Stepwise Heat-Denaturated protein introduction for tolerance induction in food allergy and pediatric Eosinophilic Esophagitis (TEHITI)

A Data Management Plan created using DMPonline.be

Creators: Jonathan Cremer, n.n. n.n., 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: n.n. n.n., n.n. n.n., Jonathan Cremer, Cheyenne Keppens

Data Manager: n.n. n.n., n.n. n.n., n.n. n.n., Jonathan Cremer, Cheyenne Keppens

Project Administrator: n.n. n.n.

Grant number / URL: T000623N

ID: 201568

Start date: 01-09-2023

End date: 01-09-2027

Project abstract:

This project aims to use heat-denatured food proteins (allergens) in children with type I (IgE-mediated) food allergy, to study whether these can of help to induce tolerance. The project
contains two clinical trials with associated research to unravel underlying immunological mechanisms paralleling tolerance induction in children with cow's milk allergy (CMA) and remission in children with cow's milk and/or hen's egg mediated eosinophilic esophagitis (pedEoE).

- WP1:
  P: Children with IgE-mediated CMA that are tolerant for 20' extensively heated cow's milk. 12 months open trial: 1/1/1:
  I: Two arms in which one arm induces cow's milk proteins gradually reducing boiling time with 5'each three months and the other arm introduces cow's milk proteins using the Flemish Milk Ladder during a 12 month protocol.
  C: Children remain using the tolerated 20' boiled cow's milk proteins for 12 months without addition of any other cow's milk protein (natural tolerance induction)
  O: Proportion of children in each arm achieving complete tolerance after 12 months (who pass the oral food challenge test with unheated cow's milk)
- Within WP2 and WP3 we will study immunological changes in T and B cell responses and basophil stimulation tests with cow's milk proteins/extracts.
- WP4:
  P: Children with pedEoE in remission on a diet free of cow's milk and/or hen's egg
  I: Progressive introduction of gradually less heated cow's milk and/or hen's egg during 8 weeks with periodic evaluation by gastroscopy and biopsy till re-occurrence of disease.
  C: None
  O: Proportion of children with pedEoE in remission after reintroduction of extensively (and later on gradually less) heated proteins from cow's milk and/or hen's egg; associated clinical and/or immunological characteristics of children with remaining remission. Changes in IgG4 production ex vivo and in vitro by B cell stimulation will be studied in parallel.

Last modified: 04-03-2024

Stepwise Heat-Denaturated protein introduction for tolerance induction in food allergy and pediatric Eosinophilic Esophagitis (TEHITI)

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Data will be generated. There will be no (re)use of data.
We will be working with personal data. We start from non-anonymized patient data, but after inclusion in the TEHITI study, the data is pseudonymized.

- Consent & assent files [paper]
- Quality of life questionnaires [paper]
- Clinical data [.docx, .xlsx]
- Cells & serum [Biobank]
- Biopsy slides [Slide boxes]
- BAT [.xlsx]
- Flow cytometry data [.fcs]
- SDS PAGE [.gif, .jpeg]
- Manuscripts [.doc, .pdf]
- Analyses [.pzfx, .ssd01, .xpt]
- Figures [.pzfx]

Total estimated volume of data: 500 GB

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Prof. Dominique MA Bullens (MD, PhD; project PI; head of Leuven Allergy & Clinical Immunology Research Group (LACI)) is the designated responsible person.
2. Storage capacity/repository
    - During the research
        - Patient personal information: local hospital patient information systems with restricted access
        - Consent & assent files: storage at LACI
        - Pseudonymised research data: KU Leuven network storage, J:drive with restricted access
        - Biological samples: storage at LACI
    - After the research
        - Patient personal information: local hospital patient information systems with restricted access
        - Consent & assent files: storage at LACI
        - Pseudonymised research data: KU Leuven network storage archive, K:drive with restricted access
        - Repository: KU Leuven RDR
        - Biological samples: storage at LACI

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

There is no planned deviation from the principle of preservation of data beyond practical reasons for biological specimens (specimen used completely for analysis, limited shelf life of e.g., immunofluorescently stained slides).

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

There are no issues concerning research data that could be used maliciously so no additional specific security measures are required besides

pseudonymization and access restriction of personal data. Personal data will only be accessible for the principal investigator and co-investigators specified on the assent/consent files.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

There are no other issues related to data management.

