**IMI2 821520 - ConcePTION**

**ConcePTION**

**WP8 - Scientific coordination, project management & sustainability**

# D8.5 Data Management Plan

| | |
|---|---|
| **Lead contributor** | Florian van der Nolle (1. UMC Utrecht) |
| | f.l.vandernolle-raven@umcutrecht.nl |
| **Other contributors** | Pieter Stolk (1. UMCU) |
| | Miriam Sturkenboom (1. UMCU) |
| | Peter Singleton (6. i-HD) |
| | Dipak Kalra (6. i-HD) |
| | Marianne Cunnington (41. GSK) |

## Document History

| Version | Date | Description | Non-contributor reviewers (if applicable) |
|---|---|---|---|
| V0.1 | 14 Feb. 2020 | First Draft | |
| V1.0 | 25 Feb 2020 | Second draft | |
| V1.1 | 5 March 2020 | First version | Managing Board |

## Summary

The ConcePTION Description of Action (DoA) includes a Data Management Plan (DMP) as deliverables D8.5 and D8.19, as part of WP8. The DMP describes the data management life cycle for all data sets that will be collected, processed or generated by the project. In addition to the ConcePTION Consortium Agreement (CA), the DMP provides a general framework regarding data management, data protection, data ownership, accessibility and sustainability requirements.

Overall, the DMP provides a description of the data management, regarding generated research data, that will be applied during the ConcePTION project including:

- Data management and strategy information per WP and/or task, including but not limited to a data summary, data collection, data storage, data processing; quality control and governance
- The possibilities of and conditions for sharing data
- The implementation of data protection requirements

The DMP is an evolving document, therefore, some aspects may be added and/or updated in later version of the document.

# Table of contents

# 1. Introduction and aim

ConcePTION aims to create a paradigm shift in how we generate and disseminate evidence on the effects of medication in pregnancy. We intend to establish a trusted ecosystem that can efficiently, systematically, and in an ethically responsible manner, generate and disseminate reliable evidence-based information regarding effects of medications used during pregnancy and breastfeeding to women and their healthcare providers. This will be achieved by generating, cataloguing, linking, collecting and analysing data from pharmacovigilance, modelling, routine healthcare, pregnant women and their children through a large network.

The ConcePTION Description of Action (DoA) includes a Data Management Plan (DMP) as deliverables D8.5 and D8.19, as part of WP8. The DMP describes the data management life cycle for all data sets that will be collected, processed or generated by the project. In addition to the ConcePTION Consortium Agreement (CA), the DMP provides a general framework regarding data management, data protection, data ownership, accessibility and sustainability requirements.

Overall, the DMP describes the data management, regarding generated research data, that will be applied during the ConcePTION project including:

- Data management and strategy information per WP and/or task, including but not limited to a data summary, data collection, data storage, data processing; quality control and governance
- The possibilities of and conditions for sharing data
- The implementation of data protection requirements

The DMP is an evolving document, therefore, some aspects may be added and/or updated in a later version of the document. The DMP will be updated with the periodic evaluation/assessment of the project and as deliverable 8.19 in M60.

In summary, the ConcePTION DMP gives guidance and provides oversight of general data management, while each study needs to provide specific data management information including, but not limited to, data capture systems, data analysis systems, data protection and data privacy measures, including a description of de-identification of data sets and access rules. In cases where the research results are not open access, a justification needs to be provided.

# 2. General principles

This report is the initial DMP for ConcePTION. As mentioned before, the DMP is a working document that will evolve during the project and will be updated to reflect project progress. The DMP will be updated with the periodic evaluation/assessment of the project and as deliverable 8.19 in M60. Additional updates will be done whenever important changes occur e.g. due to the creation of new data sets.

Procedures relating to the different data management plan aspects will be worked out as the project progresses and will be explained in more detail in an upcoming version of the DMP.

The DMP follows the 'FAIR data principle', i.e. data should be findable, accessible, interoperable and re-usable[1].

The general principles on access rules are defined in the ConcePTION Consortium Agreement (Section 8 Intellectual property – Access rights).

## Collaborative website – ConcePTION Member Area

ConcePTION makes use of one information exchange platform, the ConcePTION member area. The member area is a password secured web space were consortium member can store and exchange reports and documents. The platform is not meant to share patient research datasets. The member area is hosted by the coordinator (UMCU), contact person: Florian van der Nolle (f.l.vandernolle-raven@umcutrecht.nl).

