The Consortium on Vulnerability to Externalising Disorders and Addictions (c-VEDA): an accelerated longitudinal cohort of children and adolescents in India.

Author list: Yuning Zhang Ph.D.[1,2*], Nilakshi Vaidya M.Sc.[1,3*], Udita Iyangar Ph.D.[1], Eesha Sharma M.D.[4], Bharath Holla M.D.[5], Chirag K Ahuja M.D.[6], Gareth J Barker Ph.D.[7], Debasish Basu M.D.[8], Rose Dawn Bharath D.M.[9], Amit Chakrabarti M.D.[10], Sylvane Desrivieres Ph.D.[1], Paul Elliott Ph.D.[11], Gwen Fernandes Ph.D.[12], Amritha Gourisankar B.Sc.[13], Jon Heron Ph.D.[14], Matthew Hickman Ph.D.[15], Preeti Jacob D.M.[4], Sanjeev Jain M.D.[5], Deepak Jayarajan M.D.[5], Kartik Kalyanram M.D.[13], Kamakshi Kartik M.D.[13], Murali Krishna Ph.D.[16], Ghattu Krishnaveni Ph.D.[17], Keshav Kumar Ph.D.[18], Kalyanaraman Kumaran D.M.[17,19], Rebecca Kuriyan Ph.D.[20], Pratima Murthy M.D.[3,5], Dimitri P Orfanos Ph.D.[21], Meera Purushottam Ph.D.[22], Madhavi Rangaswamy Ph.D.[23], Sunita Simon Kupard D.N.B.[24], Lenin Singh M.D.[25], Roshan Singh Ph.D.[25], BN Subodh M.D.[8], Kandavel Thennarasu Ph.D.[26], Mireille Toledano Ph.D.[11], Mathew Varghese M.D.[5], Vivek Benegal M.D.[3*], Gunter Schumann M.D.[1,27*], The c-VEDA consortium[†]

Affiliations:

[1]Centre for Population Neuroscience and Precision Medicine (PONS), MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom;

[2]School of Psychology, University of Southampton, Southampton, United Kingdom;

[3]Centre for Addiction Medicine, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[4]Department of Child & Adolescent Psychiatry, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[5]Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[6]Department of Radiodiagnosis and Imaging, Postgraduate Institute of Medical Education & Research, Chandigarh, India;

[7]Department of Neuroimaging, Institute of Psychology, Psychiatry & Neuroscience, London, King's College London;

[8]Department of Psychiatry, Postgraduate Institute of Medical Education & Research, Chandigarh, India;

[9]Department of Neuroimaging and Interventional Radiology, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[10]ICMR-Regional Occupational Health Centre, Kolkata, India;

[11]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom;

[12]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom;

[13]Rishi Valley Rural Health Centre, Rishi Valley, India;

[14]Centre for Public Health, Bristol Medical School, University of Bristol, Bristol, United Kingdom;

[15]Bristol Medical School, University of Bristol, Bristol, United Kingdom;

[16]Foundation for Research and Advocacy in Mental Health, Mysuru, India;

[17]Epidemiology Research Unit, CSI Holdsworth Memorial Hospital, Mysuru, India;

[18]Department of Mental Health and Clinical Psychology, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[19]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom;

[20]Division of Nutrition, St John's Research Institute, Bengaluru, India;

[21]NeuroSpin, CEA, Université Paris-Saclay, Paris, France;

[22]Molecular Genetics Laboratory, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[23]Department of Psychology, CHRIST (deemed to be university), Bengaluru, India;

[24]Department of Psychiatry & Department of Medical Ethics, St. John's Medical College & Hospital, Bengaluru, India;

[25]Regional Institute of Medical Sciences, Imphal, Manipur, India;

[26]Department of Biostatistics, National Institute of Mental Health and Neurosciences, Bengaluru, India;

[27]Population Neuroscience and Precision Medicine (PONS), LIN-Charite Research Group Dept. of Psychiatry and Psychotherapy, Charite, CCM, Humboldt University, Berlin, Germany and Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China.

*Equally contributing authors; † c-VEDA consortium (www.cveda.org) authors and affiliations are listed in the supplementary materials.

Address correspondence to:

Professor Gunter Schumann, Centre for Population Neuroscience and Stratified Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London; Address: De Crespgny Park, SE5 8AF, London, United Kingdom; Telephone: +44 2078485314; E-mail: gunter.schumann@kcl.ac.uk.

and

Professor Vivek Benegal, Centre for Addiction Medicine, National Institute of Mental Health and Neurosciences, Bengaluru, India, 560029; Telephone: +91 80 2699 5275; Email: vbenegal@gmail.com;

Abstract word count: 193

Main text word count: 3470

Abstract

The global burden of disease attributable to externalising disorders such as alcohol misuse calls urgently for effective prevention and intervention. As our current knowledge is mainly derived from high-income countries such in Europe and North-America, it is difficult to address the wider socio-cultural, psychosocial context, and genetic factors in which risk and resilience are embedded in low- and medium-income countries. c-VEDA was established as the first and largest India-based multi-site cohort investigating the vulnerabilities for the development of externalising disorders, addictions, and other mental health problems. Using a harmonised data collection plan coordinated with multiple cohorts in China, USA, and Europe, baseline data were collected from 7 study sites between November 2016 and May 2019. 9010 participants between the ages of 6 and 23 were assessed during this time, amongst which 1278 participants underwent more intensive assessments including MRI scans. Both waves of follow-ups have started according to the accelerated cohort structure with planned missingness design. Here we present descriptive statistics on several key domains of assessments, and the full baseline dataset will be made accessible for researchers outside the consortium in September 2019. More details can be found on our website [cveda.org].

The need to investigate mechanisms underlying psychopathology in Low- and Medium-Income Countries (LMIC)

The simultaneous acquisition of data measuring behaviour, brain, biology and environment during critical neuro-developmental periods has advanced our knowledge of individual differences in psychiatric vulnerabilities (1-8), which provides important implications in psychiatric nosology and precision psychiatry, for better diagnosis, intervention, and prevention (8, 9). In High Income Countries (HIC), several such initiatives incorporating large sample sizes with longitudinal design and extensive data sharing have been undertaken (9, 10). Standing in contrast is the relatively few such investigations in LMIC such as India. However, the non-shared/unique genetic make-ups, distinct environmental risk constellations, and unique cultural variables that's specific to LMIC may give rise to mechanisms of psychopathology that are distinct from those in HIC, for example gene and environment may present different constitution in psychiatric aetiology in LMIC, or gene environment interaction may have different impact on outcomes, moreover, certain cultural variables (e.g., religion) may result in differential phenotypical representation of biological predispositions.

The Consortium on Vulnerability to Externalizing Disorders and Addictions (c-VEDA) was established to address this major gap. With a focus on the development of externalising disorders and addictions in adolescence, c-VEDA is the first and largest longitudinal study in India that provides a comprehensive characterisation of behaviour and psychopathology, cognition, brain structure and function, the environment, and genomics; it also contributes to the Global Imaging Genetics Initiative in Adolescence (GIGA), a global imaging-genetics consortium that harmonises information across several cohorts by linking genetic, brain, behavioural, and remote sensing satellite data to capture determinants of the physical environment such as urbanisation, pollution, and climate, that may affect mental

health outcomes in children and adolescents across cultures, environments, and ethnic groups (9).

