

# SCIENTIFIC DATA




**OPEN**  
DATA DESCRIPTOR

## The Cuban Human Brain Mapping Project, a young and middle age population-based EEG, MRI, and cognition dataset

Pedro A. Valdes-Sosa  <sup>1,2,4</sup>✉, Lidice Galan-Garcia <sup>2,4</sup>, Jorge Bosch-Bayard <sup>1,2,3,4</sup>, Maria L. Bringas-Vega  <sup>1,2,4</sup>, Eduardo Aubert-Vazquez <sup>2,4</sup>, Iris Rodriguez-Gil <sup>2</sup>, Samir Das <sup>3</sup>, Cecile Madjar <sup>3</sup>, Trinidad Virues-Alba <sup>2</sup>, Zia Mohades <sup>3</sup>, Leigh C. MacIntyre <sup>3</sup>, Christine Rogers <sup>3</sup>, Shawn Brown <sup>3</sup>, Lourdes Valdes-Urrutia <sup>2</sup>, Alan C. Evans  <sup>3,5</sup> & Mitchell J. Valdes-Sosa <sup>1,2,5</sup>

The Cuban Human Brain Mapping Project (CHBMP) repository is an open multimodal neuroimaging and cognitive dataset from 282 young and middle age healthy participants ( $31.9 \pm 9.3$  years, age range 18–68 years). This dataset was acquired from 2004 to 2008 as a subset of a larger stratified random sample of 2,019 participants from La Lisa municipality in La Habana, Cuba. The exclusion criteria included the presence of disease or brain dysfunctions. Participant data that is being shared comprises i) high-density (64–120 channels) resting-state electroencephalograms (EEG), ii) magnetic resonance images (MRI), iii) psychological tests (MMSE, WAIS-III, computerized go-no go reaction time), as well as iv.) demographic information (age, gender, education, ethnicity, handedness, and weight). The EEG data contains recordings with at least 30 minutes in duration including the following conditions: eyes closed, eyes open, hyperventilation, and subsequent recovery. The MRI consists of anatomical T1 as well as diffusion-weighted (DWI) images acquired on a 1.5 Tesla system. The dataset presented here is hosted by Synapse.org and available at <https://chbmp-open.loris.ca>.

### Background & Summary

In the past decade several neuroimaging databases (ADNI, HCP, UK Biobank, CAMCAN, ABCD, PPMI), as well as consortia (ENIGMA), have been launched. They aim to accelerate insights into neurodevelopment and physiopathology and to allow the identification of new biomarkers. An essential ingredient, lacking in many projects, is the inclusion of the electroencephalograms (EEG), an informative and direct measurement of brain activity. The EEG is cost-effective, accessible, and applicable to underserved populations throughout the world. It is not surprising, therefore, that interest in EEG has increased, and this modality has been included in new multimodal neuroimaging datasets, such as the mind-brain-body dataset CMI (Babayan *et al.*<sup>1</sup>) and the open resource for transdiagnostic research in pediatric mental health LEMON (Alexander *et al.*)<sup>1,2</sup>.

Here we present a new multimodal neuroimaging dataset<sup>3</sup> that includes, this time information from a Latin American middle-income country. It was led by the Cuban Human Brain Mapping Project (CHBMP)<sup>4</sup>, a population-based, multi-decade longitudinal brain health data-gathering effort in Havana Cuba. This ongoing project is organized by the Cuban Ministry of Public Health (MINSAP) and coordinated by the Cuban Neuroscience Center (CNEURO). The CHBMP focuses on the development of tools and health applications based on multimodal neuroimaging. This data corresponds to a young to middle-age adult population. Due to an

<sup>1</sup>The Clinical Hospital of Chengdu Brain Sciences, University of Electronic Sciences and Technology of China, Chengdu, China. <sup>2</sup>Cuban Neuroscience Center, La Habana, Cuba. <sup>3</sup>McGill Centre for Integrative Neurosciences MCIN. Ludmer Centre for Mental Health. Montreal Neurological Institute, McGill University, Montreal, Canada. <sup>4</sup>These authors contributed equally: Pedro A. Valdes-Sosa, Lidice Galan-Garcia, Jorge Bosch-Bayard, Maria L. Bringas-Vega, Eduardo Aubert-Vazquez. <sup>5</sup>These authors jointly supervised this work: Alan C. Evans, Mitchell J. Valdes-Sosa. ✉e-mail: [pedro.valdes@neuroinformatics-collaboratory.org](mailto:pedro.valdes@neuroinformatics-collaboratory.org)

Age Interval	Number of participants	%
10 < x < = 20	34	12.05
20 < x < = 30	90	31.91
30 < x < = 40	116	41.13
40 < x < = 50	29	10.28
50 < x < = 60	10	3.54
60 < x < = 70	3	1.06

**Table 1.** Distribution of participants per age interval.

official request from the Ministry of Health of Cuba (for a special study) the sampling of males according to the randomized sampling plan continued longer than that of females, resulting in a smaller number of the latter. This gender imbalance will be corrected in ongoing studies. Also two subsequent studies comprising childhood and elderly are in progress to answer scientific questions related to neurodevelopment and healthy aging. The results of these studies will be reported in due course.

The CHBMP has focused on quantitative evaluation of the EEG or “qEEG”<sup>5</sup>. qEEG helps identify brain disorders and is therefore a potentially valuable screening tool. In qEEG, the scalp recorded EEG log spectra of a proband is compared with normative log spectra using statistical parametric mapping procedure. These normative log spectra are age-dependent means and standard deviations, obtained by analyzing a large sample of healthy participants over a wide age range. An EEG normative database is thus a prerequisite for qEEG.

The need for a Cuban normative database for qEEG<sup>6</sup> thus prompted the first wave of the CHBMP, initiated in 1988. It included 211 healthy persons (ages 5 to 97 years)<sup>7</sup>. Participants were randomly selected from the population and screened by the Family Doctor system to include only healthy participants. This database was used to develop a high-resolution qEEG validated by the Cuban Health system<sup>8,9</sup>. It also prompted the development of qEEG for sources (qEEGt). Due to the unavailability of MRI in Cuba at this time, an “approximate qEEGt” was based on the average head model developed by the ICBM consortium<sup>10,11</sup>. Procedures to apply qEEG and qEEGt processing based on this dataset have been integrated into CBRAIN<sup>12</sup>.

However, average-brain template-based qEEGt does not deal adequately with individual cortical anatomy. Thus qEEGt based on each individual’s MRI, as in<sup>13</sup>, is important to validate approximate qEEG but also for integrative multimodal neuroimaging studies. This motivated the second wave of the CHBMP (from 2004 to 2008) as one of the projects for the National Program for Disability.