[1]European Commission Horizon2020 programme. Guidelines on FAIR Data Management, v3.0, 2016. (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

# 3. Data types and formats generated or collected in ConcePTION

ConcePTION aims to create an ecosystem for the rapid and robust generation of evidence on the safety of medications in pregnancy and during lactation, using both existing and newly generated real world data.

The real world data that will be transformed into evidence is coming from various sources:

1) **Existing data**
    a. Pharmacovigilance data e.g. reports of adverse events in the mother or child following drug exposure prior to or during pregnancy, or during lactation. This data may also capture events following paternal exposure.
    b. Research data e.g. Drug exposure pregnancy registries that recruit women who are exposed to specific drug(s) and are followed up prospectively
    c. Existing health care and surveillance data:
        - *Healthcare claims databases* – created for operational health care purposes and billing of costs on defined population that is followed over time (for example drug dispensing claims)
        - *General practice databases* – electronic medical records provided by General Practitioners (GPs) on defined population that is followed-up prospectively
        - *Birth research cohorts* - pregnant women recruited during pregnancy or at birth, irrespective of exposure, and followed-up prospectively
        - *Demographic/population databases* – includes the population register, residents register, date of birth/death
        - *Linkable Registries*: relevant outcome/exposure data collected for a specific purpose when they can be linked to an underlying population file that defines follow-up time
            - Medical Birth Registries
            - Specific Disease or outcomes surveillance registries e.g. congenital anomaly registries such as EUROCAT, cancer registries, infectious disease surveillance, death
            - Child surveillance databases –growth and developmental records as measured by community child health teams/public health nurses
            - Educational databases - created for operational education administration purposed for example school results, special educational needs and attendance
            - Registry of disability - created for insurance purposes and service delivery
            - Immunization registries
            - Medical encounter databases: hospital based encounters, laboratory measurements, imaging

2) **Newly collected data**
    a. Human: human reported data on pregnancy (prospective)
    b. Milk: human milk samples
    c. Blood: human blood samples
    d. Animal: animal milk and blood data on milk transfer
    e. Human cells

The trajectory and steps of conversion of data into evidence is depicted in Figure 1 below.

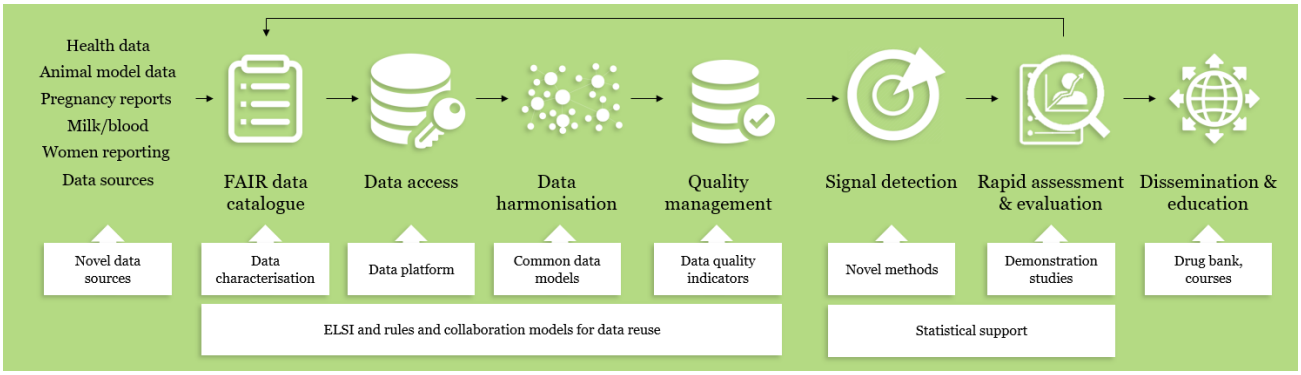


**Figure 1 Central to the ConcePTION ecosystem is a dedicated workflow that allows for identification, verification and transformation of data into evidence**

The technical infrastructure that will support the management and generation of evidence from different sources of data and the data dimension and approach are described in more detail in D7.2 (Description of the operational platform for data sharing and task management system).

The key principles in ConcePTION are that we work as a distributed data system, and that analytics go to the data rather than the data to the analytics.

