CLINIC: causes of listening difficulties in children

A Data Management Plan created using DMPonline.be

Creators: Hanne Falcone https://orcid.org/0009-0000-8862-0359, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Astrid van Wieringen

Data Manager: Astrid van Wieringen

Grant number / **URL:** T000823N

ID: 202888

Start date: 08-10-2023

End date: 21-12-2027

Project abstract:

Many children experience listening and processing difficulties (LiD), especially in the presence of background noise, despite normal hearing sensitivity. They are associated with common developmental disorders (DD), such as specific language impairment (SLI, e.g., Sharma et al., 2009), autism spectrum disorder (ASD, Schafer et al., 2020; James et al., 2022), attention deficit and hyperactivity disorder (ADHD, e.g., Blomberg et al., 2019), and learning disorders (LD, Chinn et al., 2022).

The experienced difficulties can have cascading effects on the child's academic and social development. Early identification, differential diagnostics and intervention are important to help children overcome these difficulties and reach their full potential. Some concerns about these listening and processing difficulties, such as the lack of a gold standard to diagnose LiD and age-appropriate reference data, led to the initiation of this study.

The main goal of the CLINIC project is to develop a new approach to diagnose the causes of listening difficulties (LiD) in children. This is achieved through a validated parent questionnaire (ECLiPS, Barry & Moore, 2015), multidisciplinary behavioral assessment and neural measures of auditory pathway integrity. Data from combined measures will lead to evidence-based profiles of children with LiD, which in turn will help streamline their referral pathways and diagnosis.

Last modified: 09-01-2024

CLINIC: causes of listening difficulties in children
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	**Description**	**New or reused**	**Digital or Physical**	**Digital Data Type**	**Digital Data format**	**Digital data volume (MB/GB/TB)**	**Physical volume**
References systematic review	References collected for systematic review	Reuse existing data	Digital	Compiled data	.ris	< 1 GB	
Personal data	Basic information is gathered (age, gender, mother tongue, ...) in REDCap. The data is strictly limited to the necessary data and pseudonymised before being used for further analyses.	Generate new data	Digital	Observational	.csv	< 100 MB	
ECLiPS - population study	Data obtained with the ECLiPS questionnaire during population study (study 1), collected in RedCap	Generate new data	Digital	Observational	.csv	< 1 GB	
ECLiPS - longitudinal study	Data obtained with the ECLiPS questionnaire during longitudinal study (study 2), collected in RedCap	Generate new data	Digital	Observational	.csv	< 1 GB	
CLINIC behavioral measures	Data obtained from various behavioral tests (study 2)	Generate new data	Digital	Experimental	.csv	< 1 GB	
EEG recordings	Recording of EEG (electroencephalogram) experiments	Generate new data	Digital	Experimental	.bdf	250 GB	
Processed data	Processing of recorded EEG data using custom-made Matlab scripts	Generate new data	Digital	Other	.npy, .mat, .csv, .tfrecords	50-250 GB	
Analysis scripts	Analysis scripts done in R.	Generate new data	Digital	Experimental	.r	1 GB	
Results	The outcome of this project. Results can be tables, figures and/or text explaining those.	Generate new data	Digital and non-digital	Compiled data	.pdf, .png, .svg or similar	5 GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Existing literature will be used to write a systematic review. This literature is collected from various databases such as Pubmed etc.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

The CLINIC research project consists of three studies:

1. Population study with ECLiPS questionnaire: anonymous
2. longitudinal, behavioral measures + ECLiPS questionnaire: pseudo anonymous
3. longitudinal, neural measures: pseudo anonymous.

The collection of personal data is strictly limited to the absolutely necessary data.
This study was approved by EC Onderzoek UZ/KU Leuven (S68485).