Externalising disorders such as alcohol abuse and dependence contribute substantially to the global burden of disease, and the situation is particularly concerning in LMIC. For example, globally, 10% of mortality between the ages of 15-49 years is attributable to alcohol use (11). While alcohol consumption decreased in recent years in HIC, it is increasing in LMIC, such as India (12), where alcohol-attributable mortality is almost twice the rate of HIC (13). Studies in HIC have identified both environmental (e.g., stressful life events (19), childhood abuse (20)) and genetic factors (e.g., 14-18) that convey risk and resilience for externalizing psychopathology. However, with little evidence from LMIC, it is difficult to establish if environmental and genomic risk factors are similar or distinct in industrialised nations and emerging societies (21). Conducting longitudinal imaging-genetics investigations in LMIC and compare with HIC, can help unravel the complex relationships amongst genetic and neurobiological factors that are socio-cultural-ethnic specific/relevant in externalising psychopathology.

India is in a unique position to tackle many of these scientific challenges. In addition to its distinct genetic make-ups, India has also reported relatively high prevalence rate of externalizing problems. Epidemiological studies found that 15.5% school population reported having externalising disorders (22), 30-35% men and 5% women consume alcohol (23), and the disease burden of externalizing disorders such as alcohol and drug use disorders is estimated to be increased by 25% by 2025 (24). Moreover, rapid economic growth in India has created changes in socio-economic conditions which include nutritional stress (25, 26), pollution (27), widespread socio-economic inequality and vast urbanisation, which are less common in HIC. These environmental risk factors specific to India and other emerging societies may influence trajectories leading to externalising disorders and substance misuse

during childhood and adolescence. As preventive efforts and early interventions mitigate the effects of problems and can be the most cost-effective (28), identifying these trajectories is of particular relevance for public health and prevention.

Description of the c-VEDA cohort

Supported by the Indian Council of Medical Research and the Medical Research Council UK, c-VEDA was established in 2015 as a collaborative effort from 7 Indian (**Figure 1**) and 3 UK research institutions (King's College London, Imperial College, and University of Bristol). There are three major objectives: (i) to enable investigations into the aetiology and life course of externalising disorders by characterising individuals on a great variety of environmental factors (exposome), biological characteristics, and brain structure and function, (**Table 1**); (ii) to enable comparative analysis of behavioural trajectories in childhood and adolescence across multiple cohorts by sharing a set of core assessments (**Table 1**) as well as data acquisition protocols. This sustainable platform includes cohorts from countries such as China (9) and USA (ABCD study; 36) but more specifically with three European cohorts: IMAGEN (29), the study of cognition, adolescents and mobile phones (SCAMP) (30), and the Avon Longitudinal Study of Parents and Children (ALSPAC) (31, Supplementary Material). Together these cohorts maximised comparability with c-VEDA on genomics, neuroimaging, and behavioural data across a wide range of age, e.g., SCAMP's large proportion of participants with south Asian ethnic origin grew up in the UK may enable better differentiation between ethnic/genetic effects and environmental factors; (iii) to generate a large dataset of individuals at baseline within a relatively short period of time using a combination of *accelerated longitudinal design* (32) *and planned missing data design* (33). In specific, we recruited participants within a wide age range (6-23 years old) at baseline, and randomly assigned them to either of the 2 follow-ups, 1 (Follow-up I) or 2 years (Follow-up II) after their baseline assessment (**Figure 1b**). This approach permitted us to

efficiently collect three waves of data that spans a long important developmental period while simultaneously reducing the cost of measurement, and increasing compliance by reducing fatigue from respondents, thus reducing the number of missing data. The study protocol was approved by the ethics committees of the National Institute of Mental Health and Neuroscience (NIMHANS) in Bengaluru, India (Item No. VII, SI. No. 7.08, Behavioural Sciences) and all regional collaborating institutions. The Indo-UK collaboration was approved by the Health Ministry Screening Committee of the Ministry of Health and Family Welfare, Government of India. The study was conducted in accordance with the Declaration of Helsinki (1964 and later versions).

Nine thousand and ten participants were recruited between November 2016 and May 2019 from 7 data acquisition sites in 5 geographical regions (**Figure 1a**). To account for the different socio-cultural and geographical backgrounds, we effortfully recruited participants (1) from both urban and rural areas with an agricultural as well as industrial environment (e.g., coal mining community in Kolkata, rural villages near Kolkata, Imphal, and Rishi Valley), (2) with familial risk for externalizing disorders and addictions (i.e., children of patients with psychiatric diagnosis such as substance use disorders through addiction outpatient units in Bengaluru-NIMHANS and Chandigarh), and (3) with environmental risks such as toxic exposures (coal mines, indoor and outdoor smoke), poor socio-economic status (slum-dwellers near Bengaluru-NIMHANS), and insurgency and inter-ethnic violence (e.g., politically conflicted area near Imphal) (**Figure 1a).** Two to five recruiters per site approached participants, and research purposes and involvement were explained to both the parent(s) and child/adolescent. Informed consent was obtained from parents of those under 18 (assent forms from participants), and participants over the age of 18. Potential participants were excluded if they (1) exceed the 3 recruitment age bands (C1: 6-11 years old; C2: 12-17 years old; C3: 18-23 years old), (2) have extreme physical or mental disability preventing

participation; (3) are blind and/or deaf; (4) have any siblings already enrolled in the study, and (5) have difficulties (e.g., too far from data acquisition centres) or not willing to attend follow-up assessments. Six of the seven sites each recruited a random subsample (total N=1278) for neuroimaging data acquisition using resting state functional magnetic resonance imaging (rsfMRI), diffusion tensor imaging (DTI) and structural MRI (sMRI).

While neuroimaging and genotyping data will be made available in February 2020 (**Appendix 1**), behavioural data was published on the IMAGEN databank on 26th June 2019 (DOI: 10.25720/veda-c13h). Overall and site-specific descriptive is presented in **Table 2**. Of the 9010 participants, 47.8% were boys. Our sample covered a wide range of social class (Caste) and religion: just under half (42.8%) were from the general class (a social group that do not qualify for reservation benefits and other affirmative action schemes operated by the government of India), and 68.3% were Hindu. The majority (68.7%) of participants lived in family-owned houses, and a larger proportion lived in urban areas (54.9%) relative to rural areas (38.3%). Across sites, majority of the participants were from nuclear families (72.2%). Lifetime school enrolment rate was 86.8% (**Table 2**). Demographics across sites showed similar patterns with expected deviations due to the planned recruitment strategies (**Table 2**).

According to the planned missingness design described above, all baseline participants were randomised into two groups (FU-I, FU-II) based on their age, gender, data acquisition site, date of baseline assessment, and MRI participation. Risk groups (e.g., familial psychopathology, adverse experiences) were not taken into account in randomisation because they were not considered as confounding factor, but rather key determinants for phenotypes of investigation. Python script used for randomisation can be found at https://github.com/cveda/cveda_databank/tree/master/follow_up. Participants in each group were invited to attend a telephone assessment one (FU-I) or two years (FU-II) post baseline assessments. Additionally, the neuroimaging subsample were invited to institution-based

assessments using a more extensive assessment battery, alongside MRI scans using the same Standard Operation Procedure (SOP) as baseline (**Table 1**).

FU-I started in November 2017 and FU-II in November 2018. While noting that both waves of follow-ups are still on-going and follow-up rates will change with time, as of June 2019, overall 82.7% (n=2322) from the FU-I group and 68.8% (n=708) from the FU-II group have completed their follow up assessments by telephone.

Description of assessments used in c-VEDA

A detailed list of assessments is outlined in **Table 1**. Wherever possible we used instruments that have been validated across the age groups. All assessments were translated in regionally appropriate languages, and administered using Psytools (https://www.delosis.com/psytools/overview.html).