As in the first wave of the CHBMP, the participants were recruited from the general population using stratified randomized sampling. This yielded 2,019 candidates who were screened, Family Nurses and Doctors, and extensive clinical, neurological, psychological, and neuroimaging evaluations to exclude participants with brain disorders, addictive habits, or exclusionary health conditions, resulting in a final sample of 282 “functionally healthy” participants. The recording protocol included high-resolution EEG, T1, MRI, DWI, and psychological tests such as MMSE, Wechsler Adult Intelligence Scale (WAIS-III), computerized reaction time, as well as the collection of blood samples for a genome-wide association study (GWAS) to be described in a separate further publication.

Making this dataset open is part of the Cuba-Canada-China (CCC) and the Global Brain Consortium (GBC) strategy to integrate EEG neuroimaging as an essential component of multimodal Neuroimaging, and to serve as a “translational bridge” for resource-limited scenarios. This is possible by integrating CHBMP efforts into the MNI neuroinformatics ecosystem<sup>14</sup>, based on the CBRAIN processing portal<sup>15</sup> for the processing modules and the use of the LORIS database system for data storage and open access.

## Methods

**Participants.** The Cuban Human Brain Mapping Project second-wave database contains neuroimaging, medical and cognitive data from, 282 “functionally healthy” participants from the general Cuban population between ages 18–68, ( $31.9 \pm 9.30$ ), comprising 87 ( $36.5 \pm 10.43$ ) females and 195 ( $29.9 \pm 7.97$ ) males. Details of the number of participants in each interval of age is presented in Table 1.

Details of the demographically variable frequency of gender, self-referenced handedness, and educational level are presented in Table 2, and stored in the file **Handedness.csv** and **Demographic\_data.csv**, which also included the weight (lb) of the participants.

**Ethics.** This study was carried out following the Declaration of Helsinki<sup>16</sup>. The experimental protocols were approved by the Ethical Review Committee of the Cuban Neuroscience Center using the guidelines of the Public Health (MINSAF from Spanish) and the Science, Technology, and Environment Ministries of the Republic of Cuba (CITMA from Spanish). Participants included in the study signed a written consent after a house visit by the Primary Health care nurse who gave a verbal and written explanation of the purpose, risks, and benefits of the study. Participants also were informed about the confidentiality of their personal information as well as full access to the best diagnostic and therapeutic procedures in case any kind of illness were detected that might prevent them to form part of the normative dataset. Additionally, they were verbally informed about their right to obtain any clinical, psychological, and neuroimaging results. Finally, participants were informed about further publications which would result from the project, with the guarantee of anonymization and control of the privacy of their personal information.

**Recruitment and exclusion criteria.** The recruitment procedure is summarized in Table 3 and subsequently detailed.

	Categories	N (%)
Gender	Female	87 (30.85%)
	Male	195 (69.14%)
Handedness by preference	Right	238 (84.39%)
	Left	26 (9.21%)
	Ambidextrous	8 (2.38%)
	Missing	10 (3.54%)
Education level	Primary school	16 (5.67%)
	Secondary School	72 (25.53%)
	High school	145 (51.41%)
	University	39 (13.82%)
	Missing	10 (3.54%)

**Table 2.** Demographic description of the sample.

	Activity	Sample size
Stage 1	Selection of one municipality with the closest match in gender, age, ethnic characteristics of the general Cuban population. This was also a municipality with a large proportion of persons originally from other provinces according to the MINSAP and Statistics National Office. La Lisa municipality was selected. A random sample of 30,000 participants from the total population was selected based on the ID card serial numbers.	30,000
Stage 2	From the 30,000 participants selected in the previous stage, a random subsample of 2019 was selected stratified by age, sex, ethnic origin, educational level, and socioeconomic status.	2019
Stage 3	The Doctor and nurses of the Family Program evaluated the willingness to participate and any health conditions of the candidates from the subsample (N = 2019). In detail: 1) The Family nurses visited each household to provide both printed and verbal information about the project objectives and enquired about the preliminary willingness to participate in the study. 2) Those willing to participate attended an interview at their local Polyclinic center for further explanation and signed the Informed Consent. 3) The Health Questionnaire was then applied by the nurse to carry out a further screening of pathology, cognitive complaints, use of pharmaceutical agents, heavy smoking, etc. 4) According to the exclusion criteria, Table 3, a further selection of the subsample was carried out for the next stage. Participants with health problems remained in the overall study, and continue to be followed up to this day with further clinical evaluation and treatment. However, they were excluded from the normative part of the project.	1439
Stage 4	At the local Polyclinic center psychiatrist/neurologist applied: • A standardized and structured psychiatry tool, the MINI International Interview (MINI) • MINIMENTAL State Evaluation (MMSE) • The Edinburg Handedness Questionnaire to classify dexterity A psychometrist applied the: • The Wechsler Adult Intelligence Scale (WAIS) III	530
Stage 5	At the local Polyclinic center, blood samples for genetic studies were extracted and additional measures: the blood pressure, body temperature, frequency and heart rate, body weight and height were obtained (data not included in this version of the database). The weight in pounds was included in demographic file. Digital EEG recording was the final step at the designated Polyclinics	530
Stage 6	MRI collection was carried out at the CIMEQ Hospital followed by an MRI evaluation by a neuroradiologist. Due to a request of the Ministry of Health the number of males selected at this stage was more than double of females.	394

**Table 3.** General procedure for recruitment.

**Stage 1:** In agreement with the Ministry of Public Health (MINSAP), and due to logistical constraints, a single municipality of the province of Havana City was selected for the study, together with the whole structure of the Family Doctor and Polyclinic Centers. Towards this end, a study was carried out with a committee from MINSAP and the National Office for Population Studies to assess the distribution of the following variables of all inhabitants in every municipality in the Province of Havana City: ethnicity, sex, province of origin, and socioeconomic status. Based on these distributions, the Municipality of La Lisa, was selected for the study since it had the closest match to the general Cuban population. A sample of 30,000 inhabitants in this region was randomly selected from the National Identity Card registry.

**Stage 2:** From the original roster of 30,000, a random subsample of N = 2019 was then selected for further processing, is stratified by age, gender, socio-economic status.

