DATA MANAGEMENT PLAN: CICADA-ME: Coronavirus Intersectionalities: Chronic Conditions and Disabilities and Migrants and other Ethnic minorities

1. Type of study: Mixed methods, with secondary data analysis of existing text sources and cohort and household surveys, a new national UK survey in 3 waves over 15 months, including social network analysis, interviews, participatory co-create workshops.

2. Types of data: Data generated from the project will consist of: a) photo, video and audio digital files; b) digital (RedCap) and paper-based survey; c) consent forms that participants sign; d) content logs and transcripts of audio/video files saved electronically in Word and stored in NVivo and in KWIC and social media analysis software; e) NVivo 'projects'; f) systematic review screening and data extraction records; g) maps drawn by participants and scanned into digital form for analysis; h) numerical data analysed using Python, SPSS and SNA software.

3. Personal data: The focus of this study is on those with physical, cognitive or mental health disabilities, and in the intersectionality of demographic features: gender, socioeconomic status, networks, ethnicity and migrant status as well as health status, some of which are protected features. This means the study will need to take extra measures to protect data whilst noting that survey data will be anonymised unless survey participants are selected using survey logic to be invited for interview. This will require participants to actively opt in for further information which will be sent electronically via their emails together with an interview consent form. Other than this, personal data (email address) is required to send out repeat waves of the survey. The data will be handled entirely via RedCap and will not be used outside these uses. Therefore there are no foreseeable risks to privacy. Without this data, the repeat waves and interviews cannot go ahead.

Personal identifying information will be redacted from project files; the redacted files will link to other related data through unique codes. Depersonalised data will be replaced with descriptive metadata in square brackets, for example a person's name may be replaced by [name of family member].

Recorded audio and visual data cannot be fully de-identified and its sharing will be particularly restricted; it will be treated as personal data and therefore deleted sooner than other research data.

4. Format and scale of the data: Data will be collected from approximately 210 participants for the interviews and workshops, and the aim is for 5,000 for each survey wave but the final total may exceed this. Informed consent will be sought from participants to participate in the study, to have their data permanently archived for unspecified future use in research and education (subject to ethical approval for new research studies using their data), and for selected excerpts to be used in disseminations. Paper-based data will be stored as hard copies and scanned for digital storage also. Anonymised digital media data will be saved as .mpg or WAV format files, anonymised content logs and transcripts as Word files, .sav will be used for SPSS files. Digital images will be stored as JPEGs. A database(s) with .csv output will also be used to enable data analysis, and for archiving for data sharing. These file formats are chosen as they are accepted standards and in widespread use. At project end, Word documents will be converted to plain text and PDF-A and long-term preservation of all data will be carried out in accordance with procedures at UCL.

5. Data quality and standards: Data quality will be ensured through multiple steps. Study and data processing and management, data entry and archiving processes will ensure transparency and consistent operating procedures. The investigators will keep a field record of the settings, conditions, and any issues and unexpected events during data collection in electronic files and hard copy where relevant. Participant consent forms and eligibility criteria will be checked before qualitative data collection. Comprehensive metadata will be entered in the project database for each data item, including specific consents and conditions of use. Quantitative data entry completed by one investigator will be checked by another investigator, who will also re-run all analyses. We will receive feedback from our governance groups (advisory, PPI and steering). Discussion, reflection on learning points and reviews of the work will be undertaken.

6. Managing, storing and curating data: The data will be anonymised and stored electronically in password protected computer folders on secure firewalled servers, accessed via password protected computers, by the core team. Paper copies of data will be stored in locked filing cabinets in locked offices, linked to digital copies via a unique identifier code. Signed paper-based consent forms will be also saved under lock and key, but separately from other paper-based data so these can never be linked. We will use a Data Safe Haven (a secure system for handling personal identifiable information). Only those researchers directly involved in the research project will have access to the data. Data will routinely be backed up on the

UCL main server. All data will be double checked before storage to ensure there is no personally identifiable data. All data will be identified only by a unique code. All identifying 'metadata' (names of people and places and other similar information) will be deleted from recordings and transcripts. Transcript texts will be saved as text files when this has been done, to remove hidden code that could enable original information to be recovered, and will then be reformatted. Staff need to check that filenames are correctly assigned, and that the correct file format is used, before storing data in study folders or in longer term storage.

Participants who do not consent to their data being shared will not be excluded from the study but we will ensure that their data are not shared outside the research team. Participants may end their involvement at any time up to the point where their data has been integrated into the analysis or resulting outputs. If they require it, we will destroy all their data, which will be done securely according to UCL protocols. If data have been used in disseminations before withdrawal, we will advise participants that this information cannot be withdrawn.

7. Metadata standards and data documentation: All data will be identified using a unique identifier per participant, with suffixes to signal date type (eg IS_GP02_04_data type – where IS – study name or site, GPNN = participant, _NN = interaction number (interview, workshop 1, workshop 2) with that participant), thus making it possible to link any data relating to the same person(s). Data will be stored with 1) detailed descriptions about the protocol, 2) detailed descriptions about procedures used in the data pre-processing and analyses; 3) other relevant metadata; 4) thematic analysis coding process memos.