Stepwise Heat-Denaturated protein introduction for tolerance induction in food allergy and pediatric Eosinophilic Esophagitis (TEHITI)

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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
IAC	Informed assent & consent	Generate new data	Physical	Observational	NA	NA	25 binders
QoL	Questionnaires investigating quality of life	Generate new data	Physical	Observational	NA	NA	**25 binders**
CD	Clinical data	Generate new data	Digital	Observational	.docx, .xlsx	<100MB	
BAT	Basophil activation tests	Generate new data	Digital	Experimental	.csv, .pdf, .fcs, .wsp	<1TB	
BR1	Flow cytometric identification of BR1 cells	Generate new data	Digital	Experimental	.fcs, .wsp, .pdf, .csv	<1TB	
TR1	Flow cytometric identification of TR1 cells	Generate new data	Digital	Experimental	.fcs, .wsp, .pdf, .csv	<1TB	
BR1CK	Cytokine analysis of cultured PBMC, BR1 subset	Generate new data	Digital	Experimental	.txt, .csv, .xlsx, .png, .pdf, .pzfx	<100MB	
TR1CK	Cytokine analysis of cultured PBMC, TR1 subset	Generate new data	Digital	Experimental	.txt, .csv, .xlsx, .png, .pdf, .pzfx	<100MB	
IgG4Biop	IgG4 analysis in biopsy tissue sections	Generate new data	Physical/Digital	Experimental	images	<100MB	2x 100-slide boxes
SDS PAGE	SDS PAGE analysis	Generate new data	Digital	Experimental	.gif, .jpeg	<100MB	

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:

Not applicable.

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 data will contain sensitive personal data about the subject and the subjects' health condition. Human bodily materials will also be collected and processed and may result in genetic data being analysed.

Ethical Committee Research UZ/KU Leuven will be the central EC that will give approval after they consulted the local ECs: EC Maria Middelares Ghent, EC AZ Sint-Jan Brugge, EC AZ Bonheiden and EC Jessa Hospital

Both the EC and the Clinical Trial Center (CTC) will review the study in detail, in accordance with all applicable regulatory requirements. Reference number s67001 was assigned to the study including pedEoE children of which blood samples and biopsies will be taken (IAC, CD, BR1, BR1CK, IgG4Biop). The reference number s68299 for the study including IgE-mediated cow's milk allergic children from which blood samples will be taken awaits approval, but has been approved by the CTC (IAC, QoL, CD, BR1, BR1CK, TR1, TR1CK, SDS PAGE). Lastly, s68469 was assigned to the follow-up study of the TEHITI study; also approved by CTC awaiting EC final approval.

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

Privacy Registry Reference: KWS (Klinisch Werk Station) UZ Leuven

We start from non-anonymized patient data, which - after inclusion in the TEHITI-study (upon signing assents and informed consent) - will be pseudonymised and each patient is from then on only identifiable by a unique TEHITI ID-code. Only the treating physician holds the code, which is kept in a secure place, and is able to go back to the patient chart, when needed. This is also clearly stipulated in the informed consent form.

We will further stay in contact with Toon Boon for optimizing the way how to deal with these personal data.

Name and address of participants, including contact information of their parents, will be collected (dataset IAC). Further personal information will include date of birth, initials and hospital-defined identification number assigned to data subjects (CD).

Sensitive personal data will be collected in datasets QoL, BAT, BR1, TR1, BR1CK, TR1CK, IgG4Biop. This will include health data (description of characteristics of physical features of the body, medical history, medical test information) and genetic data.

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.

- No

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.

- Yes

Third party agreements with each of the participating centers within work packages 1, 2, 3 and 4 are made by CTC and signed by UZ Leuven and the legal representatives of the separate centers. UZ Leuven is legal owner of project results. Results are confidential during the trial. Restrictions on publication and authorship are part of the clinical protocol of work packages 1, 2, 3 and 4 signed by each partner.

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.

- No

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).

1. Subjects will receive a unique code (pseudonymisation) based on the prefix TEHITI and a 3-digit number according to the order of inclusion (WP1), a letter identifying the study site (L: Leuven, H: Hasselt, B: Bonheiden, Br: Brugge, G: Gent) and a 3-digit individual code. A spreadsheet of the identification code will be kept separately in an encrypted network folder of LACI (J:\GBW-0517_TEHITI). All necessary steps to remove direct identifiers in the data (e.g., name, address, etc.) will be taken (CD).
2. Research methods and practices (including the assent and informed consent process) will be fully documented as hard copies, as well as a blank copy of the informed consent form as text file. The informed assent and consent files (IAC), as well as the Quality of Life questionnaires (QoL), will be stored in binders with a cover, describing the content of the binder. For WP1, one binder will contain the files from five patients for each of the collaborating hospitals separately. Additionally, one binder will contain the informed assent and consent files from all healthy age-matched controls. For WP4, one binder will again contain the files from five patients for each of the collaborating hospitals separately.
3. Raw experimental data will be collected per patient for WP1, WP2, WP3 and WP4 separately, including a text file (.txt) with a clear description of what the data represent and how they were generated. The input files used for the test will be kept inside the same folder. The name of the folder will contain the study acronym, the conditions (each step of the gradual cow's milk tolerance cascade) and a description of the material. Datasets BAT, TR1 and BR1 will be collected with flow cytometry. Moreover, datasets TR1CK and BR1CK will be collected by Olink or Mesoscale Discovery.
4. The codebook of each coworker will contain information on date of puncture, study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.