Family Doctors (**Stage 3**) then examined the participant's records to exclude persons whom they already had ascertained to have health issues. All the remaining participants were visited by the Family nurses who left a printed description of the project and gave a detailed verbal explanation of its aims. As usual for population studies in Cuba, it was explained that there would be no payment for the study, but if a participant needed to be absent from the workplace, the local government guaranteed this as a fully paid day. They were also informed about all data acquisition protocols as well as safety measures with a special focus on MRI acquisition and safety. To a great degree, the success of this project was due to the close contact of the Family Doctor and Nurse with the local population, as well as the abundant information provided from the media to the general public about the Cuban Neuroscience Center and its project. This explains a 93% initial willingness to participate in the project. For those participants that gave written consent, a health questionnaire was applied for further screening, and

Criteria	Description
Medical conditions	Malignant systemic disease requiring chemotherapy, diabetes, thyroid dysfunction (hyper or hypothyroidism), rheumatic disease, muscular dystrophies, liver cirrhosis, sickle cell anemia, Wilson's disease, lupus erythematosus (SLE), malnutrition, cardiovascular diseases, arterial hypertension, infectious diseases, AIDS, respiratory diseases, history of reactions to medications with hypersensitivity type I, current pregnancy or lactation, permanent metal appliances, cardiac pacemakers or other types of metal in any part of the body (fixed prosthesis, fragments of a bullet, etc.)
Neurological disease	Chronic systemic diseases of the Central Nervous System (CNS), seizures or other attacks, malignant expansive processes of the SNC-radiotherapy, more than one loss of consciousness, cerebrovascular accidents, transient cerebral ischemia, migraine or frequent headaches, severe neuropathies, history of cranial trauma with or without loss of consciousness.
Psychiatric Disorder	History of previous psychiatric treatments, tics, stuttering, intellectual disability, suicide attempt, depression, sociopathic behavior, anxiety disorders, panic attacks, schizophrenia, obsessive-compulsive disorder, drug use and abuse, alcoholism, hyperactivity and attention deficit, dementia.
Prenatal and perinatal antecedents	Difficult pregnancy as indicated by the obstetrician, preterm delivery, gestational hypertension, diabetes, obesity, early membrane rupture in pregnancy, placenta previa, retroplacental hematoma, other infections.
Sleep disorders	Nocturnal terrors, somnambulism, and others.
Familiar pathological background	Epilepsy, neurodegenerative diseases, multiple sclerosis, Wilson's disease, schizophrenia, manic depressive disorder.
Drug addiction	Alcohol abuse according to DSM-IV definition, non-legal drugs, smoker of more than 12 cigarettes per day, drugs affecting the CNS, direct and chronic exposure to toxic substances such as pesticides, heavy metals, phosphorus, and organic solvents.
Neurological physical examination	Any abnormality in the neurological physical examination (hypertonia, hypotonia, asymmetry of reflexes, decrease in visual acuity, nystagmus, etc.).

**Table 4.** Exclusion Criteria.

Cognitive	WAIS below 70 MMSE below 24
Visual EEG	Abnormalities in the background EEG activity characterized by any of the following: a) out of range average amplitude or dominant frequency, b) absence of normal modulation patterns, c) inadequate anterior/posterior organization according to age, d) paroxysms, e) lack of reactivity to eye-opening and/or hyperventilation procedures.
MRI	Neuroradiological reports of lesions or atypical findings.

**Table 5.** The exclusion criteria according to cognitive, EEG, and MRI examination.

consequently, 580 persons were excluded at this stage from the normative study. In this, as well as in subsequent stages, all participants that did not continue in the normative study followed a separate workflow to ensure specialized diagnostic and intervention by units of the health system, with the same protocol as those continuing in the study. The exclusion criteria used for this stage are listed in Table 4. The most prevalent health conditions to exclude participants were diagnosed with metabolic syndrome, psychiatric conditions, personal history of severe illnesses, and sensory and motor disabilities.

In Stage 4, participants continuing in the study were examined at the polyclinic by specialists in Neurology and Psychiatry, to rule out chronic diseases (e.g. addictions, including heavy smoking) or any disorders of the nervous system that would invalidate their participation. Neurological examination was performed following the procedure described in the guidelines published by the U.S. Department of Health and Human Services in 1997<sup>17</sup> and the Mini-Mental State Examination<sup>18</sup> (MMSE) for global cognitive screening. The Mini-International Psychiatric Interview was used for psychiatric evaluation<sup>19</sup> (Spanish version) and the Weschler Adult Intelligence Scale (WAIS) III for intelligence.

In Stage 5, EEG recordings were carried out at the polyclinic. Other measures were obtained the same day such as anthropometric (height) and blood pressure. Additional blood samples were extracted, following a joint protocol from CNEURO with the National Center for Medical Genetics [https://www.ecured.cu/Centro\\_Nacional\\_de\\_Gen%C3%A9tica\\_M%C3%A9dica](https://www.ecured.cu/Centro_Nacional_de_Gen%C3%A9tica_M%C3%A9dica) for further genetic studies of the Cuban population (to be published separately).

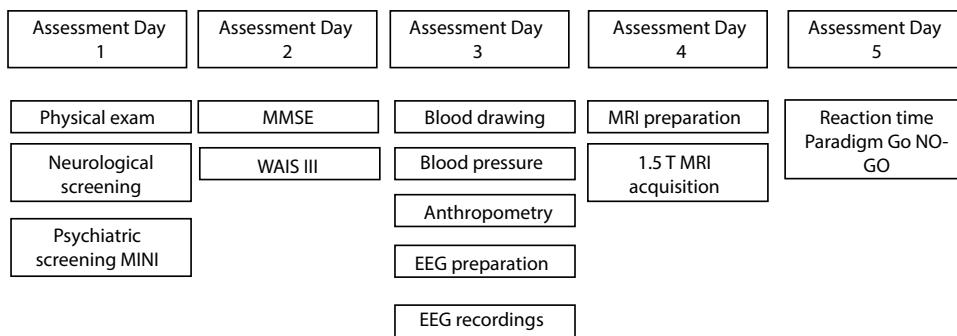
During stages 4 and 5 specific exclusion criteria were employed (see Table 5).

In Stage 6, the MRI studies were carried out at the Center for Medical and Surgical Research la Habana, Cuba (in Spanish: Centro Investigaciones Médico-Quirúrgicas, CIMEQ [https://www.ecured.cu/Centro\\_de\\_Investigaciones\\_M%C3%A9dico\\_Quir%C3%BArgicas\\_](https://www.ecured.cu/Centro_de_Investigaciones_M%C3%A9dico_Quir%C3%BArgicas_)

**(CIMEQ).** Additionally, a psychiatrist/psychometrist applied the computerized Reaction Time test at the end of the study for a subsample (N = 56).

The participants who presented hypertension during this research study were included in a separate study and underwent more specific evaluations such as carotid flow, white matter hyperintensities, eye fund, optic and blood vessel impairments, and a set of extra measurements. The analysis of the results of this hypertension study was partially published in<sup>4</sup>.