Environmental measures. We assessed social, familial, and interpersonal environment, which included self-reported psychosocial stressors, family violence, social discrimination, ownership of assets, distance from main road, food security, nutrition and exposure to environmental toxins, and biomass energy use. We have also collected data on migration status and addresses of previous residences, from which remote sensing satellite data can be linked. Other early environmental exposures such as complications during pregnancy and nutrition were also recorded.

Neuroimaging. The neuroimaging subsample was recruited from 6 (out of 7) sites and scanned in 4 scanning centres using five 3T MRI scanners (**Figure 1-a**). The scanning parameters and sequences used in rsfMRI, DTI and sMRI scans can be found in https://cveda.org/standard-operating-procedures/. These were designed to match those in IMAGEN, with minor updates to allow for changes in technology over time (e.g., Phase Encoding polarity techniques for DTI). After the MRI sequences had been frozen, a reference dataset had been chosen for each scanner, and reference parameters extracted from its

DICOM files. Data acquired from each site were all uploaded to a central database after on-site quality control (QC) involving script-based assessment of protocol compliance and artefact profiling. An independent team then compared meta-data in DICOM files of new datasets to the meta-data of the reference dataset to screen for significant deviation bi-weekly, visual inspection of image quality were also performed during this process. Prior to each data release standardised pre-processing was also applied. Detailed QC and pre-processing SOP can be found in each data release alongside imaging data.

Genomics. Standardised acquisition of whole blood was carried out in all participants at baseline, and for the deep phenotyping subsample during in-person follow-ups. All biological materials (plasma, buffy coat, red cells, tempus blood) were processed immediately after acquisition and stored locally short-term, before being transferred and stored long-term centrally at NIMHANS. This biobank allows for DNA and RNA extraction, as well as analysis of proteomic and metabolomics. Blood was chosen for its stability for DNA and RNA extract over long period of time, as well as its suitability for multimodal-omics analyses (e.g., genotype, methylation, gene expression), as well as the comparability of results derived from peripheral blood with other studies investigating behavioural-omics. Acquisition protocols (SOP), including amounts of blood drawn were adapted for each age group can be found in **Appendix 2**.

Neurotoxins. Plasma and urine samples were collected for analysis of environmental neurotoxins (See **Appendix 2** for SOP), in particular plasma lead and urinary arsenic, cotinine (tobacco metabolite) and metabolites of volatile organic compounds as markers of exposure to vehicular and biomass fuel smoke.

Cognition and Behaviour. We characterised a wide range of cognitive measures, such as executive control, emotion recognition, decision making, attention and impulsivity; behaviour and clinical phenotypes were indexed using the Mini International

Neuropsychiatric Interview (MINI) and Strengths and Difficulties Questionnaire (SDQ), as well as an extensive characterisation of behavioural measures related to externalising behaviour and psychopathology including substance use behavioural addictions, eating disorders and mobile phone use (**Table 1**).

Key findings from the baseline study

As the first overview of the study, here we present data on experiences of childhood adversity and psychopathology (**Table 2**). At baseline, overall 46.2% (n=4145) participants had experienced frequent (defined as many times) childhood adversity of any given type, a lower rate compared to 77.7% reported from ALSPAC (35). Overall, the most prevalent type of adversity in c-VEDA was living with an alcohol and drug abuser(s) in the same household (26.1%), while in ALSPAC parental psychopathology topped all childhood adversities (42.7%). Amongst the five types of childhood maltreatment assessed (emotional, physical, sexual abuse, emotional and physical neglect), emotional abuse was the most prevalent overall (9.4%), similar to IMAGEN (5.3%) and ALSPAC (19.3%). The seven study sites showed expected variation in adverse experiences recorded (**Table 2**), partly due to the different recruitment strategies applied in each site. For any adversity experienced, Chandigarh reported the highest rate of adverse experiences (60.3%) amongst all sites, and RV the lowest (26.6%). Five of the seven sites reported living with an alcohol and drug abuser(s) in the same household to be the most prevalent type of adversity (Bengaluru-NIMHANS 89.4%, Mysuru 9.9%, Chandigarh 48.5%, Kolkata 17.8%, and RV 10.6%); emotional neglect was the most prevalent in Mysuru (12.5%), and community violence in Imphal (44.1%). Amongst all types of childhood maltreatment, three sites (Bengaluru-NIMHANS 23.4%, Imphal 10.6%, and RV 3.6%) reported emotional abuse being the most prevalent, whereas emotional neglect was most prevalent in the remaining 4 sites (Bengaluru-SJRI 12.5%, Mysuru 5.0%, Kolkata 5.4%, Chandigarh 6.6%).

At baseline, 3.3% of the participants reported experiencing current major depressive episode, 4.9% reported current anxiety disorders, and 3.5% reported current ADHD (**Table 2**). Current alcohol and substance abuse/dependence were reported by 0.5% and 0.9% of participants respectively. Similar to childhood adversity, sites also reported varied rates in psychopathology (**Table 2**). We visualised behavioural and psychological outcomes assessed using the Strength & Difficulties Questionnaire alongside participants from ALSPAC. c-VEDA participants exhibited a trend of scoring higher in all difficulties and prosocial behaviour before the age 14 (**Figure2**). Variations between those who live in urban versus rural areas, and those with and without experiences of childhood adversities are also presented in Figure 2.

Strengths and Challenges

Strength. The c-VEDA study offers for the first time a comprehensive neurobehavioural characterisation in a LMIC of a large number of children and adolescents, in addition to its inclusive environmental measures. Being the first and largest of its kind in India, c-VEDA can serve as a normative database of Indian children, adolescents and young adults for highly valuable investigations such as genome-wide association studies on psychiatric traits and generation of normative age-specific brain atlases within Indian population. Additionally, as each study site presents uniqueness in their sociocultural and environmental characteristics due to the wide-spread recruitment strategy, c-VEDA also enables direct comparisons between different groups within India, potentially addressing novel research questions.

The study design permits three waves of data collection spanned a wide age to be achieved within short period of time, and the planned missingness design would allow parameters of interest to be estimated without bias. The rich longitudinal dataset based outside Western societies that's comparable with many similar cohorts worldwide (e.g., the

Adolescent Brain Cognitive Development Study (36), and similar initiatives in China (9), IMAGEN, ALSPAC) provides unique opportunities to investigate sociocultural and biological foundation for the manifestation of externalizing disorders and addictions. For example, the investigation of the heterogeneity of developmental trajectories into externalising disorders in HICs and LMICs, and the bio-psycho-social mediators or moderators to these trajectories can shed light to the recently rising field of precision psychiatry and global mental health.

Challenges. While the combination of sequential cohort and planned missing data design renders setting up cohort very efficient, it also poses some statistical challenges. First, the planned missing design randomised baseline participants into two groups to be followed up separately, meaning the traditional two-wave analysis (comparing the same group of people at baseline then follow-ups) would have reduced power. However, a variety of longitudinal models, either within a Structural Equation Modelling or Multilevel Modelling framework, can take full advantage of such data. For example, using a joint model one might examine the longitudinal interplay between alcohol use and antisocial behaviour through adolescence. In addition, these models, through their use of a maximum-likelihood approach to missing data, based on a Missing At Random assumption, can demonstrate a high level of statistical power for a fraction of the monetary and time costs of following all individuals for the whole time period. Notably, there will nevertheless be missing observations in either of the follow-ups that are not “planned” and likely to be missing at random (MAR). The Structural Equation Model Framework offers a number of maximum-likelihood (ML) based alternatives, such as full information ML (FIML) approach, which estimates a likelihood function for each individual based on variables present, and produces unbiased parameter estimates and standard errors (37, 38). Moreover, our study design has been rarely applied to neuroimaging data. Although limited knowledge exists for model fitting using neuroimaging

data of such structure, some suggested that mixed-longitudinal models with an autoregressive covariance structure modelling could be useful (39, 40).