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

For flow cytometric data, the MIFlowCyt metadata standard will be used.

For other datasets, metadata will be provided as README in .txt format describing the dataset in general, supporting information such as protocols in .docx and .pdf, tabular information as .csv, raw numerical data output in .csv, data analysis files in .xlsx and .pzfx, and data files from laboratory equipment in their respective output format if no open format (.txt or .csv) is available.
Supporting information will detail materials and methods used, equipment settings and quality control (if applicable), experimental and biological controls. In case laboratory equipment provides ways to attach metadata to output (such as imaging files), this is completed at time of experiment.

The used medical terminology (e.g. cow's milk allergy and pediatric eosinophilic esophagitis) is according to the standards in the allergy field, as represented by the European Academy of Allergy & Clinical Immunology (EAACI) and Pediatric Gastroenterology, Hepatology and Nutrition (PGHN).

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

Where will the data be stored?

All patient personal information will be stored on the UZ Leuven IT systems and local hospital IT systems, accessible only to a restricted number of individuals (attending physicians) on a need-to-know basis.

All pseudonymized experimental data, metadata and protocols will be stored on the KU Leuven network drives:

- Active documents and data will be kept on the network J:-drive, more specifically J:\GBW-0517_TEHITI.
- Large, unchanging data files may be kept on the network K:-drive, more specifically K:\GBW-0041_KlinImm\TEHITI. This is done to save costs associated with the J:-drive.

How will the data be backed up?

Data and systems backup on UZ Leuven systems is managed by UZ Leuven ICT.
The KU Leuven network drives are managed as follows in the primary KUL data center in Heverlee:

- For J:-drive:

Backups are made using “snapshot” technology, which is the online storage of incremental data changes. For the standard backup regime, as specified below, 10% of the requested storage capacity will be reserved:

- An 4-hourly backup (at 8 AM, 12 PM, 4 PM and 8 PM) the last 6 of which are stored on our servers
- A daily backup, at midnight, the last 6 of which are stored on our servers
- A weekly backup, Saturday night at midnight, the last 12 of which are stored on our servers

The end user can use his own Windows PC to restore files to an older version using the “previous versions” function. According to the backup system above, it is possible to go back in time up to 12 weeks (~3 months).
For the purpose of “business continuity” or “disaster recovery”, a mirror (exact copy) of all data is created in the second ICTS data center. The data are copied every hour to the second data center. In the event that the primary storage unit is corrupted, the ICTS team can get this copy online within the hour. The mirror in the second ICTS datacenter falls under type 1 storage.

- For K:-drive:

Automatic version management of the files. Version management is done using "snapshot" technology, where the previous versions of the changed files are kept online in a snapshot on the same storage system.

- by default, 1 snapshot is taken daily and is kept for 14 days. So you can go back to previous versions of the file up to 14 days.
- end users can restore older files themselves from within their Windows PC via the "previous versions | previous versions" functionality.

Mirror:

- a mirror (an exact copy) of the data is provided in the second ICTS data center for “business continuity” or “disaster recovery” purposes;
- a file is copied to the second data center as soon as it is written to a drive. ICTS can put the copy online within an hour in case of disaster with the primary storage;

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

All patient personal information is collected for the purpose of the patients' treatment, therefore it is kept according to Belgian law in the UZ Leuven systems.
All experimental data is kept on the KU Leuven network storage. The ICTS department provides ample storage capacity in their data centers, and the project-specific allocated space can be increased upon simple request at a fixed cost per space required.
During the project, the data managers will monitor the storage space required and request extra capacity from ICTS when necessary. This process is relatively quick (a few days at most) and saves overall cost as compared to reserving the entire estimated storage capacity from the start.
Jonathan Cremer, co-worker of the Allergy and Clinical Immunology Research Group, is in charge of adequate management of this storage capacity.

