2017***). We will further minimize physiological sources of variance by leveraging the cardiovascular reactivity mapping and other physiological information. Finally, we propose a principal components analysis (***Bari et al., 2019***) to determine the core structure of FC across sessions. We hypothesize that behavioral variance will decay over the project span, allowing us to characterize residual variability sources. Further, we will also comparatively investigate the estimated head motion patterns during dMRI and fMRI to isolate biases derived from the algorithmic implementation, building on the work of ***Bolton et al.*** (***2020***). Finally, we will use multivariate analysis of variance (MANOVA) to investigate 'scanner effects' on SC and FC with data collected on the clinical devices. In sum, this study will characterize the reliability of SC and FC networks extracted from d/fMRI, respectively, and determine structured biases between data collected from different scanners with otherwise homogeneous settings. Such a characterization is critical to bridging the gap separating network analysis from more reliable network neuroscience and, in the longer term, the definition of robust biomarkers for clinical application.

Methods

Data

Participant and ethical review. The participant (author O.E.) is a left-handed white Hispanic male, aged 40 at the onset of the study. The participant has a history of anxiety disorder and sporadic cluster headaches, but is otherwise generally healthy. The participant has mild myopia in both eyes (-0.5 right, -0.75 left). The study protocol was submitted to the Swiss BASEC (Business Administration System for Ethics Committees) and received positive approval from the corresponding local Ethics Committee on May 30, 2022 (project 2022-00360).

Standard Operating Procedures (SOPs). All the data collection and methodological implementation details are comprehensively specified in the SOPs document (for a description of SOPs and their relevance in more detail, refer to ***Etzel et al., 2022***). The SOPs are developed with Markdown (a popular markup language for documentation) and leverage Git for version control and to protect sensitive information (e.g., usernames and passwords) by a design that keeps secrets separate (and inaccessible) from the public repository. The SOPs are available at axonlab.org/hcph-sops.

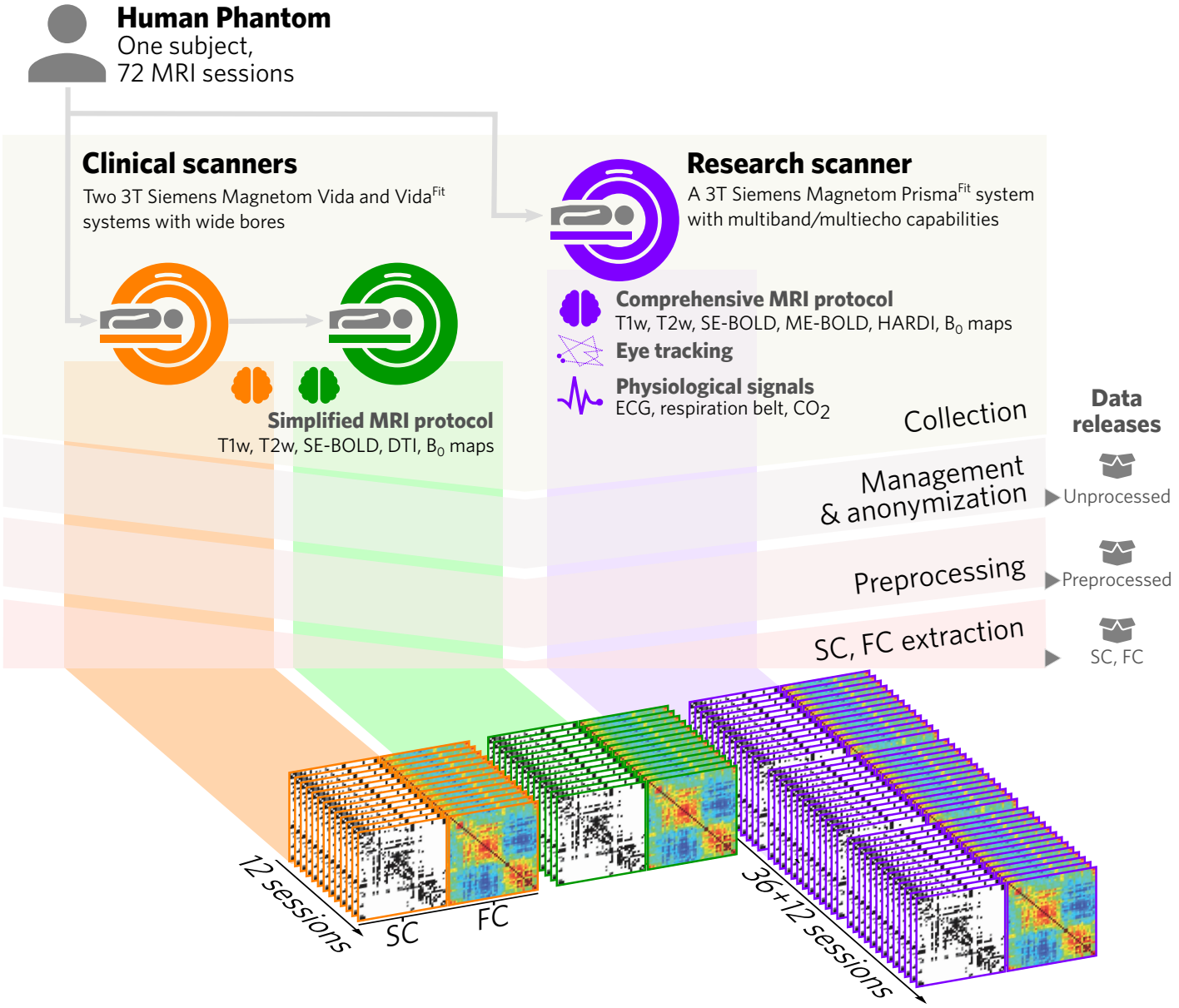

Figure 1. Characterizing the reliability of edges in connectivity matrices extracted from a single human phantom. Leveraging three 3.0 T scanners available at Lausanne University Hospital, a single human subject will undergo 72 MRI sessions to evaluate the variability of structural and functional connectivity (SC, FC, respectively). One scanner will accommodate a 'reliability' MR acquisition protocol, including a number of physiological signals and eye-tracking information that will allow for sophisticated denoising and a more reliable identification of signal with non-neural origins. Data will then undergo standardized and pre-registered management, preprocessing (with ***MRIQC***, ***Esteban et al.*** (***2017***), ***fMRIPrep***, ***Esteban et al.*** (***2019***), and ***dMRIPrep***, ***Joseph et al.*** (***2021***)), and connectivity extraction. Each of these steps will be culminated with corresponding data releases using the ***BIDS*** (Brain Imaging Data Structure; ***Gorgolewski et al., 2016***) specifications to maximize reusability. A total of 36 pairs of SC and FC matrices will then be fed into analyses of variance to identify and characterize the reliability of edges. An additional set of 36 sessions (12 per scanner) of an adapted MRI protocol suitable for clinical practice will allow to investigate scanner effects and to estimate generalization of the results on data from other scanners. Abbreviations are spelled out at the bottom of Table 4.

Table 2. 'The HCPh dataset' will fill a number of gaps within currently available 'dense imaging' datasets. A number of datasets have focused on acquiring large amounts of MRI data on a limited number of individuals. However, these datasets have typically been acquired with the advance of our understanding of the human brain as a goal. Conversely, the HCPh is designed to fully understand and characterize the reliability of functional and structural connectivity analyses, from the acquisition of MRI data to the extraction of network information. Therefore, the HCPh focuses on the assessment and optimization of our research instruments, henceforth contributing to making currently existing datasets more valuable and to increasing the trustworthiness of neuroimaging results. Several openly accessible datasets are not included in this table, because they only contain fMRI (e.g., ***Noble et al., 2017***; ***Choe et al., 2015***; ***Gordon et al., 2017***; ***Seeliger et al., 2019***) or dMRI (e.g., ***Cai et al., 2021***; ***Froeling et al., 2016***; ***Seider et al., 2022***) data, and therefore they cannot be used in multi-modal (SC and FC) analyses.

Name	N	S	Anat.	dMRI	RSfMRI	Other	Comparative overview
StudyForrest	20	3	1 T1w 1 T2w	1	n/a (only task-based fMRI)	Time-of-flight MRI (angiography), Susceptibility-weighted MRI	The 'StudyForrest' (***Hanke et al., 2014***) is probably one of the earliest datasets acquiring long sessions of few subjects (dense sampling). The dataset is characterized by the fMRI acquisition during prolonged stimulation with an auditory feature film ('Forrest Gump'). MRI data were acquired on a single 7.0 T device, along with completing structural MRI and conventional diffusion tensor imaging (DTI). With respect to the proposed dataset, the StudyForrest does not provide RSfMRI and there are no repeated scans for structural and dMRI data.
MyConnectome	1	110	10 T1w 11 T2w	15	84 × 10 min	Respiration recordings Genomic information	'MyConnectome' (***Poldrack et al., 2016***) stands out as an antecedent to the proposed dataset. Indeed, ***MyConnectome*** first proposed the human phantom approach with the goal of integrating several imaging and non-imaging derived '-omes' for a single individual in a single analysis. Whereas ***MyConnectome*** is designed to address neuroscientific questions, the HCPh has the intent of contributing to the reliability of MR Connectomics and corresponding methodological instruments by investigating the neuroimaging pipeline. Comparatively, ***MyConnectome*** provides only 15 repetitions of single-shell dMRI suitable for conventional DTI (as opposed to the HCPh, which will allow more sophisticated modeling sensitive to the varying diffusivity of tissue); fMRI was acquired with single-echo fMRI, which limits the opportunities for further investigation of physiological confound sources; acquisition was carried out with only a respiration belt recordings, and there was interest in several fMRI tasks that were also included in the protocol.
HBN-SSI	13	14	14 T1w 1 T2w	13	13 × 10 min 2 × 12 × 10 min (naturalistic)	Actimetry sensor, relaxometry MRI, multi-echo MPRAGE, magnetization transfer MRI	The Healthy Brain Network Serial Scanning Initiative ('HBN-SSI'; ***O'Connor et al., 2017***) was the first openly available dataset acquired to evaluate inter-individual differences and reliability across scan conditions and sessions of RSfMRI. Repeated dMRI acquisitions are comparable to those planned under the HCPh for the 'clinical' protocol, providing moderately high-angular-resolution and multi-shell data. The HBN-SSI is an ideal dataset to investigate the generalization of the results of the proposed study to new individuals, as well as to distant scanning settings. Limitations to such future lines will however certainly be encountered, e.g., the absence of physiological recordings during the HBN-SSI scanning, and the technical distance of the 1.5 T scanner of HBN-SSI (Siemens Avanto, 32-channel head coil) with respect to our 3.0 T scanners (Table 1).
IBC	12	5	1 T1w 1 T2w	1	n/a (only task-based fMRI)	T2 FLAIR	The Individual Brain Charting ('IBC'; ***Pinho et al., 2018***) is very oriented towards task-based fMRI data. As such, the dataset does not include RSfMRI, and the single dMRI data acquired per subject is a conventional DTI.
NSD	8	30–40	6-10 T1w 2-3 T2w	4	20×5 min or 36×5 min depending on the subjects	Eye-tracking	The Natural Scenes Dataset ('NSD'; ***Allen et al., 2022***) is a recent resource investigating the reliability of high-resolution fMRI in understanding brain function, specifically the vision system. With that scope, the NSD provides a limited set of high-angular-resolution and multi-shell dMRI data acquired at 3.0 T. The fMRI data were acquired on a 7.0 T device. Such high-field scanners are much scarcer than 3.0 T devices and are rarely used in clinical settings. The main focus for the fMRI acquisition was on visual localizers and the NSD task, thus marginal (although present) on RSfMRI. Finally, the NSD collected eye-tracking, but no other physiological recordings as compared to the HCPh.
Day2Day	6	40–50	40–50 T1w & T2w	40–50	40–50	T2w of hippocampus MRS (MR spectroscopy) Covariates	The ***Day2Day*** dataset (***Filevich et al., 2017***) had eight healthy participants, of which only six concluded all the planned sessions, who underwent brain scanning 40–50 times over the course of six months. Comparatively, ***Day2Day*** provides single-shell dMRI suitable for conventional DTI (as opposed to the HCPh, which will allow high-angular resolution modelling). Sessions including RSfMRI had a limited length 5 min 8 s, way below broad recommendations of 20 min for reliability (***Noble et al., 2017***). Participants were requested to keep their eyes closed during all image acquisitions. The study made a remarkable effort collecting covariates, which included weather, diet, sleep, smoking, mood, menstrual cycle, etc.