8. Analysis:

An initial descriptive statistical analysis of the survey data will be performed and updated with each wave for rapid dissemination using RedCap survey tools. Subsequent modelling and network analysis work will also be undertaken to determine important associations and possible causal pathways in the data. For this analysis further researchers may be given access to cleaned anonymized data and all relevant records will be updated accordingly and requests made to the Data Haven, with relevant staff training, before this is done.

9. Dissemination: Survey data will only be reported in summary aggregate form and there will be no possibility that individuals will be identifiable. Interview data anonymized extracts will be disseminated in study findings reports, presentations and publications but care will be taken to ensure no individual can be identified and no sensitive data is shared that could be harmful to participants.

10. Data preservation strategy and standards: Data (converted to suitable open formats for long term preservation as in 1.3) will be deposited for archiving and re-use according to UCL protocols existing at the time. Where permission for archiving has not been granted by participants, in line with UCL policy, all paper records will be held for up to 25 years in central archives, and electronic data stored on the data server for 5 years and subsequently on storage media such as external hard drives and DVDs for 20 years.

11. Formal information/data security standards: Data will be managed in accordance with the University's policies, and legal and relevant institutional ethical approvals (including the Data Protection Act 2018 and GDPR. The lawful basis for processing personal data will be 'public task'). The PI has conducted a Data Privacy Impact Assessment (DCIA) and registered the study for these purposes.

12. Main risks to data security: The main risk is a lack of confidentiality from a data breach or from matching data to personally identifiable information. All data will be held and handled in strict accordance with GDPR and site data protection policies to minimise this. All data for each participant will be referred to by an anonymous code from start of the study. All data will be identified and coded by this code only. Participants actual names will never be used or linked to the data. One file will have a list of participants' names, contact information, and their associated participant numbers. This file will be protected with a password and saved in a password-protected fileshare on the University server. Access to this file will be limited to senior investigators, including investigators listed in this proposal, a PDRA and an administrative assistant. Data will be fully anonymised, which includes removing any directive identifiers (e.g., names) and reducing the precision of variables that can be used as indirective identifiers (e.g., date of birth). Second, completed consent forms that have participants' signatures will be saved separately from other paper-based questionnaires under password-protected main door entry and lock and key. All consultants and collaborators will sign agreements with confidentiality clauses. No patient identifiers will be released in publications/journal articles that could lead to the identification of study participants. Archived data will be checked for anonymisation before sharing; raw data will never be shared but will remain in the UCL safe haven. Other anonymised data will be freely shared. Given the richness of our data and its potential to

address gaps in knowledge, our data have considerable potential to benefit other research groups across the world as well as practitioners.

13. The study team's exclusive use of the data and regulation of other data sharing: Only the direct research team involved in data collection and analysis will have access to the raw data. Prior consent from participants will be sought regarding routine auditing of research records. The PI will decide which users require read-only and read-write access. Off-campus access will be via the Citrix portal. External users who need access to the data will apply for a University username and then be assigned to the appropriate access group. Highly sensitive data will not be available from off-campus.

The study team will have exclusive access to the data during the lifetime of the research. This will enable the aims and objectives of the project to be achieved.

The anonymised qualitative data from interviews and workshops, and anonymised quantitative data from surveys (converted to suitable open formats for long term preservation) will be deposited for archiving and re-use according to UCL protocols existing at the time. These data will be available on request to appropriate (according to UCL archive protocols) professionals and researchers 12 months after end of the study, and for up to 25 years. Archived data will be checked for anonymisation before sharing; raw data will never be shared but will remain in the UCL safe haven. Data that are considered by the custodian to be sensitive and not in the public interest will not be shared despite anonymisation. Other anonymised data will be freely shared according to extant UCL protocols. The custodian of the data to whom requests may be made is Professor Carol Rivas, c.rivas@ucl.ac.uk. Where permission for archiving has not been granted by participants (the option of data re-use is provided in the consent form), in line with UCL policy, all paper records will be held for up to 25 years in central archives, and electronic data stored on the data server for 5 years and subsequently on storage media such as external hard drives and DVDs for 20 years.

14. Where data collection is undertaken by **co-applicant institutions or collaborators**, these will be contractually bound to follow the same due care and data management protocols or their institutional equivalents. Relevant contracts drawn up by UCL and approved by the funders will specify duties and responsibilities.

Data collected at field sites e.g. by co-researchers, will be immediately transferred to password-protected storage locally, to be collected by a UCL team member, or preferably transferred immediately to the UCL Data Safe Haven.

15. Responsibilities: The CI will be responsible for conduct of the study, day-to-day management and decision-making and will have overall responsibility for implementing the data management plan. The lead researchers will be responsible for routine supervision of data collection, transcribing and dataset development and will seek advice from Research Data Managers and IT services where relevant. Data extraction, processing and inputting for the dataset will be undertaken by PDRAs/RAs. Staff involved in the project at participating institutions will be responsible for following data management procedures. Appropriate action will be taken should any breaches or potential for breaches occur; this might include anything from retraining to dismissal depending on the seriousness. Staff must advise the PI immediately of any potential issues or any need to amend either protocol or data management document.

Changes from v1:

Python added to software

Clarification of data sharing after the study in point 13