**Procedure.** Workflow. Examinations for final participants (Stages 4–6) were carried out in a five-day schedule. See Fig. 1.

**Fig. 1** Flow-chart of 5 days assessment.

Measurement	Number
EEG	250. 170 participants with 64 and 80 with 120 channels.
MRI	203
WAIS	167
MMSE	156
RT	56
EEG + MRI	171
EEG + MRI + WAIS	142
EEG + MRI + MMSE	132

**Table 6.** Number of participants with each type of measurement.

The final sample count for this study ( $N=282$ ) included all the participants who completed all the requisites, after all the steps. The total participants included in each measurement and the conjunction between modalities can be found in Table 6.

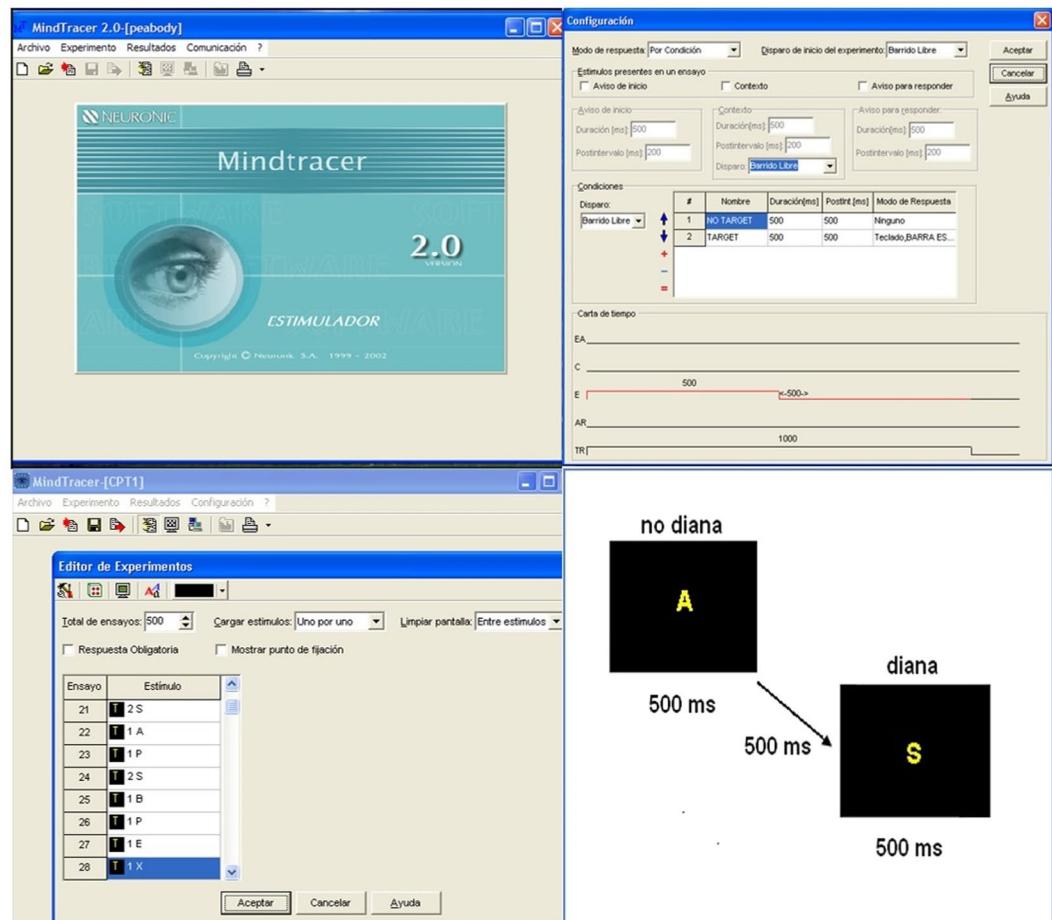
**Psychological tests.** **MMSE.** The Mini-Mental State Examination MMSE<sup>18</sup> is a quick and easy measure of cognitive functioning that has been widely used in clinical evaluation and research involving patients with dementia. In our study, the MMSE was employed as a screening test to exclude participants with cognitive impairment. The total score of the participants is available in the file **MMSE.csv** with also the individual items for 52 subjects.

**Intelligence test.** Intelligence was assessed using the fully validated and translated to the Spanish language version of the David Wechsler Adult Intelligence Scale WAIS-III printed and distributed in Mexico by El Manual Moderno (<https://www.worldcat.org/title/wais-iii-escalaweschler-de-inteligencia-para-adultos-iii/oclc/54053545>). This scale provided scores for a Full-Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ) along with four secondary indices: PO, PS, VC, and WM. The subtests included in each index were as follows: PO: picture completion, block design, matrix reasoning; PS: digit-symbol coding and symbol search; VC: vocabulary, similarities, information, comprehension; and WM: arithmetic, digit span, letter-number sequencing.

The intelligence raw measures were scored according to the official normative data included in the printed version of WAIS-III. However, to avoid cultural bias, they were subsequently standardized with information from the Cuban sample to produce scores of the specific performance, adjusted for age for our population. The results included subtests, the verbal and performance subscales and full scale IQ for all the subjects, available in the **WAIS\_III.csv** file. This dataset was employed in one study about how the white matter (FA-tracts based) predicts fluid and crystallized intelligence<sup>20</sup>.

**Go No-Go test.** For a subset of 56 participants, reaction times were recorded using a Go No-Go paradigm which consisted in a visual attention task, implemented using the psychophysiology software for cognitive stimulation Mindtracer<sup>21</sup> (N\_P-SW 1.3 v.2.1.0.0 Neuronic S.A.).

The task consisted of 500 trials, 25% in GO condition and 75% in NO-GO condition, where the stimuli was a set of letters: P B X E A S. The instructions for the participants were: Simple Reaction Time (RT1) consign: "Press the spacebar when "S" appears at the screen". Complex Reaction time (RT2) consign: "Press the spacebar only when the letter "S" appears preceded by the letter "A". The two tasks were presented to all the participants in the same consecutive order. For a description of the facilities of the software see Fig. 2. The results included in the file **Reaction\_Time.csv** are: mean, standard deviation and skewness of the simple and complex reaction time tasks in milliseconds.