Table columns: N is the number of subjects; S is the number of sessions per subject; 'Anat.' indicates the number of *anatomical* (T_1-weighted –T1w– or T_2-weighted –T2w–) imaging assets per subject, across all sessions; 'dMRI' and 'RSfMRI' indicate the number of diffusion MRI and resting-state functional MRI assets per subject across all sessions, respectively; 'Other' indicates additional data included in the dataset which are of particular interest comparatively to the HCPh.

Table 3. Session schedule for the 'reliability' and 'standard' imaging protocols. Two MRI protocols have been designed with consideration to data 'harmonization' between the devices, based on their technical capabilities. For instance, one RSfMRI run will be acquired with the same parameters across the three devices. Protocols are designed for a total session duration under 90 min. The specific imaging parameters corresponding to all the relevant MR schemes, as well as abbreviations are found in Table 4.

Reliability Imaging Protocol (36 sessions × one scanner)	mm:ss	**Standard Imaging Protocol (12 sessions × three scanners)**	mm:ss
Head scout	00:14	Head scout	00:14
FoV Manual positioning	01:00	FoV Manual positioning	01:00
T1w (anatomical reference)	5:41	T1w (anatomical reference)	5:41
DWI (dMRI, structural connectivity)	33:52	EPI BOLD (B_0 field mapping)	00:30
DWI (B_0 field mapping)	00:54	BOLD (single-echo RSfMRI, eyes open)	20:09
GRE (B_0 field mapping)	02:38	GRE (B_0 field mapping)	03:08
EPI BOLD (B_0 field mapping)	00:27	BOLD (multi-echo RSfMRI, eyes open, **only BH07060**)	10:09
BOLD (positive-control task fMRI)	03:07	DWI (dMRI, structural connectivity)	27:31
BOLD (multi-echo RSfMRI, naturalistic movie)	20:09	DWI (B_0 field mapping)	01:10
BOLD (breath-holding task fMRI)	6:00	T2w (anatomical reference)	8:24
T2w (anatomical reference)	5:10		
Total Acquisition Time	79:12	Total Acquisition Time	77:56

MRI protocols and session execution. We will acquire two different MRI protocols, a 'reliability' protocol for execution in the research scanner; and a 'standard' protocol. Table 3 introduces the specific MRI acquisitions, duration, and execution order of both protocols, and Table 4 expands the MRI protocol with the specific parameters and settings of acquisition. The reliability protocol is designed to characterize the variability sources and associated distribution spreads. The standard protocol is a projection of the reliability protocol into the specific capabilities and settings available in the scanning room of the full-time clinical devices (that is, more generally available settings, hence 'standard'). The standard protocol is acquired to address the generalization of our reliability characterization across scanners. MRI will be performed on a regular schedule, subject to the availability of the scanners (Table 1) and the participant. All sessions of the standard protocol (12 per device) will be acquired interspersing devices. Circadian effects will be accounted for by scheduling sessions at the same time and day of the week whenever possible. Nonetheless, the day and time of acquisition will be recorded and included in the statistical modeling as a confounding factor (***Vaisvilaite et al., 2022***). The participant will consume a fixed amount of caffeine at pre-designated times over the 8 h before each session, and will be fed. The times of prior caffeine and food intakes will be noted in every session, since both factors have effects on FC (***Poldrack et al., 2015***). The participant will minimize the consumption of substances that may introduce further variability in the analysis for the span of the acquisition, and will report if such substances (e.g., 5-HT_{1A} receptor agonists for cluster headaches) have been used. The participant will keep a sleep diary during the data collection span, reporting the duration and quality of their sleep. This sleep diary, diet, and fluid (water) intake, as well as mood tracking, will be implemented with Google Forms to be filled in by the participant at prescribed times. The order of MRI sequences in the protocols is presented in Table 3, and follows a design considering several relevant aspects: the value of the specific acquisition (e.g., while all the sessions must have a T1w image, the corresponding T2w image is placed near the end of the session in case an early stop is necessary), maintenance of the subject's arousal state (e.g., a control task requiring engagement of the participant will be executed by the end of the session, separate from the rest of fMRI scans), and synergy between runs (e.g., a breath-holding task will be executed immediately after the RSfMRI; ***Stickland et al., 2021***).

Sample size. All MRI sessions aggregated across scanners will be acquired over the span of 48 weeks or less to minimize the impact of physiological and structural changes on the longer-term timescale. Sessions where either the dMRI and/or the RSfMRI acquisitions do not meet quality-control inclusion criteria established before acquiring the data (and pre-registered within the SOPs) will be repeated after the initially-planned scanning period. The sample size of analyses varies depending on each specific analysis' setup. Analyses addressing the reliability characterization of SC (section Reliability characterization of parametric diffusion maps and structural connectivity) will have a sample size of 36 experimental units in one group, which can be doubled if the redundancy of data is exploited, as the dMRI sequences will acquire the same diffusion orientations with opposed polarities. In the reliability characterization of FC (Reliability characterization of functional connectivity), the analyses will also have a sample size of 36 experimental units. The work of ***Noble et al.*** (***2017***) presented evidence that RSfMRI sessions of 20 min (as those proposed in this report) or longer showcase relatively 'good' test-retest reliability with five or more sessions (refer to Figure 2 of ***Noble et al., 2017***). Conversely, when testing generalization onto the clinical protocol, the sample size will be n=12 units per each of the three groups (corresponding to the three scanners).

Reliability imaging protocol. A reliability protocol is defined for the BH07060 scanner (Table 1), totaling 36 sessions of ∼120 min duration each (including a preparatory overhead of 30 min per session). The protocol will be carried out with a 64-channel head coil and contains multi-shell, high-angular resolution dMRI, and resting-state BOLD-fMRI obtained by means of multi-echo echo-planar imaging (EPI) for the purpose of signal variability analyses and the extraction of connectivity matrices (Table 3). These dMRI and fMRI sequences leverage simultaneous multi-slice (SMS) imaging, an EPI acceleration technique based on parallel acquisition. In addition, anatomical MRI is included for the spatial contextualization of results within-subject and with respect to standard spaces used for group inference in neuroimaging. Finally, further MRI data will be acquired for the quality assessment and correction of artifacts. Specifically, we will collect two different alternatives for mapping B_0 field inhomogeneity (often referred to as 'fieldmap'), one 'positive control' fMRI task (unrelated to the purpose of the experiment but assistive in modeling sources of spurious variability), and a breath-holding task (for comprehensive physiological modeling). The protocol specification is described in Table 3, and the full detail of the corresponding MRI sequences in Table 4. Some experiments present visual cues, which will be back-projected onto a mirror attached to the head coil, with a total distance between the participants' eyes and the screen of 102 cm. The nearsightedness prescription of the participant is small enough to allow the execution of the acquisition protocol without visual aids. The reliability protocol will also collect a wealth of physiological information, and eye tracking traces, described below in Other recordings.

Standard imaging protocol. A standard protocol is defined for the two full-time clinical scanners (see Table 1) totaling 36 sessions (12 sessions for each device). Sessions will be collated within the collection time frame of the reliability sample. This standard protocol (fully described in Table 3) is estimated to have a duration of 68 min

plus a preparatory overhead of 15 min per session. Several factors impose limitations with respect to the reliability protocol: (i) collection will be carried out with a more generally available 32-channel head coil; (ii) the two scanners feature wider cores for clinical purposes; and (iii) neither of the scanners is equipped to execute SMS sequences. The protocol specification, described in Table 4 alongside full-detail of the corresponding MRI sequences will be openly distributed with the data release of the unprocessed dataset.

Structural connectivity: diffusion MRI. High-angular-resolution, multi-shell data uniformly covering the full sphere of the Q-space (***Caruyer et al., 2013***) will be acquired as prescribed in Table 4. These diffusion orientations will be acquired in both opposing polarities within a single run. The four possible phase-encoding (PE) readout directions (Anterior→Posterior, P→A, Left→Right, R→L) will be acquired sequentially per scanning device session, thereby yielding nine repetitions of each PE direction for the research scanner and three repetitions of each PE direction per clinical scanner, in consistent ordering.