## Data types per work package

The ConcePTION workplan has been divided in 8 workpackages (Figure 2). Workpackages 1-4 will deal with generation of evidence from different sources of data, WP5 will summarize and disseminate the evidence, WP6 will discuss and get feedback on acceptability of the tools/solutions from stakeholders, and WP7 is providing the ethical, IT system, and quality dimensions. WP8 deals with project management and coordination. Whereas all WP will deal with some type of data, the focus will on WP1-4 as they will deal with transformation of original health/animal/ data into evidence.

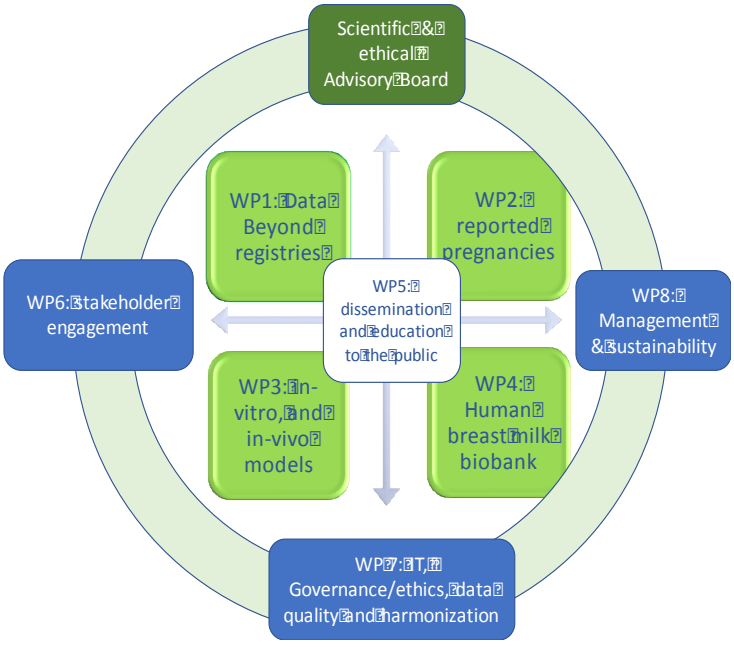


**Figure 2 Workplan according to workpackages in the ConcePTION project**

# 4. Data management and strategy

To map all foreseen data collection and to establish the data management needs, every WP is requested to complete the 'Data management and strategy survey' (tables below). The tables are based around the following topics:

- Data summary
- Data collection
- Data storage
- Data processing; cleaning, transforming and analysing
- Quality control
- Governance

In addition to these tables, every WP was allowed to provide additional comments related to data management which wasn't covert by the 6 topics above.

For this initial DMP WP1-5 and 7 were requested to complete the data management and strategy survey between September and December 2019. The surveys of these WPs can be found in Annex 1 – 6. The DMP is an evolving document and throughout the project, as more specific information about data management and strategy is known, the annexes will be updated.

---

## Data management and strategy survey

WP…….
Compiled by…..

### 1. Data summary

| |
|---|
| What is the purpose of the data collection/generation and its relation to the objectives of the project? |
| What types and formats of data will the project generate/collect? |
| What is the origin of the data? |
| Will you re-use any existing data and how? |
| What is the expected size of the data? |
| To whom might it be useful ('data utility') in the project and after wards? |

2. **Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Task | Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |

Please comment……..

3. **Please describe where the collected data will stored for each task and type of data collection**

| Task | Type of data to be collected | Where (at which physical location) is primary (original) data stored? | What software is used for storage of data? | What format and type of data standards will you use to store the data? | How will you make the primary data accessible to other partners in consortium. Are there restrictions? |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |

*Please make a data localization and format organogram .*

4. **Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |

Please explain if you wish……

5. **How will you conduct quality control?**

| Task | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| | | | | |
| | | | | |

6. **Governance please complete for each participating partner**

| Partner | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| | | | | |
| | | | | |

**Please feel free to provide additional comments**

| |
|---|
| |

## Responsibilities of the data owner(s)

The data owners of the respective research projects and dataset are responsible to comply with all legal and ethical requirements for data collection, handling, protection and storage. This includes adherence to regulations, guidelines such as (but not limited to) the EU clinical trial directive 2001/20/EC, Good clinical practice (GCP) and Good Pharmacoepidemiology Practice (GPP), as applicable.