Another challenge is brought about by the relatively low consent/assent rate for MRI participation in the lower age band (6-11 years old) compared to the higher age bands (12-17, 18-23 years old), and the low follow-up rate amongst neuroimaging subsample with particularly low rates amongst younger children, which may pose potential power issues when estimating effect sizes in complicated models. One must consider the missing data mechanism carefully above and beyond the planned missingness design, such as reasons for participants missing certain wave(s) of data or being excluded due to bad data quality, and it is recommended that sensitive analysis should be performed by fitting multiple models and examining the similarities of different estimates (41).

Additionally, what also presents challenges for analysing c-VEDA data is the uniquely complex social/ethnic/religious background in India, which may be intertwined to have an impact on genetic population stratification (42). Study site in our study is potentially a confounding factor that has an influence on both genetics and environmental exposures. Various statistical analysis strategies can be applied to control for population stratification for both overall sample and for each study site, depending on the research question. For example, when investigating environmental influences on brain and behaviour collectively, besides controlling for site effects, one can also first examine environmental influences on brain and behaviour by site or within subgroups of similar genetic background regardless of site, followed up by meta-analysis to examine the overall effect. Alternatively, one can also control for genetic components extracted through e.g., Principal Component Analysis (PCA) to account for population stratification within sites to avoid false positive associations.

Data Access and Study Information

More details on the consortium can be found at www.cveda.org. For data use and collaborations, please contact principal investigators Prof. Gunter Schumann (gunter.schumann@kcl.ac.uk) and Prof. Vivek Benegal (vbenegal@gmail.com) who will review the requests together with the consortium executive committee.

c-VEDA data will be accessible to the wider scientific community in a sustained and secure manner, which offers ways of searching and querying specific data through an anonymised databank structure developed for the IMAGEN (43). Identical data are stored in both NIMHANS and IMAGEN databank. Data quality control SOP and reports on each data release can also be found at www.cveda.org. The full baseline data is expected to be made accessible in February 2020. Follow-ups' data will be made available upon completion of both waves of follow-ups in October 2021, after identical QC and pre-processing procedures to baseline data being carried out.

External researchers are invited to propose projects, which are discussed and approved by the scientific steering committee. For data access and sharing rules, please see **Appendix 1**. Upon approval of project proposals, it is recommended that proposal holders consult or work with a member of the consortium from India, who are more familiar with the socio-cultural specific aspects of the data that may not be familiar to researchers used to HIC data for example.

Funding

c-VEDA is jointly funded by the Indian Council for Medical Research (ICMR/MRC/3/M/2015-NCD-I) and the Newton Grant from the Medical Research Council (MR/N000390/1), United Kingdom.

Conflict of interest: None.

References

1. Bogdan R, Salmeron BJ, Carey CE, Agrawal A, Calhoun VD, Garavan H, et al. Imaging genetics and genomics in psychiatry: a critical review of progress and potential. Biological psychiatry. 2017;82(3):165-75.
2. Pine DS, Ernst M, Leibenluft E. Imaging–genetics applications in child psychiatry. Journal of the American Academy of Child & Adolescent Psychiatry. 2010;49(8):772-82.
3. Viding E, Williamson DE, Hariri AR. Developmental imaging genetics: challenges and promises for translational research. Development and Psychopathology. 2006;18(3):877-92.
4. Meyer-Lindenberg A. Imaging genetics of schizophrenia. Dialogues in clinical neuroscience. 2010;12(4):449.
5. Scharinger C, Rabl U, Sitte HH, Pezawas L. Imaging genetics of mood disorders. Neuroimage. 2010;53(3):810-21.
6. Durston S. Imaging genetics in ADHD. Neuroimage. 2010;53(3):832-8.
7. Domschke K, Dannlowski U. Imaging genetics of anxiety disorders. Neuroimage. 2010;53(3):822-31.
8. Bigos KL, Weinberger DR. Imaging genetics—days of future past. Neuroimage. 2010;53(3):804-9.
9. Schumann G, Benegal V, Yu C, Tao S, Jernigan T, Heinz A, et al. Precision medicine and global mental health. Lancet Glob Health. 2019;7(1):e32.
10. Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. 2014;8(2):153-82.
11. Griswold MG, Fullman N, Hawley C, Arian N, Zimsen SR, Tymeson HD, et al. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2018;392(10152):1015-35.
12. Manthey J, Shield KD, Rylett M, Hasan OS, Probst C, Rehm J. Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. The Lancet. 2019.
13. Organization WH. Global status report on alcohol and health. Geneva: WHO; 2011. World Health Organization Violence Against Women: Intimate partner and sexual. 2013.
14. Stacey D, Bilbao A, Maroteaux M, Jia T, Easton AC, Longueville S, et al. RASGRF2 regulates alcohol-induced reinforcement by influencing mesolimbic dopamine neuron activity and dopamine release. Proceedings of the National Academy of Sciences. 2012;109(51):21128-33.
15. Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proceedings of the National Academy of Sciences. 2011;108(17):7119-24.
16. Dong L, Bilbao A, Laucht M, Henriksson R, Yakovleva T, Ridinger M, et al. Effects of the circadian rhythm gene period 1 (per1) on psychosocial stress-induced alcohol drinking. American Journal of Psychiatry. 2011;168(10):1090-8.
17. Clarke T-K, Laucht M, Ridinger M, Wodarz N, Rietschel M, Maier W, et al. KCNJ6 is associated with adult alcohol dependence and involved in gene× early life stress interactions in adolescent alcohol drinking. Neuropsychopharmacology. 2011;36(6):1142.
18. Blomeyer D, Treutlein J, Esser G, Schmidt MH, Schumann G, Laucht M. Interaction between CRHR1 gene and stressful life events predicts adolescent heavy alcohol use. Biological psychiatry. 2008;63(2):146-51.
19. Quinlan EB, Cattrell A, Jia T, Artiges E, Banaschewski T, Barker G, et al. Psychosocial stress and brain function in adolescent psychopathology. American journal of psychiatry. 2017;174(8):785-94.