Functional connectivity: resting-state fMRI. Protocols include one RSfMRI run per session (Table 3). The duration of this RSfMRI acquisition will be set to 20 min to ensure a high likelihood that at least 15 min of data per session will survive head-motion scrubbing in preprocessing (***Gordon et al., 2017***). The RSfMRI run of the reliability protocol will be acquired with multi-echo EPI to allow more comprehensive physiological modeling; and while showing a naturalistic image movie (***Finn and Bandettini, 2021***), without sound. The movie clip will display a time-lapse of a view over the sea in the village of Mundaka (Spain) and will not contain objects moving or appearing in the foreground (such as birds flying or people so close that their face is recognizable). The participant will not be exposed nor have access to the 'Mundaka clip' before the first session of the study. The clip will match the duration of the run and will start playing with the first valid BOLD volume (i.e., after calibration and discarded 'dummy' scans). This design with the same clip displayed in every session seeks the appearance of habituation (that is, diminishing responses with repeated stimuli) that we hypothesize will reduce the signal variability elicited by behavior, permitting more specific modeling of other variability constituents. In other words, habituation is anticipated to possibly reduce behavioral variability, rather than the target of our investigation. In the case of the standard protocol, RSfMRI will be acquired with a single-echo EPI scheme that can be executed on the clinical devices, and the participant will maintain their eyes open throughout, as the two clinical scanners lack the settings for movie projection.

Physiological modeling of fMRI: breath-holding task. A breath-holding task paradigm is included in the reliability protocol to enable a comprehensive physiological modeling with cardiovascular reactivity mapping. Cardiovascular reactivity can reliably be mapped with a capnograph in combination with a nasal cannula to measure the CO_2 levels in exhaled air (***Pinto et al., 2021***). In combination with the respiration belt measurements, we plan to model a respiratory response function (***Birn et al., 2008***) that can then be evaluated for reliability across sessions, and for regressing out the cardiovascular artifacts from the BOLD signal (***Bulte and Wartolowska, 2017***) before extracting FCs. The breath-holding task has been designed following several recommendations from the review of ***Pinto et al.*** (***2021***). The baseline breathing block will be paced at a frequency pre-established by measuring the individual's normal breathing frequency during the piloting stage of the project (***Bright and Murphy, 2013***). A period of apnea will follow immediately after expiration to maximize the reliability of the cerebrovascular response and robustness to involuntary confounding factors (e.g., involuntary Valsalva maneuver), and hence overall easier to model (***Scouten and Schwarzbauer, 2008***). The CO_2 signal from a gas analyzer (ML206, ADInstruments) will be corrected for the latency added by the length (8,4 m) of the tube connecting to the nasal cannula (***Moreton et al., 2016***). Color-coded visual cues with homogeneous brightness (to minimize the activation owing to visual response) will instruct the subject to breathe in and out at the pre-determined individual's normal frequency for 25 s ensuring the block ends with a full expiration cue, then followed by an apnea (breath hold) period of 15 s (***Bright and Murphy, 2013***). A total of five blocks will be executed, as it has been reported that this paradigm is robust with three repetitions (***Lipp et al., 2015***). An additional block will be set up at the beginning of the task, but the participant will be instructed to visualize and dismiss visual cues (i.e., normally breathe), to allow modeling of the visual response and its removal from the cardiovascular reactivity mapping as well as ensuring the gas analyzer is in a stable measurement regime. A pre-determined visual cue will inform the participant that the test block has concluded and, therefore they should engage in the breath-holding task. After the last apnea block, a normally-paced respiration block of 25 s will be appended to account for the signal lag of the gas analyzer and the tubing (5–25 s), and to allow full recovery of the normal breathing by the participant. To maximize the information that can be extracted from the task, it will be executed contiguously after and using the same settings of the multi-echo RSfMRI run, which will also be used for cardiovascular reactivity mapping without respiratory challenge (***Stickland et al., 2021***). Moreover, it has been shown that using multi-echo fMRI improves cerebrovascular response and increases its reliability across sessions (***Moia et al., 2021***). To convert from gas concentration into exhaled CO_2 pressure, the atmospheric pressure in the control room (where the gas analyzer will be placed) will be measured with a barometer before each session.

Quality assessment fMRI run: positive control task. We will acquire a brain activation ('positive control') task paradigm with well-known, robust expected responses (***Harvey et al., 2018***). Such reliable responses enable controlling for the correctness of the data management and processing pipelines at several checkpoints (***Niso et al., 2022***). We derived our implementation of the positive control task from the 'eye-movement' variant of the tasks proposed by ***Harvey et al.*** (***2018***). The task is implemented with blocks of three stimuli: visual grating pattern, finger-tapping commands (left/right), and gaze movement. The stimuli presentation order and their realization are randomized for every block and every session (e.g., the coordinates of the fixation point in the gaze-movement block or the hand of the finger-tapping block).

Other recordings. All fMRI acquisitions of the reliability protocol will be acquired while recording physiological information from a respiration belt, three electrocardiogram (ECG) leads, and a gas analyzer device (measuring the exhaled CO_2 concentration), time-locked to image acquisition. An eye-tracking system (EyeLink 1000Plus, SR Research) will record eye movements. Gaze trajectories of the right eye will be recorded using infrared light, with a sampling rate of 1000 Hz, through a mirror positioned inside the scanner bore, replacing the infrared-incompatible, standard head-coil mirror. Naked weight will be recorded at home before every session by the participant. Blood pressure (recorded with an Omron M3 Comfort device while sitting down) and body temperature will be measured before and after the scanning session. Scanning room temperature will also be recorded before and after every session. Self-reported measures (tracking sleep, mood, etc.), weight, blood pressure, scanning room temperature, and control room pressure will be logged with Google Forms.

Prior knowledge from brain atlases. One critical aspect of neuroimaging analyses is the choice of spatially-normalized anatomical references and corresponding atlases (***Ciric et al., 2022***). Standardized spatial references and corresponding atlases will be integrated into the analyses with *TemplateFlow* (***Ciric et al., 2022***). Therefore, all analyses will incorporate prior knowledge drawn from atlases defined with reference to the `MNI152NLin2009cAsym` template (***Fonov et al., 2011***). Nodes of networks formalizing SCs and FCs will be defined by projecting a probabilistic atlas called 'Dictionary of Functional Modes for brain imaging' (*DiFuMo*; ***Dadi et al., 2020***) on to the scanner's space of each d/fMRI run. All analyses will be executed at three regional scales (namely, 64, 128, and 512 'dimensions' of the atlas) to allow insights on how reliability evolves as the regional scale resolution is higher. *DiFuMo* has large coverage over the brain extent (i.e., including subcortical grey matter regions and the cerebellum); thus, it will not be combined with other specialized atlases. For FC analyses, the atlas will be resampled onto the `fsaverage` surface atlas to enable mixed vertex-wise (for samples on the cortical surfaces) and voxel-wise (for samples in subcortical and cerebellar regions) analysis. For analyses looking into the reliability of dMRI-derived scalar maps, we will use the four scales in the atlas of white-matter bundles included within the multi-scale probabilistic atlas of the human connectome by ***Alemán-Gómez et al.*** (***2022***). The scale 1 of the atlas (the coarsest) will be projected into the anatomical T1w space, and from that into each individual run through spatial transformations estimated with *dMRIPrep*.

MRI data acquisition, quality control, and preprocessing. The MRI data acquisition workflow will be implemented according to the *ReproNim* Guidelines (***Kennedy et al., 2019***). Subsequently, data will be converted into the Brain Imaging Data Structure (BIDS; ***Gorgolewski et al., 2016***). All MRI data will then be quality-assessed and preprocessed following previously published protocols (***Esteban et al., 2020***) with the corresponding *NiPreps* tools, i.e., *MRIQC* (***Esteban et al., 2017***), *fMRIPrep* (***Esteban et al., 2019***, RRID:SCR_016216), and *dMRIPrep* (***Joseph et al., 2021***). Both *fMRIPrep* and *dMRIPrep* will be set-up to use the same anatomical preprocessing and surface reconstruction from T1w and T2w images. The software versions of *fMRIPrep* and *dMRIPrep* tools will be locked on the latest minor release series available on the date of the last MRI session (e.g., 22.3.x, if the 22.3 series are already available). The software version of *MRIQC* will be locked on the latest minor available on the date of the first MRI session. All sessions will be processed with that same specific, *pinned* version. Preprocessing tools will generally be executed with default parameters, and deviations from defaults will be clearly reported. 'Analysis-grade' data, estimated nuisance time series, and estimated data transformations will be stored following the latest BIDS-Derivatives specifications available at the time.

Tensor-fitting, tractography, and extraction of structural connectivity. The analysis-grade dMRI data generated by *dMRIPrep* will then be used to estimate diffusion and Kurtosis tensors (DTI and DKI, respectively) with *MRtrix3* (***Tournier et al., 2019***), using a weighted linear least squares algorithm. From tensors, seven commonly used parametric scalar maps will be derived: fractional anisotropy (FA), mean/axial/radial diffusivity, and kurtosis (MD/AD/RD, MK/AK/RK, respectively). In parallel, orientation distribution functions (ODFs) will be reconstructed using constrained spherical deconvolution with *MRtrix3*. We will increase the sample size (to 108 units) by exploiting the redundancy of data and produce three groups of derived outcomes: (i) from all available diffusion-weighted polarities; and (ii) and (iii), from one subset of diffusion-weighted polarities and its complementary, respectively. Streamline deterministic tractography will subsequently be executed, ensuring seeding locations are anatomically

consistent across the 108 diffusion models. To achieve consistency, we will first determine the number of peaks per voxel from the ODFs maps. Then, we will project the peaks maps corresponding to each run into the T1w image's grid through a rigid body (six degrees of freedom) transformation as estimated by *dMRIPrep*. A total of two million random locations in the T1w reference will be determined within voxels in which 95% of the peaks maps have a single fiber (i.e., one peak). Those locations will then be mapped back to each run's native space to seed tractography. We will derive a track density map from the tractography results with *MRtrix3*. Results of tractography will be further regularized using COMMIT2 (***Schiavi et al., 2020***), an improvement over the original Convex Optimization Modeling for Microstructure Informed Tractography (COMMIT; ***Daducci et al., 2015***). Finally, SC matrices will be built by aggregating complete fiber tracks connecting two regions at a given scale of the *DiFuMo* atlas, after the projection of the atlas into the native (scanner) dMRI image space. SC matrices will be then weighted by the grand mean of track density values sampled across the tracks involved in the corresponding edge. As it has been previously reported (***Zalesky et al., 2016***; ***Sarwar et al., 2021***), the specific parameters in tractography largely determine the false positive and false negative rates of the resulting connectivity matrices. Therefore, the angular resolution and the number of seeds will be optimized to maximize the average repeatability of dMRI-derived scalars on the dMRI data collected as part of the piloting stage (Study timeline section). The optimized parameters will be registered in our SOPs document before the acquisition of the first MRI session.