# 5. Data sharing and secondary use of generated or collected data and evidence

The information collected from the completed 'Data Management and Strategy surveys' will be uploaded on the ConcePTION member area. This will enable easy identification of the available datasets and their respective data owners by consortium members. The data owners are responsible for appropriate findability outside the consortium.

## Remote data plaform

The DoA describes requirements for a remote data plat form:

*"A state-of-the-art digital research environment with ISO certified and GDPR proof services for remote collaborations will be subcontracted and operated. Access to the application server will be only allowed using two-factor authentication. The environment will be able to host multiple research projects, each with its own secured area to share data and results and provide access through remote desk tops clients. The infrastructure will offer several analytical tools (e.g. R, SQL database, Shiny, Stata) word processing software, and utilities"*

Deliverable D7.2 (Description of the operational platform for data sharing and task management system) describes the remote data platform in more detail. The report will describe the requirements of the data platform and will also review and test multiple solutions.

# 6. Protection of personal data

The collection, handling storage and exchange of personal data will be conducted in a secure manner, through secure channels. In addition, this will happen under the applicable international, IMI and national laws and regulations. Only data of relevance for the proposed research will be collected, no excess data will be stored. ConcePTION researchers commit to the highest standers of data security and protection in order to preserve the personal rights and interests of study participants. They will adhere to the provisions set out in the:

- Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR)[2]
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)[3]

Deliverable D7.20 (Templates and guidance for local and central Data Privacy Impact Assessments for data sources and repository for continuous collection of completed forms and approvals) will explain the protection of personal data in more detail.

[2] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&rid=1
[3] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058&rid=1

# 7. Ethical aspects

All ConcePTION participants are committed to the highest standards of ethics and privacy protection. Every partner will ensure that all staff working on the ConcePTION project complies to all relevant National, International and EU legislation relating to the processing of personal data, will protect privacy and maintain confidentiality, the legal and ethical use of human cells/tissues and animal experimentation.

The ConcePTION consortium has performed an ethics self-assessment (DoA, Section 5). In addition, an Ethics Advisory Board will be established to ensure that potential ethical issues are actively monitored and dealt with accordingly throughout the project.

**Responsibilities of the Ethics Advisory Board**
The Ethics Advisory Board is an advisory board to the ConcePTION project in general and to the Managing Board in particular. The Ethics Advisory Board will advise the Managing Board upon request of the Project Leader together with the Coordinator and provide non-binding advice to the General Assembly and the Managing Board as decision making support. The Ethics Advisory Board will not have decision authority in the project but will provide advice and feedback on the activities and results of ConcePTION.

The Scientific and Ethics Advisory Board will be responsible for:

a. Reviewing the proper application of the ethical rules by the Beneficiaries;
b. Providing advice to the Beneficiaries, the General Assembly and the Managing Board on ethical issues; and
c. Providing advice on the compliance with European ethical laws and regulations and with different guidelines, laws and regulations of countries where studies are being performed.

# Annex 1: Data management and strategy survey WP1

## 1. Data summary

**What is the purpose of the data collection/generation and its relation to the objectives of the project?**

*To develop methods and access to observational data for better harmonized and standardized generation of evidence on drug safety in pregnancy and perform demonstration studies*

**What types and formats of data will the project generate/collect?**

Four types of data source are foreseen

1) EUROmediCAT registries (registries of congenital anomaly with information on medication exposure in the first trimester of pregnancy)
2) Electronic health record data, education data and civil registration data

3) Prospective cohort data collected for research purposes

4) Population-based registries other than EUROCAT, e.g. Cerebral Palsy registers, cancer registries)

5) Linkages between one or more of the above categories e.g. euromedicat registries linked to electronic healthcare data, or the linked Scandinavian electronic healthcare data.

Note: Prospective cohorts and pregnancy registries which are collected for the sole purpose of medication safety monitoring will be captured by WP2

The data types to be collected include:

- Metadata to describe databases (in a data source catalog)
- Aggregate data to be held on a data platform, resulting from data analysis

N.B. Individual data will be held within the contributing data source institutions only.