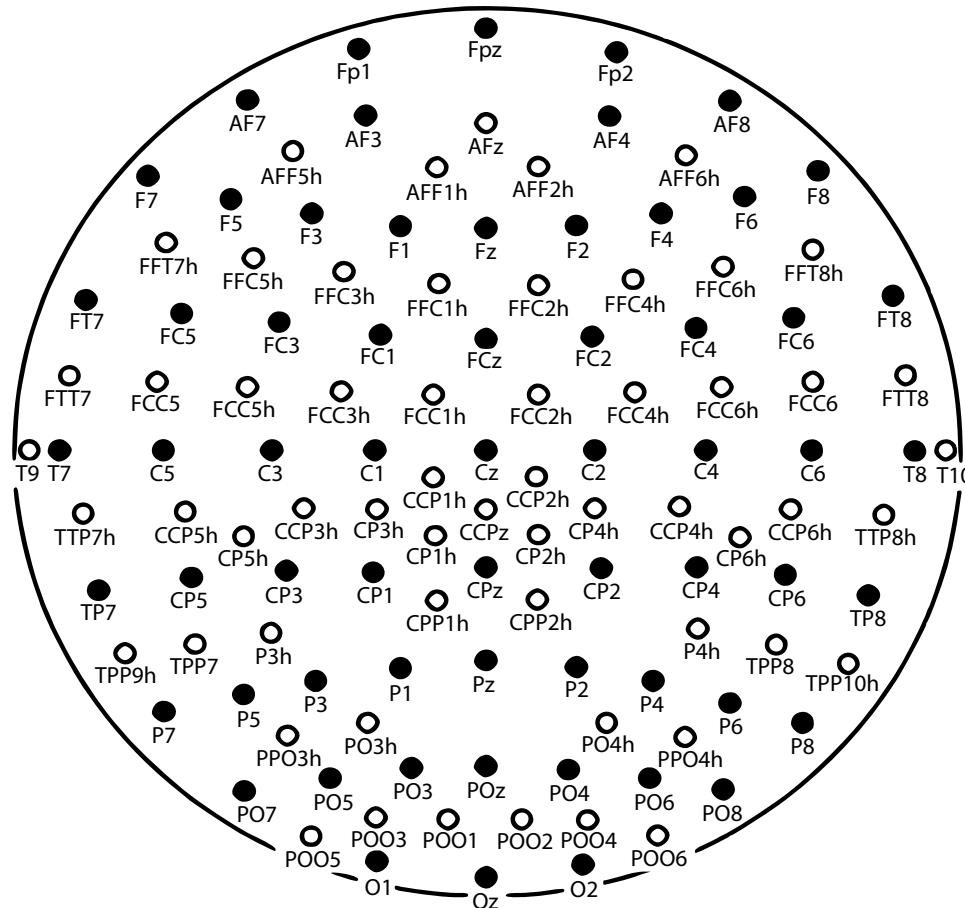
**Fig. 2** Top left: The main display of the software MINDTRACER for the preparation and presentation of the stimuli of the reaction time task. Top right.- Software options to design the experimental task. Bottom left.- Description of the different trials. Bottom right: Design of the presentation of the complex stimuli.

Equipment	MEDICID 5
Number of channels	64/120
Gain	20,000/1000
Impedances	<5 KΩ.
Line Frequency	60 Hz
Amplifiers bandwidth	0.5 to 50 Hz
Sampling frequency	5msec./200 Hz.
Room Temperature	20 to 22°C.

**Table 7.** EEG equipment & recordings technical parameters.

**EEG recordings.** Resting-state EEG was recorded using the digital electroencephalograph system MEDICID 5-with 64 and 128 electrodes with differential amplifiers and gain of 10,000. The amplifiers used three filters: 1) Low cutoff (-3dB, high-pass): first order (6dB/octave) 2) High Cutoff (-3dB, low-pass): Butterworth, second order (12 dB/octave) and 3) Line filter with a unit frequency response.

Electrodes were placed according to the 10-10 International System with a customized electrode cap. Linked earlobes were used as the EEG reference. Electrode impedances were considered acceptable if less than 5 KΩ. The bandpass filter parameters were 0.5–50 Hz and 60 Hz notch, and a sampling period of 200 Hz. The EEG was recorded in a temperature and noise-controlled room while the participant was sitting in a reclined chair. All individuals were asked to relax and remain at rest during the test to minimize artifacts produced by movements and to avoid excessive blinking. The participants received instructions to have enough sleep the previous night, take breakfast, and wash the hair before attending this appointment. See Table 7 for a summary of the technical parameters of the EEG. The structure of raw EEG recording was generated in the default format of the MEDICID neurometrics system (\*.plg extension), which later is converted to standard BIDS format<sup>22</sup>. See the data records section.



**Fig. 3** Schematic representation of the position of the electrode. In black the subset of electrodes employed in the 64 channels (maximum 62 cephalic electrodes). In white, the other electrodes were used for the 120 channels montage (maximum 120 cephalic electrodes). Both configurations use three additional channels which are employed to record the Electro-oculogram and Electrocardiogram.

**Electrode placement.** Two different montages were employed, one with 64 channels and other with 120 channels as illustrated in Fig. 3 with different colors black (64) and white (120) to identify each montage. The nomenclature of the electrodes employed in the MEDICID system and their standardization is included in online-only Table 1 “Nomenclature of electrodes position.xls”

**Description of the EEG protocol comprising the following participant condition:**

1. Baseline: resting-state EEG with closed Eyes (state A), 10 minutes
2. Reactivity: this test consisted of the consecutive opening and closing eyes with an interval of 12 seconds. Open eyes (state B), 5 minutes, where the participant was instructed to look at a point, keeping the pupils fixed.
3. Hyperventilation (HPV): Dividing it in the first minute HPV1 (state C), the subject was instructed to start taking air through the nose and to breathe deeply. The second minute HPV2 (state D), and HPV3 (state E), this last one minute with more frequent and shallow inspirations. Total 3 minutes.
4. Recovery (state F): The last step is the recovery of the patient after the HPV, which lasted around one and a half minutes, but was recorded for 2 full minutes.

Note that the subject's recordings were monitored continuously by the technician, to avoid contamination of the EEG with the electromyogram interference, other changes in the direct current level due to sweating, and also to prevent drowsiness. Any of these artifacts were annotated online by the technician.