Extraction of functional connectivity. FC will be extracted from whole-brain, RSfMRI time-series generated with *fMRIPrep*, sampled on the cortical mid-thickness surface, and on a regular, volumetric grid for the subcortical and cerebellar regions and stored in a mixed volume and surface format (CIFTI2). Time points at the onset of large head motions will be removed and replaced by interpolation from contiguous time points using cubic B-Spline kernels, whenever the corresponding framewise displacement (***Power et al., 2014***) calculated by *fMRIPrep* is above a given threshold. We will replicate and report the analyses with three different values for this theshold (0.2 mm, 0.5 mm, and 0.8 mm) covering the most typical range of values for this procedure. A subset of the confounds produced by *fMRIPrep* as well as temporal discrete cosine regressors, will be projected on the orthogonal of the signal space (***Friston et al., 1994***) and simultaneously (***Lindquist et al., 2019***) removed from the mixed volume and surface (CIFTI2) time series. Temporal filtering and confound removal by residualization will be implemented with *Nilearn* (***Abraham et al., 2014***) version 0.9. Subsequently, signals will be extracted and averaged parcel-wise, with the parcellation schemes described before. FC will be estimated as the inverse covariance matrix (***Smith et al., 2011***) of the extracted and averaged time-series estimated by a graphical lasso implemented with *Scikit-Learn* (***Abraham et al., 2014***). L_1 regularization will enforce sparsity in the estimated matrices with the goal of producing more specific FC by effectively reducing false positive connections. This approach has the shortcoming of likely increasing the false negative rate, which we hypothesize can be limited by leveraging data redundancy (i.e., session repetition), an advantage that is typically unavailable to connectivity studies implementing more standard FC extraction methodologies. Nonetheless, any alternative, exploratory analyses with more standard FC extraction settings will be reported if the sparsity of the FC matrices impedes a successful characterization of reliability or the comparison between scanners.

Reliability characterization of parametric diffusion maps and structural connectivity

First, we will visually evaluate the reliability of dMRI-derived scalar maps. A visual assessment of the within-scanner variability of SC will follow. These two first evaluations will only use the scalar maps derived from the 36 diffusion models, that is, those using the full data in each session. Finally, we propose a comprehensive Bayesian inference framework for which the total sample of 108 SCs will be applied.

Reliability of dMRI-derived scalar maps. To identify structured biases and bundle-wise robustness, we will calculate the average of each scalar map in each region of interest defined by the atlas. Those summaries will be reported with Bland-Altman (BA; ***Bland and Altman, 1986***) plots, graphically showing the difference with respect to the average of two repeated measures to investigate their agreement.

Hierarchical clustering analysis of within-scanner variability. The lower triangle of each $N_n \times N_n$ SC matrix ($N_n \in \{64, 128, 512\}$ corresponds to the number of nodes in the SC and depends on the scale of the parcellation) will be unraveled to a $1 \times C$ row vector ($C = N_n \cdot (N_n - 1)/2$ connections). All row vectors corresponding to SCs will be stacked together in a $N_{SC} \times C$ matrix ($\mathbf{G}$; $N_{SC} = 36$ is the number of SC matrices from the reliability dataset). Then, the C columns of the matrix will be reordered via a hierarchical cluster analysis using Ward's minimum variance method with matrix rows as features, to index together as columns with similar connectivity patterns. Finally, the rows (N_{SC} connectivity vectors) will then be sorted using the same hierarchical clustering to identify what sessions are more similar. A dendrogram representation of the reordered N_{SC} sessions will then be generated, and represented with a heatmap of each session's time of day, average head-motion summarized by framewise displacement, room temperature at start, and PE direction (***Begnoche et al., 2022***). To corroborate the hierarchical analysis, we will use t-distributed Stochastic Neighbor Embedding (t-SNE; ***van der Maaten and Hinton, 2008***)

to project the N_{SC} connectivity vectors of C features into a plane, and color the projected samples with the same factors as for the dendrogram. These analyses will be implemented with *Scikit-Learn*, pinned on a specific version available at the start of the collection effort.

Bland-Altman analysis generalized to multiple repetitions. All the pairwise subtractions (i.e., all combinations of two) of the N_{SC} connectivity vectors will be calculated and normalized as a percent change over the grand mean value of that particular connection. In the pool of $N_n \cdot (N_n - 1)/2 \cdot N_{SC}$ subtractions, we will select the ten connections with the largest pairwise differences (normalized by the average difference in the pool). BA plots of those ten connections will be represented with the average connection strength between each pair of sessions on the x-axis and the difference in connection strength between each pair of sessions on the y-axis. Likewise, we will also generate BA plots of the ten connections with the largest median and normalized differences. BA plots will be represented with three different hues denoting track length, track-averaged head motion, and track-averaged tractography density.

Within-scanner differences of structural connectomes. We hypothesize there are specific network edges that are less reliable, and will test whether there is any interaction with the PE direction of the dMRI data. Inference of SC will be derived from the Bayesian approach of ***Hinne et al.*** (***2013***), where the quantity of interest is the posterior distribution of adjacency matrices $\mathbf{G}$ given the streamlines connecting two regions obtained with tractography $\mathbf{S}$, a set of hyperparameters $\boldsymbol{\theta}$, and the edge length $\mathbf{l}_{i,j}$. The complete posterior expressing our knowledge of the SC is derived from Bayes' theorem:

$$\mathrm{P}(\mathbf{G} \mid \mathbf{S}, \boldsymbol{\theta}, \mathbf{l}_{i,j}) \propto \mathrm{P}(\mathbf{S} \mid \mathbf{G}, \boldsymbol{\theta}) \cdot \mathrm{P}(\mathbf{G} \mid \mathbf{l}_{i,j}). \tag{1}$$

Posterior distributions will be approximated employing Markov chain Monte Carlo sampling. Finally, reliability will be characterized by reporting the spread of edge-wise posterior distributions and the credible intervals of the posterior distribution of a *network* prior. These credible intervals of the posterior will also determine the probability of a type-S (sign) error of the best-fitting model. The network prior $\mathrm{P}(\mathbf{G} \mid \mathbf{l}_{i,j})$ models the probability $p_{i,j}$ that an edge between two regions (i, j) exists, assuming the individual edges are *a priori* independent:

$$\begin{aligned} \mathbf{g}_{i,j} \mid p_{i,j} &\sim \text{Bernoulli}(p_{i,j}), \text{and} \\ p_{i,j} \mid \mathbf{l}_{i,j} &\sim \text{Exponential}(\mathbf{l}_{i,j}), \end{aligned} \tag{2}$$

with $\mathbf{g}_{i,j}$ being an edge in $\mathbf{G}$. In Equation 2, we deviate from (***Hinne et al., 2013***) to introduce the edge length ($\mathbf{l}_{i,j}$) into the model as a prior on the probability $p_{i,j}$. Finally, we will examine the potential biases introduced by the diffusion weighting polarity and phase-encoding direction, modeled as random effects, with the introduction of an additional hierarchical prior. While ***Hinne et al.*** (***2013***) propose this before modeling the variability between individuals in their SCs, we will derive it to account for the different experimental conditions to determine structure biases and significant differences in the spread of the posterior distribution $\mathrm{P}(\mathbf{G}^m \mid \mathbf{S}^m, \boldsymbol{\theta}^m, \mathbf{l}^m_{i,j})$, where m indicates the analysis group. The number of posterior draws will be determined by the speed and convergence of posterior checks (Visual inspection, Gelman-Rubin statistic). Additional models will be formulated where the fixed and random effects are not modeled or hierarchical, and other variables may be considered, such as top-down initialization of priors from the coarsest scale of parcellation (64 regions) to the finest (512 regions). Hyperparameters may change if model convergence fails. Unless the registered model fails to converge as currently described, it will be included in the model comparison. Model comparison will be performed using the Watanabe-Akaike information criterion and leave-one-out information criteria. All the modeling framework will be implemented with the most recent stable release of *PyMC* (***Salvatier et al., 2016***) at the conclusion of data collection.

Reliability characterization of functional connectivity

Characterizing the reliability of FC. By consistently replaying the same movie clip (***Finn and Bandettini, 2021***) within constant experimental conditions, we hypothesize habituation will emerge in the RSfMRI scans that (i) reduce the variability of behavioral components of FC (which indeed is the principal challenge to studying reliability, ***Finn and Rosenberg, 2021***); and (ii) will temporally align some of the physiological sources of BOLD, thereby yielding larger between-trial correlations and enabling more reliable modeling (and removal) of these effects. Methodologically, we will approach the question with the analytical framework proposed by ***Bari et al.*** (***2019***), where the individual FC matrices are aggregated in a single principal component analysis (PCA), and then each FC can be expressed as a linear combination of a number of components and an adjacency matrix of residuals. Should our hypothesis be correct, habituation should consistently increase over time of the total variance explained for individual FC matrices, given a fixed number of components. We will determine such a number with the elbow plot method to select the minimum number of components. If the elbow plot yields a smooth increase in variance explained as a function of the number of components, we will select the minimum number of components that explain at least 85% of the variance. We will test the significance of habituation while accounting for confounding

effects by establishing a mixed-effects linear model on the residual matrices. To gain insight into how to best implement the model, we will first attempt a descriptive approach by plotting the edgewise residual variance along with the corresponding inter-quartile range to visually confirm the expected decrease as a function of the session index. The data points will be fitted with linear, quadratic, and negative exponential curves, all of which will be represented on the visualization, however, highlighting the best-fitting one with the highest R^2 value. Furthermore, to inspect which random predictor of no interest (among the time of day, day of the week, coffee consumption, etc.) should be included as random effects; we will plot the fits separately for bins of factor value (e.g., separately fit the sessions with respect to the day of acquisition). We will also visually assess whether some functional networks are more sensitive to habituation than others by separately plotting the edges grouped by networks at the lowest atlas scale (64 regions). Finally, we will build a baseline mixed-effects linear model with session index as a fixed effect and will add the orthogonalized predictors of no interest deemed relevant in the descriptive analysis as random effects. The significance of habituation will be tested by comparing the latter model to the same model without the sessions index as a fixed effect, using the likelihood ratio test with a statistical significance threshold of $\alpha = 0.05$. If more models are to be compared, we will also use the likelihood ratio test and all the tested models and results will be fully reported.

Characterizing scanner effects

Scanner effects on dMRI-derived scalar maps The scalar maps will be preprocessed for the within-scanner evaluation. Conversely, the variability will be computed with the coefficient of variability (CoV; Equation 3) across scan sessions for each of the seven diffusion scalar maps on a voxel-wise basis, as well as within regions of interest resulting from the atlas-based segmentation.

$$\mathrm{CoV}(\%) = \frac{\sigma}{\bar{x}} \cdot 100\%, \tag{3}$$

where σ corresponds to the standard deviation of the scalar metrics and $\bar{x}$ is the mean of the scalar metrics.