**What is the origin of the data?**

1) Data included in EUROmediCAT registries:

- Population-based registries for surveillance of congenital anomalies

2) Electronic health record data

- Data captured for medical recording or billing purposes in the process of routine care

3) Prospective cohort data set up for research purposes for example

- Rotterdam Study
- Danish Cohort
- Norwegian Cohort

4) Registries other than EUROCAT – for surveillance and research purposes

**Will you re-use any existing data and how?**

Yes, for items 1-4 all data is about re-use of data collected for routine care or specific research. Re-use will comprise

1) Use of individual level anonymized data on a local secure platform
2) Distributed analytics of data transformed locally in a CDM

**What is the expected size of the data?**

1) Not clear, but results will be shared only and should not be big in terms of Bytes

**To whom might it be useful ('data utility') in the project and after wards?**

1) Methods development for analysis and determining optimal methods and study design approaches in different scenarios within ConcePTION: will have long term value for many stakeholders – researchers, HCPs, patients, regulators, industry.
2) Enables accessible harmonized data repository to study medication safety in pregnancy and lactation for all stakeholders wanting to give advice about drug use in women
3) The results of the Demonstration projects will provide medication safety evidence to be published in peer reviewed journals and will have immediate value to stakeholders.

**2. Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|
| EUROmediCAT Registries - unlinked | University of Ulster, on behalf of individual registries who give specific permission | EUROmediCAT registry partners, DAPs | Human | Use of unlinked data and linkage of data with identifying link requires an approval. Unclear whether ethical approval would be granted for ConcePTION as a whole or on a study-by-study basis. Some registries have individualized consent but most do not. | No, only used at DAP levelbut not released. |

| | | | | | |
|---|---|---|---|---|---|
| Electronic health record data | Various – | DAPs | Human | Review by governance boards, Consent not typically required | Data is typically anonymized at point of use. |
| Prospective Cohort data | Various | DAPs | Human | Individual cohorts have individual consent requirements | |
| Registries outside of EUROmediCAT | Various | DAPs | Human | Various | |

**3. Please describe where the collected data will stored for each task and type of data collection**

Individual level data to be stored locally. Aggregate data will be made available to specific partners on the ConcePTION Platform.

**4. Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|
| EUROmediCAT Registries | EUROmediCAT registry data will be maintained in the | EUROmediCAT/ DAPs | Structural transformation will be conducted locally and | | Semantic harmonization by WP7 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | EUROCAT CDM | | semantic harmonization will be done by WP7. SAP will be generated by WP1 | | | |
| Electronic health record data | | DAPs | | | Semantic harmonization by WP7 | |
| Prospective Cohort data | | DAPs | | | Semantic harmonization by WP7 | |
| Registries outside of EUROmediCAT | Conversion to the lower level CDM | DAPs | | | Semantic harmonization by WP7 | |

Please explain if you wish:

- Structural transformation will be conducted locally and semantic harmonization (harmonization of content) will be done by WP7. WP1 will describe in the Statistical analysis plan for each demonstration project the variables for the level 2 CDM
- WP1 has a task to define algorithms for exposures and outcomes in the demonstration projects together with task 7.7
- People locally will convert data to the low-level CDM, WP7 will work with databases to help accomplish this.
- Each study will require a unique data analysis table, which WP1 will be defining together with WP7.
- Standardization and characterization of databases is done by WP7 while study-specific variables will be defined in collaboration with task 1.3 and the definitios task force

**5. How will you conduct quality control?**

| Type of data | Type of analyis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| EUROmediCAT Registries | | EUROmediCAT protocol regarding quality control | WP1 DP leaders | |
| Electronic health record | | | | |

| | | | | |
|---|---|---|---|---|
| data | | | | |
| Prospective Cohort data | | | | |
| Registries outside of EUROmediCAT | | | | |

**6. Governance please complete for each type of data**

| Type of data | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| EUROmediCAT Registries | ENCePP | Secure local platform | | |
| Electronic health record data | ENCePP | Secure local platform | | |
| Prospective Cohort data | ENCePP | ?? | | |
| Registries outside of EUROmediCAT | ENCePP | Secure local platform | | |

**For aggregate data, note that there are disclosure restrictions on cells with small numbers, and aggregate data formats also need approved therefore.**

**Please feel free to provide comments**

# Annex 2: Data management and strategy survey WP2

1. **Data summary**

**What is the purpose of the data collection/generation and its relation to the objectives of the project?**

*To develop methods and access to reported pregnancy data for better harmonized and standardized generation of evidence on drug safety in pregnancy*