20. Barch DM, Belden AC, Tillman R, Whalen D, Luby JL. Early Childhood Adverse Experiences, Inferior Frontal Gyrus Connectivity, and the Trajectory of Externalizing Psychopathology. Journal of the American Academy of Child & Adolescent Psychiatry. 2018;57(3):183-90.
21. Venkata JA, Panicker AS. Prevalence of attention deficit hyperactivity disorder in primary school children. Indian journal of psychiatry. 2013;55(4):338.
22. Pathak R, Sharma RC, Parvan UC, Gupta BP, Ojha RK, Goel N. Behavioural and emotional problems in school going adolescents. Australas Med J. 2011;4(1):15-21.
23. Benegal V, Nayak M, Murthy P, Chandra P, Gururaj G, Obot I. Women and alcohol use in India. Alcohol, gender and drinking problems: Perspectives from low and middle income countries. 2005;89:123.
24. Charlson FJ, Baxter AJ, Cheng HG, Shidhaye R, Whiteford HA. The burden of mental, neurological, and substance use disorders in China and India: a systematic analysis of community representative epidemiological studies. The Lancet. 2016;388(10042):376-89.
25. Khandelwal S, Reddy K. Eliciting a policy response for the rising epidemic of overweight-obesity in I ndia. obesity reviews. 2013;14:114-25.
26. Varadharajan KS, Thomas T, Kurpad AV. Poverty and the state of nutrition in India. Asia Pacific journal of clinical nutrition. 2013;22(3):326-39.
27. Jedrychowski WA, Perera FP, Majewska R, Camman D, Spengler JD, Mroz E, et al. Separate and joint effects of tranplacental and postnatal inhalatory exposure to polycyclic aromatic hydrocarbons: prospective birth cohort study on wheezing events. Pediatric pulmonology. 2014;49(2):162-72.
28. McCrone P, Craig TK, Power P, Garety PA. Cost-effectiveness of an early intervention service for people with psychosis. The British journal of psychiatry. 2010;196(5):377-82.
29. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Buchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry. 2010;15(12):1128-39.
30. Toledano MB, Mutz J, Röösli M, Thomas MS, Dumontheil I, Elliott P. Cohort profile: the study of cognition, adolescents and mobile phones (SCAMP). International journal of epidemiology. 2018;48(1):25-6l.
31. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology. 2013;42(1):111-27.
32. Duncan SC, Duncan TE, Hops H. Analysis of longitudinal data within accelerated longitudinal designs. Psychological Methods. 1996;1(3):236-48.
33. Little TD, Rhemtulla M. Planned missing data designs for developmental researchers. Child Development Perspectives. 2013;7(4):199-204.
34. World Health Organization. Adverse childhood experiences international questionnaire (ACE-IQ)2018 20/05/2019. Available from: http://www.who.int/violence_injury_prevention/violence/activities/adverse_childhood_experiences/en.
35. Houtepen LC, Heron J, Suderman MJ, Tilling K, Howe LD. Adverse childhood experiences in the children of the Avon Longitudinal Study of Parents and Children (ALSPAC). Wellcome open research. 2018;3.
36. Clark DB, Fisher CB, Bookheimer S, Brown SA, Evans JH, Hopfer C, Hudziak J, Montoya I, Murray M, Pfefferbaum A, Yurgelun-Todd D. Biomedical ethics and clinical oversight in multisite observational neuroimaging studies with children and adolescents: the ABCD experience. Developmental cognitive neuroscience. 2018 Aug 1;32:143-54.
37. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. BMC medical research methodology. 2012 Dec;12(1):46.

38. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. Biometrics. 2012 Mar;68(1):129-37.
39. Mills KL, Tamnes CK. Methods and considerations for longitudinal structural brain imaging analysis across development. Dev Cogn Neurosci. 2014;9:172-90.
40. Falk EB, Hyde LW, Mitchell C, Faul J, Gonzalez R, Heitzeg MM, et al. What is a representative brain? Neuroscience meets population science. Proceedings of the National Academy of Sciences. 2013;110(44):17615-22.
41. Matta TH, Flournoy JC, Byrne ML. Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. Developmental cognitive neuroscience. 2018;33:83-98.
42. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461(7263):489.
43. Thyreau B, Schwartz Y, Thirion B, Frouin V, Loth E, Vollstadt-Klein S, et al. Very large fMRI study using the IMAGEN database: sensitivity-specificity and population effect modeling in relation to the underlying anatomy. NeuroImage. 2012;61(1):295-303.

Fig. 1

a c-VEDA study sites.

1 Postgraduate Institute of Medical Education and Research (PGI); 2 Regional Institute of Medical Sciences (RIMS); 3 ICMR-Regional Occupational Health Centre (ROHC); 4 Rishi Valley Rural Health Centre (RVRHC); 5 National Institute of Mental Health and Neurosciences (NIMHANS); 6 St. John's Research Institute (SJRI); 7 CSI Holdsworth Memorial Mission Hospital (HMH); 8 Birth Cohort set up in collaboration with the MRC Life-course Epidemiology Unit, Southampton; * Sites recruited neuroimaging subsample. Site 4, 5, and 6 were all scanned at site 5 (Siemens Skyra, Philips Ingenia). Scanners used in site 1 3, 7 are Siemens Verio, Siemens Verio, and Philips Ingenia, respectively.

b According to the accelerated longitudinal design, c-VEDA recruited participants from a wide age range (6-23 years old) at baseline; the planned missingness design further randomized baseline participants into two groups, the first group would be followed up 1-year post-baseline, and the second group 2 years post-baseline assessment.

Fig. 2

Strength & Difficulties Questionnaire subscale scores by age in c-VEDA versus ALSPAC, and amongst c-VEDA participants, those who live in urban areas versus rural areas, as well as those experienced no childhood adversity defined by the frequent scale of Adverse Childhood Experience Questionnaire, versus those experienced at least one type of childhood adversity For c-VEDA, parental reports were used for participants aged between 6 and 17, and self-reports were used for those between 18 and 23 years of age when generating this graph. ALSPAC used parental report only.

Table 1. c-VEDA measurements by age band (C1=6-11, C2=12-17, C3=18-23 years old) and comparison with IMAGEN, ALSPAC and SCAMP

		c-VEDA						
		Baseline						
DOMAIN/Measurement	**Description**	**C1**	**C2**	**C3**	**Follow-up**	**IMAGEN**	**ALSPAC**	**SCAMP**
ENVIRONMENT								
- Adverse Childhood Experiences-International Questionnaire	Abuse, neglect, violence and any serious household dysfunction	✓	✓	✓	S[1]	✓	✓	
- Alabama Parenting Questionnaire *	Parenting behaviour	✓	✓			✓	✓	
- Environmental Exposure Questionnaire	Energy, drainage, pesticides, insecticides	✓	✓	✓				✓
- Family History Questionnaire	Family history of mental illness	✓	✓	✓	✓	✓	✓	✓
- Indian Family Violence and Control Scale	Abuse experiences of married women at the hands of their partners/marital family	✓	✓	✓			✓	
- Life Event Questionnaire *	Major life events			✓		✓	✓	✓
- Pregnancy History Interview-Revised	Pregnancy history, nutrition, complications	✓	✓	✓		✓	✓	✓
- Parent Bonding Instrument	Parenting style			✓			✓	
- School Climate Questionnaire *	Peers, bullying, school environment	✓	✓	✓			✓	✓
- Socio-demographic Information and Migration questions *	Migration status	✓	✓	✓	✓	✓	✓	
- Short Food Questionnaire-Revised	Food intake & nutrition	✓	✓	✓			✓	
- Usage of digital devices questionnaire adapted from SCAMP	Screen exposure		✓	✓		✓		✓
NEUROIMAGING								
- Structural MRI – T1/T2		S	S	S	S	✓	✓	
- DTI		S	S	S	S	✓	✓	
- Functional MRI – Resting state fMRI		S	S	S	S	✓	✓	
PHYSICAL DEVELOPMENT & HEALTH								
- Anthropometry	Height and Weight; mid arm & head circumference, leg length	✓	✓	✓	S	✓	✓	✓
- Medical Problems Questionnaire	Symptoms and diagnoses of physical conditions	✓	✓	✓			✓	
- Pubertal Development Scale	Pubertal development	✓	✓		S	✓	✓	✓
BIOLOGICAL SAMPLE								
- Blood (or buccal swab)	Genetic information	✓[2]	✓[2]	✓[2]	S	✓		
- Urine	Neurotoxins	✓	✓	✓				✓