Between-scanner differences of structural connectomes. The SCs extracted from the clinical scanners will separately be modeled with the Bayesian framework previously described for the within-scanner reliability analysis. Differences in SC matrices between scanners will be evaluated by calculating the Bayes factor between the posterior distributions corresponding to each of the three groups. Bayes factors will be reported and interpreted as described by ***Jeffreys*** (Appendix B, 1998).

Determining scanner effects on functional connectivity. Since the standard protocol does not provide RSfMRI while movie-watching, we will determine scan-effects using the single-echo BOLD runs included in both reliability and standard protocols. We will first investigate differences in FC between the BOLD runs with and without naturalistic movie. To do so, we will residualize each of the 36 FC adjacency matrices (extracted from the single-echo, without-movie data) with respect to the PCA model developed above for the multi-echo BOLD data. We will determine the statistical significance of differences in residuals with a repeated-measures, multivariate analysis of variance (RM-MANOVA), with two groups (with/without movie clip) and 36 repetitions. Secondly, to test for scanner effects, we will re-execute the PCA decomposition using only the single-echo, without-movie data, and establish a subsequent RM-MANOVA with 3 groups (one per scanner device), on the residualized adjacency matrices. The statistical significance threshold will be set at the α=0.05 level for all tests.

Study timeline

The anticipated total duration of the study is 68 weeks, starting from the Stage 1 (accepted in principle) milestone. The timeline involves 'freezing' several sections of the protocol, as they are set as final (e.g., the prescription of all the parameters in the MRI protocol). The SOPs document will receive updates until the freeze of a particular section, in which an annotated git tag will be created for preservation. Deviations from the registered report will therefore be easily tracked by finding differences with respect to the git commit tagged by the corresponding freeze.

Quality assurance of the research workflow. We have acquired a small sample (n = 18 subjects, two sessions each) of healthy volunteers under the same ethical approval, with an MRI protocol including both dMRI and RSfMRI, which will be used to prepare the analysis workflow for the execution of the experiment's data. This dataset is currently being used to ascertain the feasibility of the neuroimaging workflow and take corrective actions before acquiring data. This quality assurance dataset will be released openly, but independently of this particular study.

Piloting the MRI protocols. We have already initiated the deployment of the study protocols (Table 4 and Table 3 on the three scanners involved. The study's participant is currently running pilot sessions using these protocols to ascertain the feasibility of the full protocol within the time frame of a session. Pilot sessions will be released with the dataset with a clear indication of their piloting nature, and will be excluded from final analyses. These sessions will also be applied to validate the processing workflow from data acquisition to preprocessed derivatives.

Milestone 1: MRI protocol freeze. Once the piloting stage is deemed complete, the reliability and standardized protocols will be fixed for the remainder of the study. Modifications to the protocol after this point will require comprehensive justification. The protocols will be exported to shareable formats under the ReproNim guidelines and released as part of the unprocessed data package, to ensure the reproducibility of the MR acquisitions.

Data acquisition and quality assessment. Data collection of the reliability dataset will span 18 weeks, starting after clearing Stage 1. Collection of the standard dataset will start after the reliability dataset has been completed, within the subsequent 18 weeks. Each session will be quality-assessed with *MRIQC* as described in our previous report (***Provins et al., 2023***), and prepared for archival in BIDS format within the span of two weeks after each session.

Milestone 2: Completion of data collection. The data acquisition phase will culminate when the target sample size (in terms of the number of sessions) is achieved for all three MR scanners, including repeat sessions to replace exclusions due to the substandard quality of the acquired data. A distributable package following *BIDS* will be created as the main deliverable. Facial features present in the anatomical data (T1w and T2w images) will be removed (*defacing*) to enable redistribution of the dataset on OpenNeuro (***Markiewicz et al., 2021***). However, based on our previous exploration (***Provins et al., 2022***), processing will be carried out on the original images preserving the face. This *unprocessed* data package (see Figure 1) will include all the corresponding *MRIQC*-generated reports and quality metrics for anatomical, dMRI, and BOLD images, as well as any other data derived from quality assessments that can be distributed. This package will be released at the conclusion of the study.

Implementation of the software instrumentation for processing and analysis. The data analysis scripts will make use of several SciPy (***Virtanen et al., 2020***) and R packages. The implementation of software instruments is scheduled to start at the conclusion of the piloting phase and will be finished before the start of the data analysis phase (i.e., week 37 after Stage 1 or earlier). Software execution on high-performance computing resources (e.g., SLURM scheduling) will also be trialed out during this phase, before analysis. To maximize the overall replicability of the study and preserve provenance information of all derived results, the execution will make use of containerized and versioned environments and the *DataLad*'s 'containers-run' extension. All the source code will be tested on the piloting data.

Milestone 3: Completion of data preprocessing. Preprocessing with *fMRIPrep* and *dMRIPRep* will be initiated upon completion of data collection, and a second checkpoint for the quality assessment of derivatives will be executed (***Provins et al., 2023***). Both preprocessed data, diffusion-derived parametric maps and the extracted SC & FC matrices will also be organized according to *BIDS Derivatives* and prepared as a distributable package for distribution upon finalization of the study.

Milestone 4: Analysis code freeze. Before initiating data analysis, all the analysis code will be locked to a particular commit of the git repository, and preserved with an annotated git tag. The code will then be packaged and made publicly available under the Apache 2.0 license (deliverable).

Data analysis. 'Analysis-grade' connectivity data will be retrieved once data collection is concluded (that is, starting on week 37 the latest) and analyzed following the anticipated protocols over the period of five months. The writing and edition of the Stage 2 report will be executed subsequently to the finalization of each analysis.

Milestone 5: Collection of results. The results and a comprehensive report of deviations from the registered study will culminate in the data analysis phase. The analysis code will again be tagged, including bugfixes and registration deviations from the previous freeze, as well as new code for the generation of figures and tables of the report. The main deliverable of this milestone is the draft of the final manuscript to be submitted to initiate Stage 2 of the study.

Stage 2 submission. Submission of the final results is scheduled before week 68 or earlier.

References

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics. ; 2014, 8. doi:10.3389/fninf.2014.00014.

Alemán-Gómez Y, Griffa A, Houde JC, Najdenovska E, Magon S, Cuadra MB, Descoteaux M, Hagmann P. A multi-scale probabilistic atlas of the human connectome. Scientific Data. ; 2022, 9(1):516. doi:10.1038/s41597-022-01624-8.

Allen EJ, St-Yves G, Wu Y, Breedlove JL, Prince JS, Dowdle LT, Nau M, Caron B, Pestilli F, Charest I, Hutchinson JB, Naselaris T, Kay K. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature Neuroscience. ; 2022, 25(1):116–126. doi:10.1038/s41593-021-00962-x.

Arnatkeviciute A, Fulcher BD, Bellgrove MA, Fornito A. Where the genome meets the connectome: Understanding how genes shape human brain connectivity. NeuroImage. ; 2021, 244:118570. doi:10.1016/j.neuroimage.2021.118570.

Bargmann CI, Marder E. From the connectome to brain function. Nature Methods. ; 2013, 10(6):483–490. doi:10.1038/nmeth.2451.

Bari S, Amico E, Vike N, Talavage TM, Goñi J. Uncovering multi-site identifiability based on resting-state functional connectomes. NeuroImage. ; 2019, 202:115967. doi:10.1016/j.neuroimage.2019.06.045.

Bassett DS, Sporns O. Network neuroscience. Nature Neuroscience. ; 2017, 20(3):353–364. doi:10.1038/nn.4502.

Begnoche JP, Schilling KG, Boyd BD, Cai LY, Taylor WD, Landman BA. EPI susceptibility correction introduces significant differences far from local areas of high distortion. Magnetic Resonance Imaging. ; 2022, 92:1–9. doi:10.1016/j.mri.2022.05.016.

Bijsterbosch J, Harrison SJ, Jbabdi S, Woolrich M, Beckmann C, Smith S, Duff EP. Challenges and future directions for representations of functional brain organization. Nature Neuroscience. ; 2020, 23(12):1484–1495. doi:10.1038/s41593-020-00726-z.

Birn RM, Smith MA, Jones TB, Bandettini PA. The respiration response function: The temporal dynamics of fMRI signal fluctuations related to changes in respiration. NeuroImage. ; 2008, 40(2):644–654. doi:10.1016/j.neuroimage.2007.11.059.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. ; 1986, 327(8476):307–310. doi:10.1016/S0140-6736(86)90837-8.

Bolton TAW, Kebets V, Glerean E, Zöller D, Li J, Yeo BTT, Caballero-Gaudes C, Van De Ville D. Agito ergo sum: Correlates of spatio-temporal motion characteristics during fMRI. NeuroImage. ; 2020, 209:116433. doi:10.1016/j.neuroimage.2019.116433.

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford JA, Adcock RA, Avesani P, Baczkowski BM, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner J, Benoit RG, et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature. ; 2020, 582(7810):84–88. doi:10.1038/s41586-020-2314-9.

Bright MG, Murphy K. Reliable quantification of BOLD fMRI cerebrovascular reactivity despite poor breath-hold performance. NeuroImage. ; 2013, 83:559–568. doi:10.1016/j.neuroimage.2013.07.007.

Bright MG, Whittaker JR, Driver ID, Murphy K. Vascular physiology drives functional brain networks. NeuroImage. ; 2020, 217:116907. doi:10.1016/j.neuroimage.2020.116907.

Buchanan CR, Pernet CR, Gorgolewski KJ, Storkey AJ, Bastin ME. Test–retest reliability of structural brain networks from diffusion MRI. NeuroImage. ; 2014, 86:231–243. doi:10.1016/j.neuroimage.2013.09.054.

Bulte D, Wartolowska K. Monitoring cardiac and respiratory physiology during FMRI. NeuroImage. ; 2017, 154:81–91. doi:10.1016/j.neuroimage.2016.12.001.

Cai LY, Yang Q, Kanakaraj P, Nath V, Newton AT, Edmonson HA, Luci J, Conrad BN, Price GR, Hansen CB, Kerley CI, Ramadass K, Yeh FC, Kang H, Garyfallidis E, Descoteaux M, Rheault F, Schilling KG, Landman BA. MASiVar: Multisite, multiscanner, and multisubject acquisitions for studying variability in diffusion weighted MRI. Magnetic Resonance in Medicine. ; 2021, 86(6):3304–3320. doi:10.1002/mrm.28926.