**What types and formats of data will the project generate/collect?**

Four types of data are foreseen

3) Available data from publicly available international spontaneous reporting systems: EUDRAVIGILANCE, FAERS, VAERS, Vigibase
4) Available spontaneous reporting data from pharmacovigilance centers locally
5) Available Pregnancy registry data & prospective cohort datasets
6) Newly collected data in ConcePTION on neurodevelopmental outcomes

**What is the origin of the data?**

Ad 1) Available data from publicly available international spontaneous reporting systems: EUDRAVIGILANCE, FAERS, VAERS, Vigibase

- EUDRAVIGILANCE: spontaneous reports sent to EMA (small molecules and biologics), from marketing authorization holders, patients, Health care professionals, pharmacovigilance centers, lawyers. Regarding all products licensed in the European Union
- FAERS: spontaneous reports sent to FDA on drugs from marketing authorization holders, patients, Health care professionals, pharmacovigilance centers. Regarding all products licensed in the USA
- VAERS: spontaneous reports sent to US-CDC on vaccines from marketing authorization holders, patients, Health care professionals, pharmacovigilance centers. Regarding all products licensed in the USA
- VIGIBASE: spontaenous reports of suspected adverse drug reactions, collected by national drug authorities in over 110 countries

Ad 2) Available spontaneous reporting data from pharmacovigilance centers locally

- LAREB: spontaneous reports including narratives, from NL potentially other pharmacovigilance sites

Ad 3) Pregnancy registry data & cohorts

- ENTIS reported pregnancy exposure & follow-up data, at different ENTIS sites in Europe
- Data collection from clinical research groups (i.e. RA, SLE, MS patients) at different institutions_
- Regulatory mandated pregnancy exposure registries at companies, based on CRO collected data
- Pregnancy & birth cohorts (independent of drug exposure), with patient reported data and follow-up (www.birthcohorts.net)

Ad 4) Cohort of women and children followed-up after pregnancy exposure specifically for ConcePTION to test new neurodevelopmental outcomes

**Will you re-use any existing data and how?**

Yes, for items 1-3 all data is about re-use of data collected for routine care, surveillance or specific research. Re-use will comprise

3) Use of individual level anonymized data on a secure platform
4) Distributed analytics of data transformed locally in a CDM

**What is the expected size of the data?**

1) Spontaneous reporting data publicly available (not restricted to pregnancy)

- EUDRAVIGILANCE: .15 million reports
- VIGIBASE: 20 million reports
- FAERS: 17 million
- VAERS: 350,000

Updates will be made during the course of ConcePTION

2) LAREB reports: 3000
3) Pregnancy registries & cohorts: unclear
4) New data: unclear

**To whom might it be useful ('data utility') in the project and after wards?**

4) Methods development for analysis: useful for ENTIS, Conception, pharmacovigilance centers, EMA, FDA, MAH, CDC, research groups
5) Accessible harmonized data repository on pregnancy reports will be useful for all stakeholders wanting to give advice about drug use in women

**2. Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|
| Int. Publicly available datasets | LAREB | UMCU, GSK??, EMA ??? | human | EV: application EMA<br>VIGIBASE: request to WHO-UMC<br>FAERS: none<br>VAERS: None | No |
| LAREB data | LAREB | | human | Non/internal governance | Yes but can be anonimized |
| Prosp.cohort data | ENTIS | ENTIS sites | human | Review of consent applicability | Yes but can be anonimized or accessed through CDM |

| | | | | | |
|---|---|---|---|---|---|
| | | | | Review by governance boards | |
| Regulatory mandated pregn. exposure | GSK/Other EFPIA depending on accessibility | Relevant EFPIA companies | human | Approval by companies, freedom of information | Can be anonimized |
| New cohort | ??University of Manchester (if Rebecca) | ?? | human | Ethics Board consent | |

**3. Please describe where the collected data will stored for each task and type of data collection**

*Please make a data localization and format organogram.*

**4. Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|
| FAERS/EUDRAVIGILANCE | Extraction Deduplication, completion mapping | Luc, HJ, Caitlin | Transformation into CDM | R | LAREB/UMCU | Public SAP Publicly available R code in repository |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | (WP7) |
| Prospective cohort data | unknown | | Transformation into CDM | R | ? | Public SAP Publicly available R code in repository (WP7) |
| LAREB data | None additional | LAREB | None | None | LAREB | |
| Newly collected data neurodevelopmental | unknown | Unknown | unknown | unknown | unknown | |
| | | | | | | |
| | | | | | | |

