Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative

Plan de gestion de données créé à l'aide de DMP OPIDoR

Créateurs : Dimitri Papadopoulos Orfanos, Édouard Duchesnay

Affiliation : Other Organisation

Identifiant ORCID : 0000-0002-3660-6640

Modèle : European Union's Horizon 2020 research and innovation programme

Numéro de subvention : 754907

Résumé du projet :

Scientific description - Bipolar disorder (BD) is a prevalent mental disorder and a leading cause of suicide. Lithium (Li) is the key mood stabilizer for prevention of BD relapse and suicide. Whilst many cases become asymptomatic with Li treatment, the majority show sub-optimal response. Identifying biomarkers for predicting Li response would enable personalization of treatment, define criteria for stratification of BD cases and further refine the clinical response phenotype. The objectives of this proposal are to (i) improve outcomes of bipolar I disorder (BDI) cases prescribed Li through the application of stratified approaches; (ii) optimize the early prediction of Li response using a set of multi-modal biomarkers (“blood omics”, Magnetic Resonance Imaging and Li7-Magnetic Resonance Imaging derived-markers); (iii) develop a multidisciplinary multinational network of experts to undertake this and future projects on personalized diagnostics and therapeutics; and (iv) implement new, powerful technologies to characterize brain Li distribution and the blood molecular signature of Li in responders and non-responders. This cutting edge approach will identify the eligibility criteria for treatment with Li in BD in terms of response, safety and tolerability. The assessment of each putative biomarker (singly and combined) will be guided by preliminary findings already obtained by R-LiNK; our expertise will allow exploratory analyses and innovative modeling of multi-modal data. Likely impacts include improved outcomes and quality of life for BDI cases; development of a screening tool for clinicians; and an evaluation of the cost-effectiveness of this stratified approach. The network will develop new avenues of research on Li mechanisms of action and disease

mechanisms; our industrial partnerships will enable development of medical devices to improve treatment adherence and patient’s autonomy, diagnostic kits, and tools based on the molecular signature in treatment responders. The data sharing strategy will follow the "as open as possible as close as necessary" principle. After a five years necessary embargo, the Executive committee will regulate the data sharing following (i) the scientific objectives of the R-LiNK consortium's members; (ii) the regulatory constraints (GPRD, etc.) and (iii) the European recommendations about openscience.

Dernière modification : 28/03/2019

Droits d'auteur

Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative - Initial DMP

1. Data summary

Two clinical studies will be conducted by the R-LiNK project.

The first study is a prospective, multicentre clinical cohort study of individuals with bipolar 1 disorder (BDI) initiating lithium and examining putative early biomarkers predicting long term response/non-response to lithium (Li) treatment. We recruit a naturalistic cohort of **300 patients** initiating Li in all phases of the disorder. We collect data for extensive baseline assessment and intensive follow-up over two years, in **15 recruitment centres** across Europe. Baseline data are used to identify predictors of long-term response/non-response to Li treatment. All components of this study allow us to establish if incorporating any measures of potential biomarkers into day-to-day clinical practice is both clinically viable and economically beneficial. Collected data are :

- **Clinical data** are recorded by recruiting centres directly into a centralised eCRF (electronic case report form) system. Patients are invited to self-report daily symptoms and medication adherence between assessments in a dedicated mobile health system. Data are exported into files in tabular format such as CSV (comma-separated values) or TSV (tabulation-separated values).
- **Biological samples** are collected before and 12 weeks after Lithium initiation. Blood samples are prepared for total blood mRNA, circulating microRNA, metabolomics and proteomic studies. For a given modality, all samples are processed by the same platform. Data are then published in a file format appropriate for each specific modality, independant of any specific programming language.
- **Neuroimaging data** are collected before and 12 weeks after Lithium initiation. The MRI protocol includes T1-weighted, FLAIR and diffusion imaging (DTI) images of the brain,
  single voxel proton magnetic resonance spectroscopy (H-MRS) and is implemented in all recruiting centres of the study. In some centres only, an additional Li7-MRI protocol is implemented to measure brain Li concentration and distribution. Imaging data are collected in DICOM format. Raw data are quality-controlled then preprocessed. Raw and proprecessed data are pusblished in NIfTI (Neuroimaging Informatics Technology Initiative) format.
- Sleep-wake cycle and activity patterns are collected who agree to wear an actiwatch for a minimum of 120 consecutive days from study entry.

The second R-LiNK clinial study is implemented in 2 investigative centres. We evaluate the feasability, safety and patients' acceptability of a home-based device to monitor salivary Li levels. Data from the first study will be re-used in the context of this clinical study, in addition to data from the device itself.