Therefore, recordings of at least half an hour were ensured. A design requirement was to have enough valid EEG to carry subsequent frequency domain analysis. For this, quasi-stationary EEG epochs were selected, each consisting of 512-time samples, or 2.56 seconds being marked online continuous EEG recordings. Due to the high density of electrodes, the number of epochs for further analysis was guaranteed to be at least 50 windows for 64 channels and 80 windows for 120 channels (For details on analysis see<sup>12</sup>)

Protocol	Description
T1w	<ul style="list-style-type: none"> <li>Patient position: First Supine Head</li> <li>Sequence: 3D MPRAGE</li> <li>FOV: <math>256 \times 256 \times 160</math> mm</li> <li>Voxel dimensions: <math>1 \times 1 \times 1</math> mm</li> <li>Inversion Time (TI): 1100 ms</li> <li>Repetition Time (TR): 3000 ms</li> <li>Echo Time (TE): 3.93 ms</li> <li>Flip Angle: 15 degrees</li> <li>Series Name: t13d_anatVOL</li> <li>Name of the sequence: tfl3d1_ns brusque and remarkable.</li> </ul>
Field Maps	<ul style="list-style-type: none"> <li>Sequence: 2D Double-echo gradient-echo field map</li> <li>FOV: <math>224 \times 224</math> mm</li> <li>Imaging matrix: <math>64 \times 64</math></li> <li>In-plane resolution: <math>3.5 \times 3.5</math> mm</li> <li>Number of slices: 40</li> <li>Slice orientation: Axial</li> <li>Slice thickness: 3.5 mm</li> <li>Slice gap: 0 mm</li> <li>Repetition Time (TR): 672 ms</li> <li>First Echo Time (TE1): 7.71 ms</li> <li>Second Echo Time (TE2): 12.47 ms</li> <li>Acquisitions: 1</li> <li>Flip Angle: 60 degrees</li> <li>Image type: Magnitude and Phase</li> <li>Name of the Series: fieldmap_pha or fieldmap_mag depending on the type</li> <li>Name of the Sequence: fm2d2r.</li> </ul>
Diffusion-weighted images (DWI) Variant 1	<ul style="list-style-type: none"> <li>Sequence: 2D Single-Shot Gradient Echo Echo-Planar-Imaging (GE-EPI)</li> <li>Parallel acquisition: none</li> <li>Multiband: none</li> <li>FOV: <math>256 \times 256</math> mm</li> <li>Imaging matrix: <math>128 \times 128</math></li> <li>Phase-Encoding direction: AP</li> <li>In-plane resolution: <math>2 \times 2</math> mm</li> <li>Number of slices: 25</li> <li>Slice orientation: Axial</li> <li>Slice thickness: 3 mm</li> <li>Slice gap: 3 mm</li> <li>Repetition Time (TR): 7000 ms</li> <li>Echo Time (TE): 160 ms</li> <li>Flip Angle: 90 degrees</li> <li>Acquisitions: 6</li> <li>Diffusion gradient scheme: Siemens MDDW</li> <li>Number of diffusion gradient directions: 12</li> <li>Echo Spacing: 1.149 ms</li> <li>Readout Time: 145.92 ms</li> <li>Image type: Magnitude</li> <li>Series Name: ep2d_diff12dir</li> <li>Name of the Sequences: ep_b0, ep_b1200 # 0...11</li> </ul> <p>This DWI sequence was repeated in a second run with the same parameters. The only difference was the position of the slices. They were translated parallel to the axis normal to the slice (axial plane) so that the single from the gaps of the first run were acquired in the second run. Consequently, the gaps of the second run occupied the regions from which the signal in the first run was acquired. In this way, the entire brain was covered with a total of 50 slices.</p>
Diffusion-weighted images (DWI) Variant 2	<ul style="list-style-type: none"> <li>Sequence: 2D Single-Shot Gradient Echo Echo-Planar-Imaging (GE-EPI)</li> <li>Parallel acquisition: none</li> <li>Multiband: none</li> <li>FOV: <math>256 \times 256</math> mm</li> <li>Imaging matrix: <math>128 \times 128</math></li> <li>Phase-Encoding direction: AP</li> <li>In-plane resolution: <math>2 \times 2</math> mm</li> <li>Number of slices: 25</li> <li>Slice orientation: Axial</li> <li>Slice thickness: 3 mm</li> <li>Slice gap: 0 mm</li> <li>Repetition Time (TR): 7000 ms</li> <li>Echo Time (TE): 160 ms</li> <li>Flip Angle: 90 degrees</li> <li>Acquisitions: 6</li> <li>Diffusion gradient scheme: Siemens MDDW</li> <li>Number of diffusion gradient directions: 12</li> <li>Echo Spacing: 1.41 ms</li> <li>Image type: Magnitude</li> </ul> <p>This sequence was not repeated.</p>

**Table 8.** MRI Protocol.

**MRI procedure.** MRI: Magnetic resonance imaging (MRI) was performed on a 1.5 Tesla scanner (MAGNETOM Symphony Siemens Erlangen Germany). Over the course of MRI data acquisition, the scanner remained stable and did not undergo any major maintenance or updates which would systematically affect the quality of data provided here. The total measurement time was 45 minutes. See the MRI protocol used in Table 8.

Table 7 summarizes the technical specifications for the acquisition of the MRI. Note that only 203 participants have useful MRI images. Concerning diffusion images, we collected 201 participants with two variants: Variant

1 was performed on 148 subjects and Variant 2 on 53 subjects. The DWI specific parameters can be found in the metadata provided by BIDS for each participant. All DWIs were visually inspected and those which presented either technical or pathological defects were discarded.

Several EEG and MRI studies using this dataset has been published. One such study demonstrated how the cortical surface area could explain the morphological connectivity of brain networks<sup>23</sup>. Other studies explained the substantial inter-individual variability on the neuroanatomical determinants of EEG spectral properties using the DWI-fractional anisotropy<sup>24</sup>. Two papers studied the human brain anatomical network via diffusion-weighted MRI and Graph Theory characterizing brain anatomical connections<sup>25,26</sup>. A general framework for the tensor analysis of single-modality model inversion and multimodal data fusion was presented using our neuroimaging data as an example<sup>27</sup>.

## Data Records

BIDS (Brain Imaging Data Structure) is the new standard for the organization and description of the datasets containing neuroimaging (MRI, MEG, EEG, iEEG, NIRS, PET) and behavioral information<sup>28</sup>.

Based on this BIDS structure, we developed a methodology with the following steps:

1. Anonymization of EEG recordings and MRI scans.
2. Defacing of the MRI scans
3. Conversion of EEG recordings and MRI to BIDS-EEG
4. Validation of the BIDS structure

**Anonymization.** We developed an application (*anomplg.exe*) to erase all the personal information stored in the EEG recordings in plg format, which could facilitate the identification of the participants. This application generated a secured copy of the personal information before its elimination.

The anonymization of the MRI neuroimages was performed using the script Dicat.py V 1.2 <https://github.com/aces/DICAT> developed by MCIN (McGill Center for Integrative Neuroscience) Montreal, Canada. All the data was anonymized with DICAT in 4 folders (ID, Patient\_Name, Sex, Birth\_Date).