Caruyer E, Lenglet C, Sapiro G, Deriche R. Design of multishell sampling schemes with uniform coverage in diffusion MRI. Magnetic Resonance in Medicine. ; 2013, 69(6):1534–1540. doi:10.1002/mrm.24736.

Chang C, Cunningham JP, Glover GH. Influence of heart rate on the BOLD signal: The cardiac response function. NeuroImage. ; 2009, 44(3):857–869. doi:10.1016/j.neuroimage.2008.09.029.

Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage. ; 2010, 50(1):81–98. doi:10.1016/j.neuroimage.2009.12.011.

Chen JE, Lewis LD, Chang C, Tian Q, Fultz NE, Ohringer NA, Rosen BR, Polimeni JR. Resting-state "physiological networks". NeuroImage. ; 2020, 213:116707. doi:10.1016/j.neuroimage.2020.116707.

Choe AS, Jones CK, Joel SE, Muschelli J, Belegu V, Caffo BS, Lindquist MA, Zijl PCMv, Pekar JJ. Reproducibility and Temporal Structure in Weekly Resting-State fMRI over a Period of 3.5 Years. PLOS ONE. ; 2015, 10(10):e0140134. doi:10.1371/journal.pone.0140134.

Ciric R, Thompson WH, Lorenz R, Goncalves M, MacNicol EE, Markiewicz CJ, Halchenko YO, Ghosh SS, Gorgolewski KJ, Poldrack RA, Esteban O. TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. Nature Methods. ; 2022, 19:1568–1571. doi:10.1038/s41592-022-01681-2.

Cosgrove KT, McDermott TJ, White EJ, Mosconi MW, Thompson WK, Paulus MP, Cardenas-Iniguez C, Aupperle RL. Limits to the generalizability of resting-state functional magnetic resonance imaging studies of youth: An examination of ABCD Study® baseline data. Brain Imaging and Behavior. ; 2022. doi:10.1007/s11682-022-00665-2.

Dadi K, Varoquaux G, Machlouzarides-Shalit A, Gorgolewski KJ, Wassermann D, Thirion B, Mensch A. Fine-grain atlases of functional modes for fMRI analysis. NeuroImage. ; 2020, 221:117126. doi:10.1016/j.neuroimage.2020.117126.

Daducci A, Dal Palú A, Lemkaddem A, Thiran JP. COMMIT: Convex Optimization Modeling for Microstructure Informed Tractography. IEEE Transactions on Medical Imaging. ; 2015, 34(1):246–257. doi:10.1109/TMI.2014.2352414.

Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. Proceedings of the National Academy of Sciences. ; 2006, 103(37):13848–13853. doi:10.1073/pnas.0601417103.

Douaud G, Lee S, Alfaro-Almagro F, Arthofer C, Wang C, McCarthy P, Lange F, Andersson JLR, Griffanti L, Duff E, Jbabdi S, Taschler B, Keating P, Winkler AM, Collins R, Matthews PM, Allen N, Miller KL, Nichols TE, Smith SM. SARS-CoV-2 is associated with changes in brain structure in UK Biobank. Nature. ; 2022, p. 1–17. doi:10.1038/s41586-022-04569-5.

Doucet GE, Lee WH, Frangou S. Evaluation of the spatial variability in the major resting-state networks across human brain functional atlases. Human Brain Mapping. ; 2019, 40(15):4577–4587. doi:10.1002/hbm.24722.

Dubol M, Epperson CN, Sacher J, Pletzer B, Derntl B, Lanzenberger R, Sundström-Poromaa I, Comasco E. Neuroimaging the menstrual cycle: A multimodal systematic review. Frontiers in Neuroendocrinology. ; 2021, 60:100878. doi:10.1016/j.yfrne.2020.100878.

Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLOS ONE. ; 2017, 12(9):e0184661. doi:10.1371/journal.pone.0184661.

Esteban O, Ciric R, Finc K, Blair R, Markiewicz CJ, Moodie CA, Kent JD, Goncalves M, DuPre E, Gomez DEP, Ye Z, Salo T, Valabregue R, Amlien IK, Liem F, Jacoby N, Stojić H, Cieslak M, Urchs S, Halchenko YO, et al. Analysis of task-based functional MRI data preprocessed with fMRIPrep. Nature Protocols. ; 2020, 15:2186–2202. doi:10.1101/694364.

Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods. ; 2019, 16(1):111–116. doi:10.1038/s41592-018-0235-4.

Etzel JA, Brough RE, Freund MC, Kizhner A, Lin Y, Singh MF, Tang R, Tay A, Wang A, Braver TS. The Dual Mechanisms of Cognitive Control dataset, a theoretically-guided within-subject task fMRI battery. Scientific Data. ; 2022, 9(1):114. doi:10.1038/s41597-022-01226-4.

Filevich E, Lisofsky N, Becker M, Butler O, Lochstet M, Martensson J, Wenger E, Lindenberger U, Kühn S. Day2day: investigating daily variability of magnetic resonance imaging measures over half a year. BMC Neuroscience. ; 2017, 18(1):65. doi:10.1186/s12868-017-0383-y.

Finn ES, Bandettini PA. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. NeuroImage. ; 2021, 235:117963. doi:10.1016/j.neuroimage.2021.117963.

Finn ES, Rosenberg MD. Beyond fingerprinting: Choosing predictive connectomes over reliable connectomes. NeuroImage. ; 2021, 239:118254. doi:10.1016/j.neuroimage.2021.118254.

Finn ES, Scheinost D, Finn DM, Shen X, Papademetris X, Constable RT. Can brain state be manipulated to emphasize individual differences in functional connectivity? NeuroImage. ; 2017, 160:140–151. doi:10.1016/j.neuroimage.2017.03.064.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nature Neuroscience. ; 2015, 18(11):1664–1671. doi:10.1038/nn.4135.

Fonov VS, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage. ; 2011, 54(1):313–327. doi:10.1016/j.neuroimage.2010.07.033.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping. ; 1994, 2(4):189–210. doi:10.1002/hbm.460020402.

Froeling M, Tax CMW, Vos SB, Luijten PR, Leemans A. "MASSIVE" Brain Dataset: Multiple Acquisitions for Standardization of Structural Imaging Validation and Evaluation. Magnetic Resonance in Medicine. ; 2016, p. n/a–n/a. doi:10.1002/mrm.26259.

Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF. Precision Functional Mapping of Individual Human Brains. Neuron. ; 2017, 95(4):791–807.e7. doi:10.1016/j.neuron.2017.07.011.

Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA, Hanke M, Keator D, Li X, Michael Z, Maumet C, Nichols BN, Nichols TE, Pellman J, Poline JB, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data. ; 2016, 3:160044. doi:10.1038/sdata.2016.44.

Gratton C, Laumann TO, Nielsen AN, Greene DJ, Gordon EM, Gilmore AW, Nelson SM, Coalson RS, Snyder AZ, Schlaggar BL, Dosenbach NUF, Petersen SE. Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. Neuron. ; 2018, 98(2):439–452.e5. doi:10.1016/j.neuron.2018.03.035.

Greenwell S, Faskowitz J, Pritschet L, Santander T, Jacobs EG, Betzel RF. High-amplitude network co-fluctuations linked to variation in hormone concentrations over menstrual cycle. bioRxiv. ; 2021, p. 2021.07.29.453892. doi:10.1101/2021.07.29.453892.

Griffa A, Baumann PS, Thiran JP, Hagmann P. Structural connectomics in brain diseases. NeuroImage. ; 2013, 80:515–526. doi:10.1016/j.neuroimage.2013.04.056.

Griffa A, Ricaud B, Benzi K, Bresson X, Daducci A, Vandergheynst P, Thiran JP, Hagmann P. Transient networks of spatio-temporal connectivity map communication pathways in brain functional systems. NeuroImage. ; 2017, 155:490–502. doi:10.1016/j.neuroimage.2017.04.015.

Hagmann P. From diffusion MRI to brain connectomics. PhD. Thesis, École Polytechnique Fédérale de Lausanne; 2005.

Hagmann P, Cammoun L, Gigandet X, Gerhard S, Ellen Grant P, Wedeen V, Meuli R, Thiran JP, Honey CJ, Sporns O. MR connectomics: Principles and challenges. Journal of Neuroscience Methods. ; 2010, 194(1):34–45. doi:10.1016/j.jneumeth.2010.01.014.

Hanke M, Baumgartner FJ, Ibe P, Kaule FR, Pollmann S, Speck O, Zinke W, Stadler J. A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. Scientific Data. ; 2014, 1(1):140003. doi:10.1038/sdata.2014.3.

Harvey JL, Demetriou L, McGonigle J, Wall MB. A short, robust brain activation control task optimised for pharmacological fMRI studies. PeerJ. ; 2018, 6:e5540. doi:10.7717/peerj.5540.

Hermundstad AM, Bassett DS, Brown KS, Aminoff EM, Clewett D, Freeman S, Frithsen A, Johnson A, Tipper CM, Miller MB, Grafton ST, Carlson JM. Structural foundations of resting-state and task-based functional connectivity in the human brain. Proceedings of the National Academy of Sciences. ; 2013, 110(15):6169–6174. doi:10.1073/pnas.1219562110.

Hinne M, Heskes T, Beckmann CF, van Gerven MAJ. Bayesian inference of structural brain networks. NeuroImage. ; 2013, 66:543–552. doi:10.1016/j.neuroimage.2012.09.068.

Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, Hagmann P. Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences of the United States of America. ; 2009, 106(6):2035–40. doi:10.1073/pnas.0811168106.

Horien C, Shen X, Scheinost D, Constable RT. The individual functional connectome is unique and stable over months to years. NeuroImage. ; 2019, 189:676–687. doi:10.1016/j.neuroimage.2019.02.002.

Jeffreys H. The Theory of Probability. OUP Oxford; 1998.

Joseph MJE, Pisner D, Richie-Halford A, Lerma-Usabiaga G, Mansour S, Kent JD, Keshavan A, Cieslak M, Dickie EW, Tourbier S, Satterthwaite TD, Poldrack RA, Veraart J, Rokem A, Esteban O. dMRIPrep: a robust preprocessing pipeline for diffusion MRI. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 30 Virtual Meeting; 2021. p. 2473. https://docs.google.com/document/u/2/d/1ocamAFP2OGnUlUooL9gxu5CExqiCS4Le3caHWM_8E04/edit?usp=drive_web&ouid=104994410621175933959&usp=embed_facebook.

Kennedy DN, Abraham SA, Bates JF, Crowley A, Ghosh S, Gillespie T, Goncalves M, Grethe JS, Halchenko YO, Hanke M, Haselgrove C, Hodge SM, Jarecka D, Kaczmarzyk J, Keator DB, Meyer K, Martone ME, Padhy S, Poline JB, Preuss N, et al. Everything Matters: The ReproNim Perspective on Reproducible Neuroimaging. Frontiers in Neuroinformatics. ; 2019, 13. doi:10.3389/fninf.2019.00001.