		c-VEDA						
		Baseline			Follow-up			
DOMAIN/Measurement	**Description**	**C1**	**C2**	**C3**		**IMAGEN**	**ALSPAC**	**SCAMP**
COGNITION								
- Balloon Analogue Risk-Taking Task	Risk-taking behaviour	✓	✓	✓		✓		
- Corsi Block Tapping Test (PEBL)	Visual-spatial attention & working memory	✓	✓	✓		✓	✓	✓
- Digit Span (PEBL)	Verbal attention & working memory	✓	✓	✓		✓	✓	✓
- Emotional Recognition Task	Emotion recognition	✓	✓	✓			✓	
- 27-item Monetary Choice Questionnaire (Now-or-later test)	Reward processing & decision-making	✓	✓	✓		✓		
- Social Cognition Rating Tools in the Indian Setting	Theory of Mind	✓	✓	✓			✓	
- Stop Signal Task	Response inhibition (Impulse control)	✓	✓	✓		✓	✓	
- Trial Making Test (PEBL)	Visual attention and task shifting	✓	✓	✓		✓	✓	✓
- Wisconsin Card Sorting Test (PEBL)	Cognitive flexibility	✓	✓	✓				
BEHAVIOUR & PSYCHOPATHOLOGY								
- Alcohol, Smoking and Substance Involvement Screening Test *	Substance use and related problems	✓	✓	✓	✓	✓	✓	✓
- Adult Temperament Questionnaire	Temperament and personality			✓		✓	✓	
- Early Adolescent Temperament Questionnaire	Temperament to social-emotional functioning		✓			✓	✓	
- Adolescent Attachment Questionnaire	Attachment to parents; caregiving experience		✓			✓	✓	
- Big Five Personality Test *	Five factors of personality			✓		✓		
- Childhood Behaviour Questionnaire	Behavioural problems	✓					✓	
- Mini-International Neuropsychiatric Interview *	Clinical interview on psychiatric disorders	✓	✓	✓	S			
- Strengths and Difficulties Questionnaire *	Emotional symptoms, conduct problems, hyperactivity/ inattention, peer relationship problems, prosocial behaviour	✓	✓	✓	✓	✓	✓	✓

N.B. * Shared core assessments amongst all GIGA cohorts.[1] S=Neuroimaging subsample; [2] Saliva samples were collected only when blood samples were not available or possible;

Table 2. c-VEDA sample characteristics and descriptive of key variables, overall, and by site.

	Overall	Study sites						
		Bengaluru (NIMHANS)	Bengaluru (SJRI)	Mysore	Imphal	Chandigarh	Kolkata	Rishi Valley
	N=8999	n=1883	n=1018	n=1411	n=1120	n=1267	n=1524	n=776
$N_{\text{neuroimaging subsample}}$				-	-			
Age (years)								
- Range	5.32-24.91	6.00-23.98	5.38-17.45	7.42-24.91	5.93-23.74	5.32-24.06	5.35-24.19	5.32-24.47
- Mean (SD)	14.55 (4.61)	15.19 (4.18)	12.21 (1.66)	18.46 (3.03)	15.22 (4.72)	15.16 (5.13)	12.03 (4.23)	11.91 (4.17)
- Missing [n (%)] [1]	5 (0.06)	3 (0.2)	1 (0.1)	0 (0.0)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)
Sex								
- Female [n (%)]	4699 (52.2)	41.6	72.8	63.3	51.1	32.1	60.1	49.9
- Male [n (%)]	4300 (47.8)	58.4	27.2	36.7	48.9	67.9	39.9	50.1
Caste								
- General [n (%)]	3852 (42.8)	1021 (54.2)	239 (23.5)	555 (39.3)	675 (60.3)	881(69.5)	328 (21.5)	153 (19.7)
- Other Backwards Class [n (%)]	1864 (20.7)	285 (15.1)	252 (24.8)	367 (26.0)	341 (30.4)	63 (5.0)	144 (9.4)	412 (53.1)
- Scheduled Castes [n (%)]	1357 (15.1)	348 (18.5)	212 (20.8)	169 (12.0)	40 (3.6)	88 (6.9)	409 (26.8)	91 (11.7)
- Scheduled Tribes [n (%)]	649 (7.2)	110 (5.8)	51 (5.0)	29 (2.1)	53 (4.7)	4 (0.3)	369 (24.2)	33 (4.3)
- Other [n (%)]	618 (6.9)	55 (2.9)	168 (16.5)	271 (19.2)	1 (0.1)	38 (3.0)	1(0.1)	84 (10.8)
- Missing [1]	659 (7.4)	64 (3.4)	96 (9.4)	20 (1.4)	10 (0.9)	193 (15.2)	273 (17.9)	3 (0.4)
Religion								
- Hindu [n (%)]	6150 (68.3)	1493 (79.3)	608 (59.7)	806 (57.1)	979 (87.4)	612 (48.3)	1003 (65.8)	649 (83.6)
- Muslim [n (%)]	806 (9.0)	105 (5.6)	136 (13.4)	199 (14.1)	12 (1.1)	22 (1.7)	229 (15.0)	103 (13.3)
- Sikh [n (%)]	397 (4.4)	2 (0.1)	0 (0.0)	1 (0.1)	0 (0.0)	384 (30.3)	10 (0.7)	0 (0.0)
- Christian [n (%)]	809 (9.0)	204 (10.8)	154 (15.1)	353 (25.0)	71 (6.3)	9 (0.7)	1 (0.1)	17 (2.2)
- Jain [n (%)]	14 (0.2)	2 (0.1)	7 (0.7)	3 (0.2)	0 (0.0)	2 (0.2)	0 (0.0)	0 (0.0)
- Buddhist [n (%)]	2 (0.01)	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
- Other [n (%)]	203 (2.3)	23 (1.2)	40 (3.9)	34 (2.4)	47 (4.2)	46 (3.6)	9 (0.6)	4 (0.5)
- Missing [n (%)] [1]	618 (6.9)	52 (2.8)	73 (7.2)	15 (1.1)	11 (1.0)	192 (15.2)	272 (17.8)	3 (0.4)
House ownership								
- Family's Own [n (%)]	6183 (68.7)	886 (47.1)	445 (43.7)	1113 (78.9)	1017 (90.8)	869 (68.6)	1211 (79.5)	662(82.7)
- Rented [n (%)]	2201 (24.5)	947 (50.3)	500 (49.1)	283 (20.1)	93 (8.3)	206 (16.3)	41 (2.7)	131 (16.9)
- Missing [n (%)][1]	615 (6.8)	50 (2.7)	73 (7.2)	15 (1.1)	10 (0.9)	192 (15.2)	272 (17.8)	3 (0.4)
House location								
- Rural [n (%)]	3444 (38.3)	329 (17.5)	12 (1.2)	623 (44.2)	594 (53.0)	421 (33.2)	889 (58.3)	576 (74.2)