**Defacing.** The defacing process consisted in the elimination of the section with the face of the subject inside the anatomical MRI. This prevents the identification of the subject if a posterior 3D rendering is employed with the MRI scan. The software employed was the Mri\_deface V 1.2, del FreeSurfer [https://surfer.nmr.mgh.harvard.edu/fswiki/mri\\_deface](https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface).

**Conversion to BIDS.** *EEG.* We developed an ad-hoc application (*plg2bids.exe*) to read the original EEG recordings in NEURONIC format and convert them into BIDS structure. This application is designed to read either individual EEG recordings or folders with multiple recordings and can update a current BIDS structure with new recordings.

*MRI.* The conversion of MRI neuroimages to BIDS structure was using the Dcm2Bids <https://github.com/cbedetti/Dcm2Bids> which generates the MRI BIDS structure with the original data in format DICOM.

The BIDS-EEG and MRI-BIDS structures were combined in only one structure using a software to combine EEG-MRI-BIDS (*joinbids.exe*).

**Validation.** The final step was the validation of the BIDS structure using the web Bids-validator. <https://bids-standard.github.io/bids-validator/>

The dataset is hosted at Synapse.org, see<sup>1</sup> <https://doi.org/10.7303/syn22324937> and complete access is possible after login in the system. The dataset is also imported in the Longitudinal Online Research and Imaging System (LORIS) v20.2. <https://mcin.ca/technology/loris/>

## Technical Validation

For the quality control of the EEG, the MRI and psychological tests data of this study were implemented in a workflow:

### EEG.

1. The ocular movements and other incidences were annotated by the technician during the recordings.
2. A Board of Certified clinical neurophysiologist reviewed the recorded raw EEG by visual inspection to provide:
  - a. Overall assessment of the EEG recordings quality to determine whether the recording should be repeated.
  - b. Scoring of semi-quantitative scales for abnormalities which could motivate the exclusion of the participant from the normative sample.
  - c. Selection of artifact-free EEG segments useful for further analysis in the continuous recordings.

### MRI.

1. Automatic inspection was performed to check the protocol parameters of the MRI images using in-house software which generate a file with the value of the parameters.
2. Visual inspection by several clinical radiologists to detect abnormalities to decide if the participant should be excluded.

3. As part of the MRI quality control process, several MRI T1 images studies were fixed when a wrap-around artifact (without overlapping with the head) was detected by using the in-house Quality Control MRI app.

### Psychological tests and behavioral information.

- 1) Supervision of the assessment sessions by one board-certified Clinical Neuropsychologist.
- 2) After input to the system, curation of the clinical, psychological, and demographical data was carried by a CNEURO team, assisted by statistical summary tools, to ensure quality control.

### Usage Notes

Each participant was assigned one ID with the structure **CBMxxxx** 3 characters identifying “Cuban Brain Mapping” and four digits indicating the number or order of the participant in the dataset. Note that the code is the same as the different data modalities of the subject. The dataset included BIDS files, the in-house programs, the psychological (WAIS-III, MMSE and reaction time), and the demographic and handedness data (\*.csv) available at Synapse.org. See reference<sup>3</sup>. You can visualize the data at <https://doi.org/10.7303/syn22324937>. To download them you need to be registered at the synapse.org website.

All the datasets have also been stored in the McGill Centre for Integrative Neuroscience (MCIN) network. The dataset will be available by request at <https://chbmp-open.loris.ca>

As mentioned before, due to an official request, there is a gender imbalance in the sample. This should be taken into consideration when doing statistical analysis.

While this data will provide greater ethnic and geographical diversity by providing a large multimodal dataset from a Latin-American setting, it is important to consider differences in EEG and MRI recording equipment. The interindividual variability of EEG characteristics seems to be much higher than country or equipment specific sources of variance<sup>5,29–32</sup>. However as mentioned in the section on EEG recording the characteristics of the EEG amplifiers used are provided to facilitate more detailed analyses of these factors. The issue of site and MRI scanner harmonization must be taken into consideration<sup>33,34</sup>.

We do note that this is a normative sample, selected to be functionally healthy with exclusion of any possible factor impacting brain health. The participants also had the benefit of an adequate life-long diet, free national educational and health system, most subjects with early stimulation programs during their infancy as provided by daycare centers. When comparing this data to other populations this is something to be taken into consideration. In an early multinational study<sup>29</sup> the most important factor predicting deviations from EEG normality was socio economic status. The availability of this dataset can be useful in further studies of this type, now incorporating MRI information.

### Code availability

The software developed by EAV is at <https://github.com/eduardo-aubert/CHBMP-Code> and <https://doi.org/10.7303/syn22324937>. The codes are:

- 1) Anomplg.exe: to anonymize the EEG records, deleting all the personal information stored in the EEG recordings, which could facilitate the identification of the participants.
- 2) Plg2bids.exe: to read the original EEG recordings in NEURONIC format and convert them to BIDS structure.
- 3) Joinbids.exe: to combine BIDS-EEG and MRI-BIDS into only one structure EEG-MRI-BIDS.

Received: 2 September 2020; Accepted: 12 January 2021;

Published online: 05 February 2021

### References

1. Babayan, A. *et al.* Data descriptor: A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Sci. Data* **6**, 1–21 (2019).
2. Alexander, L. M. *et al.* Data Descriptor: An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 1–26 (2017).
3. Valdés-Sosa, P. A. *et al.* CHBMP\_database. *Synapse* <https://doi.org/10.7303/syn22324937> (2020).
4. Hernandez-Gonzalez, G. *et al.* Multimodal quantitative neuroimaging databases and methods: the Cuban Human Brain Mapping Project. *Clin. EEG Neurosci.* **42**, 149–159 (2011).
5. John, E. R. *et al.* Neurometrics. *Science* (80–). **196**, 1393–1410 (1977).
6. Valdés, P. *et al.* QEEG in a Public Health system. *Brain Topogr.* **4**, 259–266 (1992).
7. Bosch-Bayard, J., Galan, L., Aubert Vazquez, E., Virues Alba, T. & Valdes-Sosa, P. A. Resting State Healthy EEG: The First Wave of the Cuban Normative Database. *Front. Neurosci.* **14** (2020).
8. Valdés, P. *et al.* Frequency domain models of the EEG. *Brain Topogr.* **4**, 309–19 (1992).
9. Szava, S. *et al.* High resolution quantitative EEG analysis. *Brain Topogr.* **6**, 211–219 (1994).
10. Evans, A. *et al.* 3D statistical neuroanatomical models from 305 MRI volumes. in *Proceedings of IEEE- Nuclear Science Symposium and Medical Imaging Conference* 1813–1817 (1993).
11. Bosch-Bayard, J. *et al.* 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clin. Electroencephalogr.* **32**, 47–61 (2001).
12. Bosch-Bayard, J. *et al.* A Quantitative EEG Toolbox for the MNI Neuroinformatics Ecosystem: Normative SPM of EEG Source Spectra. *Front. Neuroinform.* **14** (2020).
13. Park, H. J. *et al.* Statistical parametric mapping of LORETA using high density EEG and individual MRI: Application to mismatch negativities in schizophrenia. *Hum. Brain Mapp.* **17**, 168–178 (2002).
14. Das, S., Zijdenbos, A. P., Harlap, J., Vins, D. & Evans, A. C. LORIS: a web-based data management system for multi-center studies. *Front. Neuroinform.* **5**, 1–11 (2012).
15. Sherif, T. *et al.* CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Front. Neuroinform.* **8**, 54 (2014).