Laird AR. Large, open datasets for human connectomics research: Considerations for reproducible and responsible data use. NeuroImage. ; 2021, 244:118579. doi:10.1016/j.neuroimage.2021.118579.

Li X, Ai L, Giavasis S, Jin H, Feczko EJ, Xu T, Clucas J, Franco AR, Vogelstein JT, Yan CG, Esteban O, Poldrack RA, Craddock RC, Fair DA, Satterthwaite TD, Kiar G, Milham MP. Moving Beyond Processing and Analysis-Related Variation in Neuroscience. Nature Portfolio. ; 2021, (In Review).

Lindquist MA, Geuter S, Wager TD, Caffo BS. Modular preprocessing pipelines can reintroduce artifacts into fMRI data. Human Brain Mapping. ; 2019, 40(8):2358–2376. doi:10.1002/hbm.24528.

Lipp I, Murphy K, Caseras X, Wise RG. Agreement and repeatability of vascular reactivity estimates based on a breath-hold task and a resting state scan. NeuroImage. ; 2015, 113:387–396. doi:10.1016/j.neuroimage.2015.03.004.

van der Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. ; 2008, 9(86):2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html.

Maier-Hein KH, Neher PF, Houde JC, Côté MA, Garyfallidis E, Zhong J, Chamberland M, Yeh FC, Lin YC, Ji Q, Reddick WE, Glass JO, Chen DQ, Feng Y, Gao C, Wu Y, Ma J, Renjie H, Li Q, Westin CF, et al. The challenge of mapping the human connectome based on diffusion tractography. Nature Communications. ; 2017, 8(1):1349. doi:10.1038/s41467-017-01285-x.

Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, Hardcastle N, Wexler J, Esteban O, Goncavles M, Jwa A, Poldrack RA. The OpenNeuro resource for sharing of neuroscience data. eLife. ; 2021, 10:e71774. doi:10.7554/eLife.71774.

Milham MP, Vogelstein J, Xu T. Removing the Reliability Bottleneck in Functional Magnetic Resonance Imaging Research to Achieve Clinical Utility. JAMA Psychiatry. ; 2021, 78(6):587–588. doi:10.1001/jamapsychiatry.2020.4272.

Moia S, Termenon M, Uruñuela E, Chen G, Stickland RC, Bright MG, Caballero-Gaudes C. ICA-based denoising strategies in breath-hold induced cerebrovascular reactivity mapping with multi echo BOLD fMRI. NeuroImage. ; 2021, 233:117914. doi:10.1016/j.neuroimage.2021.117914.

Moreton FC, Dani KA, Goutcher C, O'Hare K, Muir KW. Respiratory challenge MRI: Practical aspects. NeuroImage: Clinical. ; 2016, 11:667–677. doi:10.1016/j.nicl.2016.05.003.

Mueller S, Wang D, Fox MD, Yeo BTT, Sepulcre J, Sabuncu MR, Shafee R, Lu J, Liu H. Individual Variability in Functional Connectivity Architecture of the Human Brain. Neuron. ; 2013, 77(3):586–595. doi:10.1016/j.neuron.2012.12.028.

Naselaris T, Allen E, Kay K. Extensive sampling for complete models of individual brains. Current Opinion in Behavioral Sciences. ; 2021, 40:45–51. doi:10.1016/j.cobeha.2020.12.008.

Niso G, Botvinik-Nezer R, Appelhoff S, Vega ADL, Esteban O, Etzel JA, Finc K, Ganz M, Gau R, Halchenko YO, Herholz P, Karakuzu A, Keator DB, Markiewicz CJ, Maumet C, Pernet CR, Pestilli F, Queder N, Schmitt T, Sójka W, et al. Open and reproducible neuroimaging: from study inception to publication. NeuroImage. ; 2022, p. 119623. doi:10.1016/j.neuroimage.2022.119623.

Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. NeuroImage. ; 2019, 203:116157. doi:10.1016/j.neuroimage.2019.116157.

Noble S, Scheinost D, Constable RT. Cluster failure or power failure? Evaluating sensitivity in cluster-level inference. NeuroImage. ; 2020, 209:116468. doi:10.1016/j.neuroimage.2019.116468.

Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. Influences on the Test–Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. Cerebral Cortex. ; 2017, 27(11):5415–5429. doi:10.1093/cercor/bhx230.

O'Connor D, Potler NV, Kovacs M, Xu T, Ai L, Pellman J, Vanderwal T, Parra LC, Cohen S, Ghosh S, Escalera J, Grant-Villegas N, Osman Y, Bui A, Craddock RC, Milham MP. The Healthy Brain Network Serial Scanning Initiative: a resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. GigaScience. ; 2017, 6(2):giw011. doi:10.1093/gigascience/giw011.

Pinho AL, Amadon A, Ruest T, Fabre M, Dohmatob E, Denghien I, Ginisty C, Becuwe-Desmidt S, Roger S, Laurier L, Joly-Testault V, Médiouni-Cloarec G, Doublé C, Martins B, Pinel P, Eger E, Varoquaux G, Pallier C, Dehaene S, Hertz-Pannier L, et al. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. Scientific Data. ; 2018, 5(1):180105. doi:10.1038/sdata.2018.105.

Pinto J, Bright MG, Bulte DP, Figueiredo P. Cerebrovascular Reactivity Mapping Without Gas Challenges: A Methodological Guide. Frontiers in Physiology. ; 2021, 11. https://www.frontiersin.org/articles/10.3389/fphys.2020.608475.

Poldrack RA, Congdon E, Triplett W, Gorgolewski KJ, Karlsgodt KH, Mumford JA, Sabb FW, Freimer NB, London ED, Cannon TD, Bilder RM. A phenome-wide examination of neural and cognitive function. Scientific Data. ; 2016, 3:160110. doi:10.1038/sdata.2016.110.

Poldrack RA, Laumann TO, Koyejo O, Gregory B, Hover A, Chen MY, Gorgolewski KJ, Luci J, Joo SJ, Boyd RL, Hunicke-Smith S, Simpson ZB, Caven T, Sochat V, Shine JM, Gordon E, Snyder AZ, Adeyemo B, Petersen SE, Glahn DC, et al. Long-term neural and physiological phenotyping of a single human. Nature Communications. ; 2015, 6:8885. doi:10.1038/ncomms9885.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage. ; 2012, 59(3):2142–2154. doi:10.1016/j.neuroimage.2011.10.018.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage. ; 2014, 84(Supplement C):320–341. doi:10.1016/j.neuroimage.2013.08.048.

Preti MG, Bolton TA, Van De Ville D. The dynamic functional connectome: State-of-the-art and perspectives. NeuroImage. ; 2017, 160:41–54. doi:10.1016/j.neuroimage.2016.12.061.

Provins C, Alemán-Gómez Y, Cleusix M, Jenni R, Richiardi J, Hagmann P, Esteban O. Defacing biases manual and automated quality assessments of structural MRI with MRIQC. In: *28th Annual Meeting of the Organization for Human Brain Mapping* Glasgow, Scotland: OSF Preprints; 2022. p. WTh566. doi:10.31219/osf.io/8mcyz.

Provins C, MacNicol EE, Seeley SH, Hagmann P, Esteban O. Quality Control in functional MRI studies with MRIQC and fMRIPrep. Frontiers in Neuroimaging. ; 2023, 1:1073734. doi:10.3389/fnimg.2022.1073734.

Rodrigue AL, Mastrovito D, Esteban O, Durnez J, Koenis MMG, Janssen R, Alexander-Bloch A, Knowles EM, Mathias SR, Mollon J, Pearlson GD, Frangou S, Blangero J, Poldrack RA, Glahn DC. Searching for Imaging Biomarkers of Psychotic Dysconnectivity. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. ; 2021. doi:10.1016/j.bpsc.2020.12.002.

Rosenberg MD, Finn ES. How to establish robust brain–behavior relationships without thousands of individuals. Nature Neuroscience. ; 2022, p. 1–3. doi:10.1038/s41593-022-01110-9.

Rosenberg MD, Scheinost D, Greene AS, Avery EW, Kwon YH, Finn ES, Ramani R, Qiu M, Constable RT, Chun MM. Functional connectivity predicts changes in attention observed across minutes, days, and months. Proceedings of the National Academy of Sciences. ; 2020, 117(7):3797–3807. doi:10.1073/pnas.1912226117.

Rosenthal G, Váša F, Griffa A, Hagmann P, Amico E, Goñi J, Avidan G, Sporns O. Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. Nature Communications. ; 2018, 9(1):2178. doi:10.1038/s41467-018-04614-w.

Sadaghiani S, Poline JB, Kleinschmidt A, D'Esposito M. Ongoing dynamics in large-scale functional connectivity predict perception. Proceedings of the National Academy of Sciences. ; 2015, 112(27):8463–8468. doi:10.1073/pnas.1420687112.

Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science. ; 2016, 2:e55. doi:10.7717/peerj-cs.55.

Sarwar T, Ramamohanarao K, Zalesky A. A critical review of connectome validation studies. NMR in Biomedicine. ; 2021, 34(12):e4605. doi:10.1002/nbm.4605.

Schiavi S, Ocampo-Pineda M, Barakovic M, Petit L, Descoteaux M, Thiran JP, Daducci A. A new method for accurate in vivo mapping of human brain connections using microstructural and anatomical information. Science Advances. ; 2020, 6(31):eaba8245. doi:10.1126/sciadv.aba8245.

Scouten A, Schwarzbauer C. Paced respiration with end-expiration technique offers superior BOLD signal repeatability for breath-hold studies. NeuroImage. ; 2008, 43(2):250–257. doi:10.1016/j.neuroimage.2008.03.052.

Seeliger K, Sommers RP, Güçlü U, Bosch SE, van Gerven MAJ. A large single-participant fMRI dataset for probing brain responses to naturalistic stimuli in space and time. bioRxiv. ; 2019, p. 687681. doi:10.1101/687681.

Seider NA, Adeyemo B, Miller R, Newbold DJ, Hampton JM, Scheidter KM, Rutlin J, Laumann TO, Roland JL, Montez DF, Van AN, Zheng A, Marek S, Kay BP, Bretthorst GL, Schlaggar BL, Greene DJ, Wang Y, Petersen SE, Barch DM, et al. Accuracy and reliability of diffusion imaging models. NeuroImage. ; 2022, 254:119138. doi:10.1016/j.neuroimage.2022.119138.