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

The UZ Leuven systems are purpose-built to keep patient information confidential. Access to the systems is only given to personnel, with multifactor authentication in place. The systems are only accessible through specially configured workstations, laptops or VPN. All actions on the network and in the systems are logged and monitored.
External partners (local hospitals) have their own patient information systems. (3 out 4 use KWS but access is restricted to the treating physician; Maria Middelares Ghent uses MM Synops). Data are locally extracted, pseudonymised and transferred to KUL via safe data transfer protocols.
Data transfer from UZ Leuven or external partners to KU Leuven is performed using 'Liquid files'.
The KU Leuven network is also closed with multifactor authentication, accessible only to authorized personnel via configured hardware or VPN. Furthermore, the folders containing the data on the network drives are set up with security groups, permitting the data managers of the

project to give or revoke access to/from specific individuals in the organization. This access will be logged in a spreadsheet maintained by the data managers.
Data transfer from KU Leuven is performed using Belnet Filesender or Globus.

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

Costs for data storage are as follows:
Personal data of participating subjects is stored in the regular clinical information systems. No extra costs are incurred for the project.
KU Leuven network J:-drive has a yearly cost of €51,9 per 100 GB. With an estimated total data size of 500 GB at the end of the project and a yearly growth of 100 GB, the data storage and backup cost during the project would be an estimated €778,5.
The project budget will cover this expense.

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 digital data will be stored for 10 years in accordance with the informed consent.
All remaining physical samples will be transferred to the UZ/KU Leuven Biobank after the study has been concluded. For s68299, the Biobank application has already been approved (BB-GEN002-FO03).

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

Personal data of study subjects will stay in the hospital clinical information systems.
Experimental data will be archived on KU Leuven network drive K:.
Upon final publication of results, datasets may be published in the KU Leuven RDR.

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

KU Leuven network K:-drive has a yearly cost of €5,69 per 100 GB. With an estimated total data size of 500 GB at the end of the project, and a projected 10 year storage, the storage and backup cost after the project has ended would be €56,90. A storage cost increase due to e.g. inflation can be expected.
After termination of the project, the PI's 'personal' budget will cover the storage costs.

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 a restricted access repository (after approval, institutional access only, …)

Raw .fcs-files will be made available after publication as specified by MIFlowCyt. Other raw data will not be made available unless upon specific request and after evaluation of the Ethical Committee (s68299, s67001, s68469).
Clinical data about the patients are confidential, subject to compliance with applicable personal data protection laws, and not publicly available.
The pseudonymized data will be available upon publication in open access journals.

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

Personal data will only be available to the head investigator and co-investigators in agreement with the child's and parent's assent and consent. Afterward, the data will be shared between the PI and co-PI's and within the research unit but always respecting the pseudonymization.
Sequence data will be uploaded in an open access repository and shared upon request.
Data without sharing restrictions will be shared through peer reviewed publications.

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

With respect to patient data and biomaterial: patient privacy is respected and permission to share data/samples is obtained by the informed consents and stipulates clearly who will have insight in the pseudonymized data. Data will not be shared with other (third) parties, not participating in the study.

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

Since we are working with sensitive data, KU Leuven RDR will be used. Sequence data will be uploaded in an open access repository and only shared upon request. Data without sharing restrictions will be shared through peer reviewed publications.

When will the data be made available?

After final publication of grouped analysis of the two work packages (WP1, WP4).

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

No specific license has been selected yet. This will be done in close collaboration with the RDM Support group at KU Leuven and will likely be a Creative Commons-type license.

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.

- Yes

All datasets published in KU Leuven RDR will have a DOI assigned.

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

Each researcher can currently deposit up to 50 GB per year in KU Leuven RDR at no charge. This will likely be sufficient for the project's datasets. Peers may use the data at no cost under condition of co-authorship. Commercial organizations are not anticipated to share data (or biomaterial) at the moment but if secondary use of material would later on be possible for them (after informing the patient/legal representative of this third party commercial use of this material), they will have to pay a fee that will be determined by LRD-KU Leuven.

6. Responsibilities

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

The PI (Dominique MA Bullens), Cheyenne Keppens and Jonathan Cremer will manage documentation and metadata during the research project.

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

Jonathan Cremer and Sabien Fevery will manage data storage and backup during the project.

Who will manage data preservation and sharing?

The PI (Bullens), KU Leuven co-PI (Schrijvers), and four local co-PI's (Leus, Coppens, Verelst, Waelkens)).

Who will update and implement this DMP?

Dominique MA Bullens, Cheyenne Keppens and Jonathan Cremer will update and implement this DMP. The PI bears the end responsibility of updating and implementing this DMP.