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Yes, we will use personal data. Depending on the sub-study (see question above), other personal data is collected.
Study 1 (Population study with ECLiPS questionnaire)
Because only four personal identifiers are requested (age, gender, spoken languages, known DD), this study's data collection will be completely anonymous.
Study 2 (longitudinal, behavioral measures + ECLiPS questionnaire) and study 3 (longitudinal, neural measures)
The personal data that will be collected in this study are: month and year of birth, gender, (native) language, any known diagnosis, and relevant medical history such as hearing history. Since the main study population consists of children (< 14 years old), only the contact details (e.g. e-mail address, telephone number) of the legal representative are collected. The file that links the code and personal identification data (subject identification log) is only accessible to authorized persons and is kept separately from other research data. For the remainder of the research, all derived data will be coded and therefore pseudonymised using a unique code.
With regard to GDPR, the project is registered in the KU Leuven "PRET-module" under the file G-2023-7359.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

We intend to provide the Flemish version of the ECLiPS questionnaire, including guidelines to score and interpret this parent-report measure, and our findings on behavioral and neural measures freely to all stakeholders interested in evaluating children on processes important for listening.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- Yes

To use the ECLiPS questionnaire a Material Agreement (MTA) between the 'user' and the University of Nottingham (UK) needs to be signed. For the Flemish version of the ECLiPS, this can be facilitated by Leuven Research and Development (KU Leuven). A similar agreement will be required for the use of the measures developed by ExpORL and implemented in the CLINIC app.

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

New lab members get a training on how to perform the experiments and how and where to store and handle the collected data. Furthermore standard guidelines and instructions are available as .pdf and/or .readme, stored on network drive. They contain best practices regarding the practical side of the experiments (set-up, parameters,...) as well as policies about how to treat the subjects, how to handle and where to keep sensitive
information, etc. In this way, the information given during the training can easily be reread, refreshed.

This, in combination with the fact that the original collected data, are stored on back-uped drives, in a standard format (BIDS, see also website bids.neuroimaging.io), should make it possible to understand and reuse the data.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The data will be stored based on the BIDS structure, an organisation structure for neuroimaging and behavioural data (see also website bids.neuroimaging.io). The BIDS format is essentially a way to structure your data / metadata within a hierarchy of folders. This makes it easy to browse from a computer, as well as to automatically parse a BIDS folder with a program. The BIDS structure makes minimal assumptions about the tools needed to interact with the data that's inside.

3. Data storage & back-up during the research project

Where will the data be stored?

New data is collected during the entire project. Data from the ECLiPS questionnaire (study 1: population study; study 2: longitudinal study in children with DD) will be collected using REDCap electronic data capture tools (Harris et al. 2009, 2019) hosted at KU Leuven. Metadata (timestamp, electronic instructions) are automatically captured during survey completion in REDCap and will not be used in the analysis of the study data. Files containing metadata will not be stored on a different server than the raw data.
Behavioral data (study 2: longitudinal study in children with DD) is exported directly from the tablet to a protected drive of the research group. This data will also be imported into the REDCap database by the site-specific employee or an applicant. The file that links the code and personal identification data (subject identification log) is only accessible to authorized persons and is kept separately from other research data.
Data in the form of EEG brain activity is collected during neural measurements. The data will be stored on KU Leuven administered drives (large volume storage and OneDrive). In order to be able to analyse the data in Matlab, some files will need (temporarily) to be stored on the encrypted PC hard drive (this since calculations from a non-local source are too slow and lead to computational failures). Once analysed, the raw data are again removed from the local hard drive. The file that links the code and personal identification data (subject identification log) is only accessible to authorized persons and is kept separately from other research data.

How will the data be backed up?

Since the data are stored on KU Leuven storage, the general ICT back-up Policy is

applied (daily automatic back-up).

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

Since the data are stored on KU Leuven servers, and these drives are expandable in blocks, the backup capacity is technically not an issue.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All included storage facilities (REDCap, protected drives) are incorporated within secured KU / UZ Leuven environments, are password-protected (including smartphone-based multi-factor identification) and are only accessible by registered collaborating researchers. RedCap is a password-secured, web-based software platform. The database is only accessible by collaborating researchers. Participants complete the survey via a survey link. They are not logging into REDCap to enter the data and they cannot see other data that has been entered into the project. Only applicants and site-specific employees are able to log into REDCap to see data and perform actions, according to permissions that have been given to them (according to their data access group).
The lab policy is that the researchers have only access to the data from the project they are involved in. Furthermore, the data for longer term storage are kept on separate drives with a) limited access (only a limited set of people have access) and b) an overwrite and delete protection (based on read-write access) in order to prevent accidental loss of these data.
All data collected throughout the study will be pseudonymized and coded. This unique code will not contain any elements that could lead to the identification of the participant, not even for the researchers involved. The file linking the code and personal identifiers will only be accessible to authorized individuals and stored in a restricted access, secure environment managed by the KU Leuven ICT facility. Access will be controlled by PI-determined access rights mediated by password protection.
In case sharing of the data within the lab is necessary/favorable, the PI will decide which data will be shared. For sharing outside the lab, the explicit consent of the subject will be asked (in writing). Pseudonymized data can be made available for further analysis in line with the terms of the ICFs. The ICF contains a section in which the participant can choose if his/her data is shared. Possibilities are:

- no sharing allowed
- sharing within KU Leuven
- sharing within the European Union
- worldwide sharing
- worldwide sharing and the data can be made available on a public database.

Only coded and pseudonymized data can be shared.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

There is standard 2TB personal OneDrive cloud storage available for KU Leuven employees. No additional costs are expected for OneDrive.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data will be retained for at least 5 year after the end of the project. Research data resulting from the current project will be discarded after 25 years according to KU Leuven/UZ Leuven research data policy regulations.

Where will these data be archived (stored and curated for the long-term)?

The data are stored on a KU Leuven, ICT managed Large Volume Storage, drives especially designed for archive storage.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The archive drive (K:) storage costs approximately 200 euro per TB per year.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

We opt for a mix of gold and green open access articles for the dissemination of our research findings. Articles that are published in journals that do not grant open access online will be self-archived after publication in the online repositories of KU Leuven.

If access is restricted, please specify who will be able to access the data and under what conditions.

All data is collected in REDCap. By using Data Access Groups (DAGs), we partially restrict viewing of data within a database. Users at each participating site are assigned to a group and will only be able to see records created by users within their group. KU Leuven applicants are able to see the full dataset.
All included storage facilities (REDCap, protected KU Leuven drives) are incorporated within secured KU / UZ Leuven environments, are password-protected (including smartphone-based multi-factor identification) and are only accessible by registered collaborating researchers. Access will be controlled by PI-determined access rights mediated by password protection.
Data will be coded and therefore pseudonymised using a random and unique code. This unique code does not contain any elements that could lead to identification of the participant, not even for the researchers involved. The file that links the code and personal identification data (subject identification log) is only accessible to authorized persons and is kept separately from other research data.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects
- Yes, Ethical aspects

Pseudonymised data can be made available for further analysis in line with the terms of the ICFs. The ICF contains a section in which the participipant can choose if his/her data is shared. Possibilities are:

- no sharing allowed
- sharing within KU Leuven
- sharing within the European Union
- worldwide sharing
- worldwide sharing and the data can be made available on a public database

Only coded and pseudonymised data can be shared.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The protocol for the Systematic Review will be published on PROSPERO. For the published results, we opt for a mix of gold and green open access. The studies described will be published in (inter)national journals, and presented at (inter)national conferences. A PhD thesis will finalize the doctoral trajectory of the junior researcher.
With support of our advisory committee and societal actors, we will communicate progress via a dedicated part of the website of ExpORL (https://www.kuleuven.be/exporl). The lay audience will be informed through social media (Linkedin), flyers and the website, and popular scientific magazines (cf our publication in EOS, e.g.). We will also share knowledge and skills with our students of Speech Pathology and Audiology Sciences and with the World Rehabilitation Alliance (of which the PI is a member).

When will the data be made available?

As soon as analyses are complete, the data can be made available.

Which data usage licenses are you going to provide? If none, please explain why.

Data will only be used within KU Leuven. If external parties are interested in the dataset, it can be obtained on request, and personal, individual agreements will be composed.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

What are the expected costs for data sharing? How will these costs be covered?

not applicable.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

All researchers directly involved in the project are responsible for data documentation and metadata. The PI (Astrid van Wieringen) bears final responsibility

Who will manage data storage and backup during the research project?

All researchers directly involved in the project are responsible for data storage and back up. The PI (Astrid van Wieringen) bears final responsibility.

Who will manage data preservation and sharing?

The PI, Astrid van Wieringen

Who will update and implement this DMP?

Hanne Falcone (PhD researcher), under the responsibility of the PI, Astrid van WIeringen