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. Network modelling methods for FMRI. NeuroImage. ; 2011, 54(2):875–891. doi:10.1016/j.neuroimage.2010.08.063.

Song H, Rosenberg MD. Predicting attention across time and contexts with functional brain connectivity. Current Opinion in Behavioral Sciences. ; 2021, 40:33–44. doi:10.1016/j.cobeha.2020.12.007.

Stickland RC, Zvolanek KM, Moia S, Ayyagari A, Caballero-Gaudes C, Bright MG. A practical modification to a resting state fMRI protocol for improved characterization of cerebrovascular function. NeuroImage. ; 2021, 239:118306. doi:10.1016/j.neuroimage.2021.118306.

Tournier JD, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, Christiaens D, Jeurissen B, Yeh CH, Connelly A. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. NeuroImage. ; 2019, 202:116137. doi:10.1016/j.neuroimage.2019.116137.

Uddin LQ. Bring the Noise: Reconceptualizing Spontaneous Neural Activity. Trends in Cognitive Sciences. ; 2020, 24(9):734–746. doi:10.1016/j.tics.2020.06.003.

Vaisvilaite L, Hushagen V, Grønli J, Specht K. Time-of-Day Effects in Resting-State Functional Magnetic Resonance Imaging: Changes in Effective Connectivity and Blood Oxygenation Level Dependent Signal. Brain Connectivity. ; 2022, 12(6):515–523. doi:10.1089/brain.2021.0129.

Vanderwal T, Eilbott J, Finn ES, Craddock RC, Turnbull A, Castellanos FX. Individual differences in functional connectivity during naturalistic viewing conditions. NeuroImage. ; 2017, 157:521–530. doi:10.1016/j.neuroimage.2017.06.027.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. ; 2020, 17(3):261–272. doi:10.1038/s41592-019-0686-2.

Vázquez-Rodríguez B, Suárez LE, Markello RD, Shafiei G, Paquola C, Hagmann P, Heuvel MPvd, Bernhardt BC, Spreng RN, Misic B. Gradients of structure–function tethering across neocortex. Proceedings of the National Academy of Sciences. ; 2019, 116(42):21219–21227. doi:10.1073/pnas.1903403116.

Zalesky A, Fornito A, Cocchi L, Gollo LL, van den Heuvel MP, Breakspear M. Connectome sensitivity or specificity: which is more important? NeuroImage. ; 2016, 142:407–420. doi:10.1016/j.neuroimage.2016.06.035.

Zamani Esfahlani F, Jo Y, Faskowitz J, Byrge L, Kennedy DP, Sporns O, Betzel RF. High-amplitude cofluctuations in cortical activity drive functional connectivity. Proceedings of the National Academy of Sciences. ; 2020, 117(45):28393–28401. doi:10.1073/pnas.2005531117.

Zuo XN, Xu T, Milham MP. Harnessing reliability for neuroscience research. Nature Human Behaviour. ; 2019, 3(8):768–771. doi:10.1038/s41562-019-0655-x.

Data and software availability statement

The source code to execute the analysis and generate the figures in the study will be made openly available under the terms of the Apache 2.0 license at https://github.com/TheAxonLab/hcph-reliability-analysis. Original data and preprocessing derivatives will be made openly available under the terms of the CC-BY license. Versioned container images created to encapsulate the analysis pipelines will also be made openly available under the terms of the MIT license, through the *Docker Hub* service and the *GitHub Package registry*. Git and Git-annex repositories containing the provenance tracking of *DataLad*'s 'containers-run' will be openly released under suitable data and software licenses under https://gin.g-node.org/oesteban.

Diversity, Equity, and Inclusiveness statement

This manuscript acknowledges the biases in citation practices regarding gender, ethnicity, and academic pedigree. Before submission, we analyzed the gender distribution of the first and last authors in our references with an automated Python script (available at https://github.com/oesteban/bibbias under an Apache 2.0 license) that programmatically checked the first name of authors against the Gender-API service (gender-api.com) to determine the likely gender of the author. Names with low certainty in the Gender-API response were manually revised to the best of our efforts in trying to find academic websites corresponding to the author. A 54% of the references have a woman as the first author, of which only 19% of the cases the last author was also a woman. The last authors are overwhelmingly males (87%). When the last author was a woman, only in 21% of the references the first author is a male (i.e., only 3% of all the references have male-first-female-last authors). During the development of Stage 2 of the report, and in the final writing of the manuscript, we will continue to actively seek to minimize the gender imbalance in our references without loss in scientific quality.

Author contributions

Data acquisition: CP, YA, EF, BF, HL, IJ. Conceptualization: CP, YA, EF BF, HL, WHT, IJ, PH, OE. Data curation: CP, YA, OE. Methodology: CP, YA, WHT, IJ, OE. Funding acquisition: OE. Resources: EF, BF, IJ, PH, OE. Project administration: OE. Supervision: OE. Pre-registered report – writing: CP, OE. Pre-registered report – review & editing: all authors.

Acknowledgments

We thank Drs. Eilidh E. MacNicol, Stefano Moia, and Thomas A. W. Bolton for their generosity in reviewing the manuscript and their critical feedback. We thank Drs. César Caballero-Gaudés and Stefano Moia for their guidance in setting up the physiological recording infrastructure, and in particular, the gas analyzer. We thank Dr. Garikoitz Lerma-Usabiaga for the recording of the 'Mundaka clip' for the RSfMRI. We thank the Laboratory for Investigative Neurophysiology (The LINE Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland), the Sense Innovation and Research Center (Lausanne and Sion, Switzerland), and Prof. Micah M. Murray for generously lending the eye-tracking equipment. We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole Polytechnique Fédérale de Lausanne (EPFL), University of Geneva (UNIGE), and Geneva University Hospitals (HUG). This project is supported by the Swiss National Science Foundation —SNSF— (#185872, OE; #185897, PH; #194260, IJ; #182602, HL). This project also receives funding from NIMH (RF1MH12186, OE), and from CZI (EOSS5/'NiPreps', OE).

Table 4. 'Reliability' and 'standard' imaging protocols across three 3.0 T scanners. The description of protocols found in Table 3 is expanded with the following detailed specifications of the MR sequences. These parameters will be optimized when corresponding during the piloting phase of the study (see section Study timeline). All the scans and the results informing the final choice of the parameter will be comprehensively reported at Stage 2.

Reliability Imaging Protocol (36 sessions × one scanner)	**Standard Imaging Protocol (12 sessions × three scanners)**
Anatomical MRI	
T1w MPRAGE, sagittal, 246 slices, matrix 320×320, 0.8×0.8×0.8 [mm^3] resolution, FOV=256 mm, FA=9 °, GRAPPA factor 2, TE/TR/TI = 2.20/1,760/900 ms.	T1w — same parameters as in the reliability protocol
T2w T2-SPACE, sagittal, 176 slices, matrix 256×256, 0.5×0.5×0.9 [mm^3] resolution, FOV=240 mm, FA variable, CAIPIRINHA factor 4, TE/TR/TI = 386/4,900/1,580 ms.	T2w — same parameters as in the reliability protocol
Diffusion MRI (dMRI)	
EPI DWI, 279 directions (12×*b*=700; 40×*b*=1,000; 90×*b*=2,000; 132×*b*=3,000; 6×*b*=0 [s/mm^2]; *b*=0 evenly interleaved, matching the 140 orientations and their antipodal directions of the standard protocol), 1.6×1.6×1.6 [mm^3] resolution, axial, 87 slices, FOV=239 mm, 146×146 matrix, TR/TE=7,800/98 ms, GRAPPA factor 2, SMS factor 3	EPI DWI, 140 directions (6×*b*=700; 20×*b*=1,000; 45×*b*=2,000; 66×*b*=3,000; 3×*b*=0 [s/mm^2]; *b*=0 evenly interleaved), 2.0×2.0×2.0 [mm^3] resolution, axial, 72 slices, FOV=240 mm, 120×120 matrix, TR/TE=7,800/92 ms, GRAPPA factor 2, no SMS acceleration
Functional MRI (fMRI)	
EPI BOLD TR=1,600 ms, TE=(12.60/33.04/53.48/73.92) ms, FA=64°, 2.2×2.2×2.2 [mm^3] resolution, distance factor 0%, 60 slices, 96×96 matrix, FOV=211 mm, GRAPPA factor 2, SMS factor 4	EPI BOLD TR=2,000 ms, TE=33.04 ms, FA=63°, 2.2×2.2×2.2 [mm^3] resolution, distance factor 0%, 60 slices, 96×96 matrix, FOV=211 mm, GRAPPA factor 2 [**Only BH07060**] EPI BOLD TR=1,600 ms, TE=(12.60/33.04/53.48/73.92) ms, FA=64°, 2.2×2.2×2.2 [mm^3] resolution, distance factor 0%, 60 slices, 96×96 matrix, FOV=211 mm, GRAPPA factor 2, SMS factor 4
B_0 field mapping	
GRE, 72 slices, TE=(4.45; 6.91) ms, TR=324 ms, matrix 120×120, 2.0×2.0×2.0 [mm^3] resolution, FOV=239 mm, FA=55 °	GRE — same parameters as in the reliability protocol
EPI DWI — 6×*b*=0 s/mm^2, otherwise same parameters as the (reliability) dMRI acquisition	EPI DWI — 6×*b*=0 s/mm^2, otherwise same parameters as the (standard) dMRI acquisition
EPI BOLD 4 measurements, TR=3,808 ms, TE=35.00 ms, FA=64°, 2.2×2.2×2.2 [mm^3] resolution, distance factor 0%, 60 slices, 96×96 matrix, FOV=211 mm, GRAPPA factor 2	EPI BOLD — same parameters as in the reliability protocol
Physiological recordings and eye-tracking	
Eye-tracking, ECG, respiration belt, CO_2 concentration	n/a — No simultaneous recordings

Abbreviations: BOLD: blood-oxygen level-dependent; CAIPIRINHA: controlled aliasing in parallel imaging results in higher acceleration; DWI: diffusion-weighted imaging; ECG: electrocardiogram; EPI: echo-planar imaging; FA: flip angle; FOV: field-of-view; GRAPPA: GeneRalized Autocalibrating Partial Parallel Acquisition; GRE: gradient-recalled echo; MPRAGE: magnetization-prepared rapid gradient-echo; PE: phase-encoding; SMS: simultaneous multi-slice; TE: echo time; TI: inversion time; TR: repetition time; T2-SPACE: T2- sampling perfection with application-optimized contrasts using different flip angle evolution; T1w: T_1-weighted; T2w: T_2-weighted.