2. FAIR data

The R-LiNK dataset is continuously augmented and improved as acquisition centres send data, curators discover and fix errors, data are processed and published. We plan on regularly releasing new versions of the R-LiNK dataset to take into account these changes while providing some stability to end users.

We will not only assign a Digital Object Identifier (DOI) to the R-LiNK dataset as a whole, but also a distinct DOI to each successive release of the dataset. A public landing page will be created for the dataset and each formal release will be described on this landing page or on its own landing page. To avoid excessive storage cost, we might not keep all successive releases of the R-LiNK dataset.

We will strive to use metadata and data organization standards whenever possible, such as the BIDS format for neuroimaging data. In other cases metadata will be available as free text in the accompanying documentation.

At this point we do not plan on providing a facility to search data by keywords.

Each recruitment site is identified by a 2-digit number and a standardized short name. Each patient enrolled in R-LiNK is identified by a 5-digit pseudonym, starting with the 2 digits of the recruitment site followed by a 3-digit incremental number.

We will use a two levels pseudo-anonimization procedure: the first level will occure at the recruiting center and the second before publication of the data.

R-LiNK data will be initally shared between members of the consortium and their collaborators. One reason is that we need some time to curate data, typically from 6 months to 1 year. This will be followed by an embargo period during which only the consortium members will contribute to the valorization of the R-LiNK data- the exact duration of the embargo needs to be discussed (5 years after the validation of the Database).

Another reason is we need to address privacy concerns in the context of the GDPR and additional national rules. Opening medical research data has always been a legal challenge for research institutions and we need to investigate the **costs of enforcing necessary restrictions in the long term**.

We keep data files sent by acquisition sites or processed by working groups on a storage system located in CEA server rooms. These files could be indexed and/or imported into a database to allow queries. The benefit of using a database needs to be further discussed, and balanced with the costs of database maintenance, especially in the long term beyond the end of the R-LiNK project.

Data release for scientific production will follow the same decision process before and after the end of the embargo. Data release will first take into account the type consent each patient has given (use of the data within the consortium only, also in collaboration with external academic partners only, also in collaboration with external private partners). Principal investigator in charge of the analysis and the draft of the scientific production will have to submit a “Data request form” to the Executive Committee. After approval, the data release will be organized.

Data will be published on a server with direct access to the storage system. The server allows only authenticated and encrypted access, either HTTPS or SSH/SFTP. The exact modalities of access to the server are decided by project management. Once access has been granted,we create an account and assign access rights in accordance with project policy and the needs of the user.

The infrastructure used to collect most of the data, operated by the CATI for imaging and AP-HP for clinical data, is not open source. On the other hand we would like scripts used to process data to be published on GitHub.

AP-HP export clinical data from the eCRF as tables. We are not aware of standards for table or column names in our research domain, therefore tables and columns need to be explained in a specific document. CEA integrate clinical data with the rest of the data and publish them as tables available in CSV or TSV format, accompanied by their documentation.

Imaging data are sent by acquisition centres to CEA in DICOM format. CEA will convert files to NIfTI format and leverage the BIDS emerging standard, specific to neuroimaging studies, to organise NIfTI files and decorate them with metadata in associated JSON files.

Genomics data are produced by the IRCSS from the biological samples. They are sent to CEA as files in standard formats, such as:

- PLINK for genotypic data,
- language-independant self-describing file formats, typically NetCDF for array-oriented data and HDF5 for more complex data structures.

Before the end of the embargo, we cannot share R-LiNK data outside members of the consortium and their collaborators, due to the sensitive nature of medical research. After the end of the embargo, we will evaluate if and how part of the R-LiNK data can be opened more widely. Minimal requirements include further data de-identification and authorization from the Ethics Advisory Board and data protection authorities.

3. Allocation of resources

Question sans réponse.

4. Data security

We will use a two levels pseudo-anonimization procedure: the first level will occure at the recruiting center and the second before publication of the data. Only recruitment sites can convert between first level pseudonym and identifying data. Recruitment sites ship only pseudnymized data. Only, the centralized database can convert between first and second level pseudonyms.

Therefore, even if a link in the processing chain is compromised, it will be impossible to trace the identity of the individual.

All data exchange will be made throug secured transfer protocol (sftp, https).

5. Ethical aspects

Question sans réponse.

6. Other

We collect and process data according to the MR-001 methodology published by CNIL, the French data protection and privacy commissioner.