- Urban [n (%)]	4940 (54.9)	1504 (79.9)	933 (91.7)	773 (54.8)	516 (46.1)	654 (51.6)	363 (23.8)	197 (25.4)
- Missing [n (%)][1]	615 (6.8)	50 (2.7)	73 (7.2)	15 (1.1)	10 (0.9)	192 (15.2)	272 (17.8)	4 (0.4)
Family structure								
- Nuclear (%)	6500 (72.2)	1708 (90.7)	612 (60.1)	1107 (78.5)	860 (76.8)	624 (49.3)	1130 (74.1)	459 (59.1)
- Joined (%)	1825 (20.3)	122 (6.5)	288 (28.3)	288 (20.4)	250 (22.3)	448 (35.4)	115 (7.5)	314 (40.5)
- Other [2] (%)	57 (0.6)	1 (0.1)	45 (4.4)	1 (0.1)	0 (0.0)	3 (0.4)	7 (0.5)	0 (0.0)
- Missing [n (%)][1]	617 (6.9)	52 (2.8)	73 (7.2)	15 (1.1)	10 (0.9)	192 (15.2)	272 (17.8)	3 (0.4)
Life-time school enrolment [3] (%)								
- Yes [n (%)]	7812 (86.8)	1632 (86.7)	933 (91.7)	1333 (94.5)	1053 (94.0)	999 (78.8)	1135 (74.5)	727 (93.7)
- No [n (%)]	562 (6.2)	200 (10.6)	12 (1.2)	63 (4.5)	57 (5.1)	75 (5.9)	109 (7.2)	46 (5.9)
- Missing [n (%)][1]	625 (6.9)	51 (2.7)	73 (7.2)	15 (1.1)	10 (0.9)	193 (15.2)	280 (18.4)	3 (0.4)
Childhood Adverse experience [4]								
- Type of adversities experienced [n (%)]								
o Emotional abuse	850 (9.4)	440 (23.4)	108 (10.6)	47 (3.3)	119 (10.6)	76 (6.0)	52 (3.4)	8 (1.0)
o Physical abuse	540 (6.0)	273 (14.5)	77 (7.6)	32 (2.3)	64 (5.7)	59 (4.7)	27 (1.8)	8 (1.0)
o Contact sexual abuse	58 (0.6)	20 (1.1)	0 (0.0)	6 (0.4)	1 (0.1)	14 (1.1)	17 (1.1)	0 (0.0)
o Parental separation/absence	703 (7.8)	174 (9.2)	70 (6.9)	56 (4.0)	223 (19.9)	62 (4.9)	42 (2.8)	76 (9.8)
o Domestic violence	1381 (15.3)	461 (24.5)	119 (11.7)	71 (5.0)	415 (37.1)	184 (14.5)	65 (4.3)	66 (8.5)
o Emotional neglect	609 (6.8)	132 (7.0)	127 (12.5)	71 (5.0)	84 (7.5)	84 (6.6)	83 (5.4)	28 (3.6)
o Physical neglect	256 (2.8)	98 (5.2)	7 (0.7)	17 (1.2)	43 (3.8)	50 (3.9)	20 (1.3)	21 (2.7)
o Bullying	107 (1.2)	40 (2.1)	2 (0.2)	24 (1.7)	15 (1.3)	23 (1.8)	2 (0.1)	1 (0.1)
o Community violence	947 (10.5)	140 (7.4)	55 (5.4)	119 (8.4)	494 (44.1)	86 (6.8)	18 (1.2)	35 (4.5)
o War/collective violence	23 (0.3)	6 (0.3)	4 (0.4)	1 (0.1)	6 (0.5)	4 (0.3)	1 (0.1)	1 (0.1)
o Alcohol/drug abuser in the household	2349 (26.1)	917 (89.4)	50 (4.9)	139 (9.9)	276 (24.6)	614 (48.5)	271 (17.8)	82 (10.6)
o Household member mental illness	402 (4.5)	131 (7.0)	9 (0.9)	15 (1.1)	99 (8.8)	133 (10.5)	10 (0.7)	5 (0.6)
o Household member imprisonment	141 (1.6)	29 (1.5)	11 (1.1)	5 (0.4)	55 (4.9)	19 (1.5)	11 (0.7)	11 (1.4)
- Number of adversities experienced [n (%)]								
o 0	4149 (46.1)	610 (32.4)	583 (57.3)	1011 (71.7)	317 (28.3)	282 (22.3)	781 (51.2)	565 (72.8)
o 1	2030 (22.6)	370 (19.6)	202 (19.8)	263 (18.6)	299 (26.7)	440 (34.7)	328 (21.5)	128 (16.5)
o 2	991 (11.0)	375 (19.9)	87 (8.5)	68 (4.8)	212 (18.9)	148 (11.7)	55 (3.6)	46 (5.9)
o 3 and more	1126 (12.5)	461 (24.1)	74 (7.3)	52 (3.8)	279 (25.0)	176 (13.9)	51 (3.5)	33 (4.2)
o Missing	703 (7.8)	67 (3.6)	72 (7.1)	17 (1.2)	13 (1.2)	221 (17.4)	309 (20.3)	5 (0.5)
- ACES score [Mean (SD)]	1.03 (1.45)	1.66 (1.73)	0.68 (1.11)	0.44 (0.94)	1.67 (1.64)	1.39 (1.47)	0.53 (0.99)	0.46 (0.99)

Strength and Difficulties Questionnaire [5] [Mean (SD)]								
- Emotional symptoms	3.28 (2.54)	2.31 (2.32)	3.59 (2.53)	3.88 (2.65)	4.12 (2.63)	3.04 (2.60)	3.94 (2.22)	2.40 (2.03)
- Conduct problems	2.46 (1.94)	2.46 (1.89)	2.10 (1.43)	2.61 (2.00)	2.21 (1.58)	2.32 (2.16)	3.59 (2.02)	1.54 (1.74)
- Hyperactivity/inattention	3.67 (2.21)	3.46 (2.11)	4.44 (1.73)	2.82 (1.94)	4.29 (2.30)	3.77 (2.45)	4.49 (1.86)	2.39 (2.23)
- Peer problems	2.52 (1.85)	1.89 (1.89)	3.28 (1.52)	2.79 (1.76)	2.50 (1.73)	2.27 (1.90)	2.97 (1.81)	2.32 (1.85)
- Prosocial behaviour	7.86 (2.37)	7.68 (3.30)	6.51 (1.98)	8.84 (1.59)	8.05 (1.85)	8.44 (1.67)	6.92 (2.05)	8.71 (1.70)
Psychopathology [6]								
- Alcohol abuse and/or dependence, current (%)	45 (0.5)	29 (1.5)	0 (0.0)	3 (0.2)	3 (0.3)	5 (0.4)	3 (0.2)	2 (0.3)
- Substance abuse and/or dependence, current (%)	84 (0.9)	56 (3.0)	0 (0.0)	1 (0.1)	9 (0.8)	13 (1.0)	5 (0.3)	5 (0.6)
- ADHD current (%) [7]	316 (3.5)	154 (8.2)	7 (0.7)	10 (0.7)	19 (1.7)	70 (5.5)	27 (1.8)	29 (3.7)
- Major depressive episode, current (%)	300 (3.3)	62 (3.3)	6 (0.6)	92 (6.5)	36 (3.2)	39 (3.1)	47 (3.1)	18 (2.3)
- Anxiety disorders, current (%)[8]	444 (4.9)	93 (4.9)	40 (3.9)	145 (10.3)	44 (3.9)	57 (4.5)	25 (1.6)	40 (5.2)

N.B. [1] Missing included "refused", "don't know", and system missings; [2] This included: staying with grandparent(s), n=4; staying with relatives (e.g., grandparents, aunt, older sister), n=7; semi-nuclear family, n=1; staying in a hostel, n=20; living in orphanages, n=23; Staying with extended family, n=6; [3] This is assessed using the question "did the subject never enrol/discontinue/drop out of school or college"; [4] Adverse childhood experiences in c-VEDA are measured using the Adverse Childhood Experiences International Questionnaire, the frequency version; [5] Participants aged between 6 and 17 years old used parental report, and 19-23 used self-report; [6] Measured by the Mini International Neuropsychiatric Interview for Children and Adolescent (M.I.N.I KID; C1 and C2 age bands) and the Mini-International Neuropsychiatric Interview (M.I.N.I; C3 age band); [7] Current ADHD in C3 age band is measured by the Adult ADHD Self-Report Scale; [8] Measured using M.I.N.I. KID and M.I.N.I., combining panic disorder-lifetime, panic disorder-limited symptom attacks lifetime, panic disorder current, agoraphobia current without history of panic disorder, panic disorder without agoraphobia current, separation anxiety disorder (M.I.N.I KID only), social anxiety disorder current, obsessive-compulsive disorder current, and generalised anxiety disorder current; SD=Standard deviation.

a. c-VEDA study sites.