16. World Medical Association. World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. *Bull. World Health Organ.* **79**, 373–374 (2001).
17. Department, U. H. and H. *Neurological Single Examination*. (1997).
18. Folstein, M., Robins, N. & Helzer, J. The mini-mental state examination. *Arch. Gen. Psychiatry* **40**, 812 (1983).
19. Lecrubier, Y. *et al.* La Entrevista Neuropsiquiátrica Internacional Reducida (MINI). Una entrevista diagnóstica estructurada breve: fiabilidad y validez según la CIDI. *Eur. Psychiatry Spanish Ed.* **5**, 13–21 (1998).
20. Góngora, D., Jahanshahi, M., Vega-hernández, M., Valdés-sosa, P. A. & Bringas-vega, M. L. Crystallized and fluid intelligence are predicted by microstructure of specific white-matter tracts. *Hum. Brain Mapp.* **1–11**, <https://doi.org/10.1002/hbm.24848> (2019).
21. Borrego Hernández, M., Díaz-Comas Martínez, L. & Bobes León, M. A. Mindtracer 2.0, Sistema De Estimulación Para Estudios Cognitivos. in *BioInformatica2007 (CD)BIO030*. ISBN: 978-959-286-002-5. 1–10 (2007).
22. Pernet, C. R. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. data* **6**, 103 (2019).
23. Sanabria-Díaz, G. *et al.* Surface area and cortical thickness descriptors reveal different attributes of the structural human brain networks. *Neuroimage* **50**, 1497–510 (2010).
24. Valdés-Hernández, P. A. *et al.* White matter architecture rather than cortical surface area correlates with the EEG alpha rhythm. *Neuroimage* **49**, 2328–39 (2010).
25. Iturria-Medina, Y. *et al.* Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *Neuroimage* **36**, 645–660 (2007).
26. Iturria-Medina, Y., Sotero, R. C., Canales-Rodríguez, E. J., Alemán-Gómez, Y. & Melie-García, L. Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. *Neuroimage* **40**, 1064–1076 (2008).
27. Karahan, E., Rojas, P. A., Bringas-vega, M. L., Valdes-Hernandez, P. A. & Valdes-Sosa, P. A. Tensor Analysis and Fusion of Multimodal Brain Images. *Proceeding IEEE* **103** (2015).
28. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
29. Harmony, T. *et al.* EEG maturation on children with different economic and psychosocial characteristics. *Int J Neurosci* **41**, 103–113 (1988).
30. John, E. R., Harmony, T., Prichep, L. & Valdes, P. Analysis of electroencephalographic maturation. in *Machinery of the mind*, Birkhauser, Boston, 360–375 (Birkhauser, Boston, 1990).
31. Koenig, T. *et al.* Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *Neuroimage* **16**, 41–48 (2002).
32. Hu, S., Ngulugulu, A., Bosch-Bayard, J., Bringas-Vega, M. L. & Valdes-Sosa, P. A. Multinational qEEG developmental surfaces. *bioRxiv* (2019).
33. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).
34. Yamashita, A. *et al.* Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS Biol.* **17**, e3000042 (2019).

## Acknowledgements

The second wave of the CHBMP was carried out during the National Program of Disabilities initiated by the Ministry of Public Health in 2004 and directed by Vice-Minister Marcia Cobas Ruiz. Our gratitude to all the Cuban participants who volunteered to participate. Very special thanks to the nurses and the staff of the Polyclinics “Elpidio Berovides” and “Aleida Fernandez” of the municipality of La Lisa, La Habana Cuba who were indispensable to carry out this project. We would like to thank the contributions of principal investigators, researchers, and staff personnel of the Cuban Neuroscience Center and former members of the CHBM project who are not authoring this paper: Pedro A. Valdes Hernandez, Yasser Iturria Medina, Lester Melie-Garcia and Gertrudis Hernandez Gonzalez. This project was carried out in collaboration with the National Center for Clinic Genetic directed by Beatriz Marcheco. The authors would like to thank the support from the National Natural Science Foundation of China (NSFC) projects (81861128001, 61871105, 61673090, and 81330032) and the CNS Program of UESTC (No. Y0301902610100201). This work is part of the Cuba China Canada (CCC) axis platform of neuroinformatics. JBB, SD, CM, ZM, LCM, CR, SB and ACE were supported by Brain Canada (243030), the Fonds de recherche du Québec (FRQ) HBHL FRQ/CCC Axis (246117) and Helmholtz (252428) during the realization of this work. JBB, SD, CM, ZM, LCM, CR, and ACE were also supported by the CFREF/HBHL HIBALL. Special thanks to Ayelen Bosch-Castro who helped with the preparation of the figures.

## Author contributions

P.V.S. was the principal investigator of this project. P.V.S. and M.V.S. with J.B.B. and L.G.G. conceived this project, directed the data collection, and wrote the first version of this article. M.L.B. designed and directed the neuropsychological assessment and the analysis of cognitive data. I.R.G. and E.A.V. were in charge of the cooperation to upload the dataset in the MCIN inside LORIS. E.A.V. did the processing and visualization of neuroimaging data; T.V.A. was in charge of the protocols, recordings, and visual inspection of the EEG; L.V.U. was in charge of logistics and organization of the study. A.C.E. gave the strategic and institutional support of MCIN, and S.D., C.M., Z.M., L.C.M., C.R. and S.B. were the Canadian team in charge of the CBRAIN platform and LORIS software integration of the CHBMP data. P.V.S., L.G.G., J.B.B. and M.L.B. wrote the final version of the paper.

## Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

## Additional information

**Correspondence** and requests for materials should be addressed to P.A.V.-S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021