Location	Language	Cohort characteristics		
		Urban or Rural	General or high-risk	Main recruitment sources
Chandigarh [1,*]	Punjabi, Hindi	Urban, rural	High risk, general	Hospital de-addiction services, schools, slums
Imphal [2]	Manipuri	Rural	General	Politically conflictual areas, hospitals, schools
Kolkata [3,*]	Bengali, Hindi	Rural	High risk	Eastern Coalfields and local villages
Rishi Valley [4,*]	Telegu	Rural	General	Local villages
Bengaluru [5,*]	Kannada, Tamil	Urban	High risk, general	Addiction outpatient services
Bengaluru [6,*]	Kannada, Tamil	Urban	General	Local schools
Mysore [7,*]	Kannada	Urban	General	Siblings of existing birth cohorts[8], hostels, educational institutions

b. Accelerated longitudinal cohort design with planned missingness.

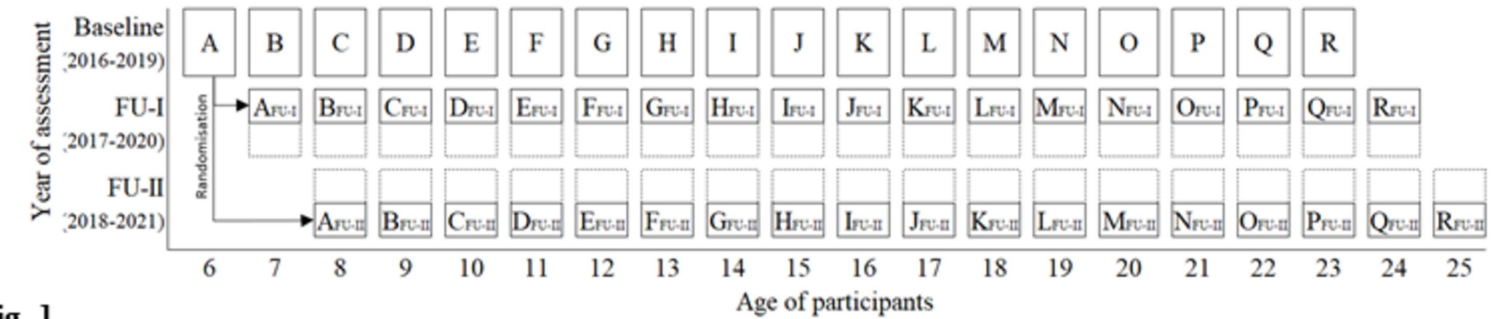

Fig. 1

a [1] Postgraduate Institute of Medical Education and Research (PGI); [2] Regional Institute of Medical Sciences (RIMS); [3] ICMR-Regional Occupational Health Centre (ROHC); [4] Rishi Valley Rural Health Centre (RVRHC); [5] National Institute of Mental Health and Neurosciences (NIMHANS); [6] St. John's Research Institute (SJRI); [7] CSI Holdsworth Memorial Mission Hospital (HMH); [8] Birth Cohort set up in collaboration with the MRC Life-course Epidemiology Unit, Southampton; * Sites recruited neuroimaging subsample. Site 4, 5, and 6 were all scanned at site 5 (Siemens Skyra, Philips Ingenia). Scanners used in site 1 3, 7 are Siemens Verio, Siemens Verio, and Philips Ingenia, respectively. **b** According to the accelerated longitudinal design, c-VEDA recruited participants from a wide age range (6-23 years old) at baseline; the planned missingness design further randomized baseline participants into two groups, the first group would be followed up 1-year post-baseline, and the second group 2 years post-baseline assessment.

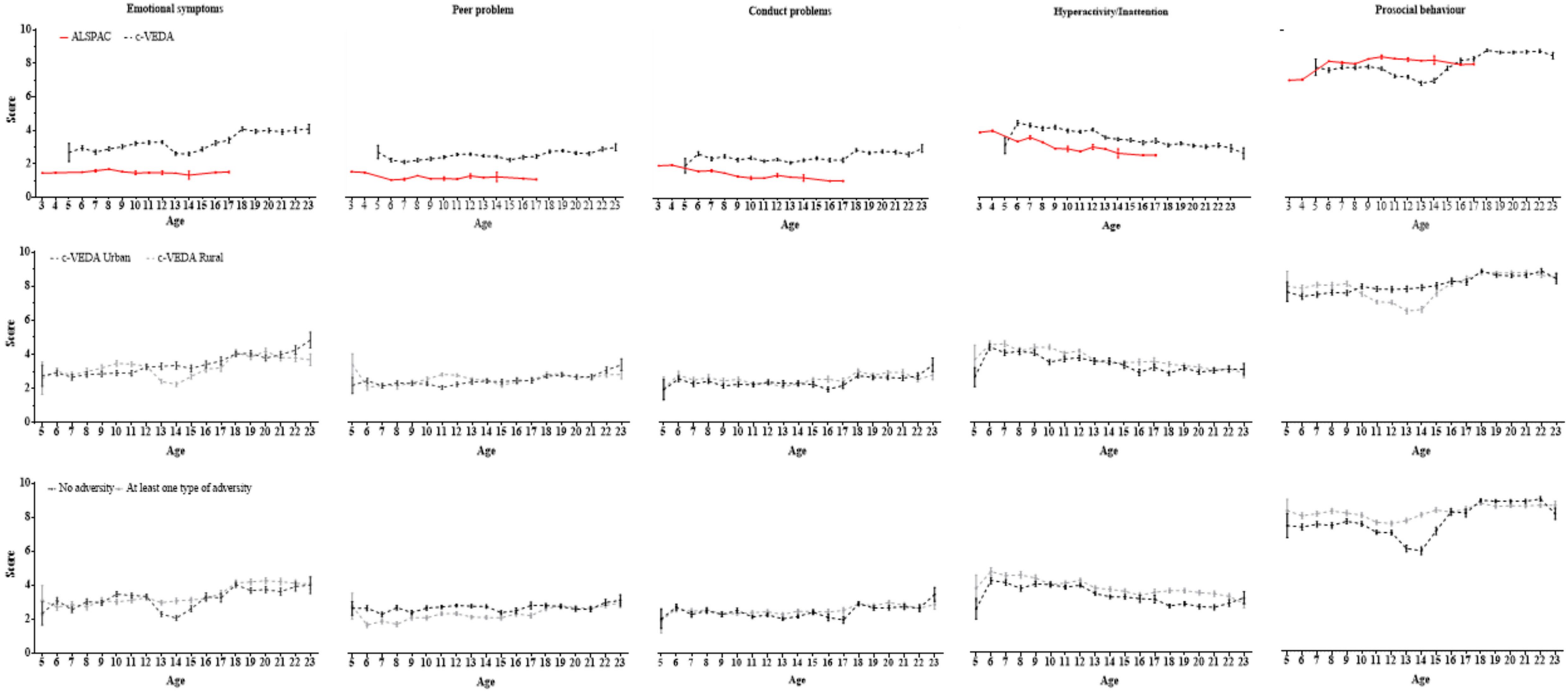

Fig. 2

Strength & Difficulties Questionnaire subscale scores by age in c-VEDA versus ALSPAC, and amongst c-VEDA participants, those who live in urban areas versus rural areas, as well as those experienced no childhood adversity defined by the frequent scale of Adverse Childhood Experience Questionnaire, versus those experienced at least one type of childhood adversity For c-VEDA, parental reports were used for participants aged between 6 and 17, and self-reports were used for those between 18 and 23 years of age when generating this graph. ALSPAC used parental report only.