Please explain if you wish……

**5. How will you conduct quality control?**

| Type of data | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| Int. Publicly available datasets | Completeness description (indicators)<br>Signal detection | Yes (indicator protocol T7.6)<br><br>Yes (demonstration projects) | LAREB/UMCU…… | Review by second group<br>Double coding |
| LAREB data | Manual reading of | all exported reports | LAREB | to be discussed |

| | | | | |
|---|---|---|---|---|
| | narratives, deduplication | compatible with R3 and EMA validation rules | | |
| Prosp.cohort data | Description of completeness of data (indicators)<br>Signal detection | Yes (indicator protocol T7.6)<br>Yes (demonstration projects) | …….. | |
| Regulatory mandated pregn. exposure | Description of completeness of data (indicators) | Yes (indicator protocol T7.6) | | |
| New cohort | Study | Yes (demonstration project) | ? | ? |

**6. Governance please complete for each type of data**

| Type of data | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| Int. Publicly available datasets | ENCePP | Secure platform (WP7) | Secure platform (ISO/GDPR), UMCG | UMCG |
| LAREB data | ENCePP | Secure platform | --to be discussed-- | p.vanderhorst@lareb.nl |
| Prosp.cohort data | | | | |
| Regulatory mandated pregn. exposure | | | | |
| New data | | | | |

**Please feel free to provide comments**

# Annex 3: Data management and strategy survey WP3

## 1. Data summary

| |
|---|
| **What is the purpose of the data collection/generation and its relation to the objectives of the project?**<br>To develop, characterise, validate and apply a non-clinical testing platform for reliable prediction of drug concentrations in human breast milk (and plasma; milk/plasma ratio) along with systemic drug exposure in breastfed infants. |
| **What types and formats of data will the project generate/collect?**<br>*In silico* data, *in vitro* cell culture data and *in vivo* animal data<br>Clinical data to be used will come from WP4<br>DOC, XLS, CSV, PDF, TXT, PPT, JPG, ZIP, PNG, EPS, AVI, MPG |
| **What is the origin of the data?**<br>Experimental *in silico*/*in vitro*/*in vivo* animal data |
| **Will you re-use any existing data and how?**<br>Yes, literature data, data obtained in related external projects (e.g. opportunistic sampling in infants), and clinical data generated in WP4 |
| **What is the expected size of the data?**<br>1Tb |
| **To whom might it be useful ('data utility') in the project and after wards?**<br>All that need to predict drug exposure in human milk to inform risk assessment on medication use during breastfeeding (e.g. Health Care Industry: pharmaceuticals, physicians; Health Authorities, Patients, Academic Researchers, etc...). |

2. **Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Task | Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|---|
| 3.1 | Literature lactation data | UNIBO, TEVA | KUL, NVS, Lilly Covance Ellegaard, BioNotus | Human and animal | None | No |
| | Literature data from existing in vitro/in silico models | UNIBO, TEVA | KUL, NVS, Lilly | Human and animal | None | No |
| 3.2 | In vitro animal and human cell culture data | KUL, NVS | UNIBO, BioNotus, Ellegaard, Covance | Human and animal | None (commercially available cell lines) | No |
| 3.3 | In vivo experiments animal data | UNIBO, Ellegaard | BioNotus, Covance, Lilly, NVS, Teva | Animal | Yes: animal ethics committee | No |
| 3.4 | In silico data generation | KUL, NVS | BioNotus, UNIBO, CHUT, UNIGE, Teva | From in vitro data described in 3.2 and In vivo animal data in 3.3 | No | No |
| | Clinical data from literature studies, related external studies (outside ConcePTION) and/or WP4 No data collection in WP3. | KUL, NVS | BioNotus, UNIBO, CHUT, UNIGE, Teva | Human | No (literature) Ethics commitee (via WP4) | No Maybe yes (either anonymized but better coded) |

| Task | Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|---|
| 3.5 | Literature data on predicted infant drug exposure | KUL, NVS | UNIBO, UNIGE, BioNotus, Teva | Human and animal | No (literature) | No |
| | Infant data may become available via related external studies (outside ConcePTION).<br>No infant data collection is planned in WP3/WP4 | KUL, NVS | UNIBO, UNIGE, BioNotus | Human | • Ethics committee (via WP4)<br>• 'External' data collection should have been approved by EC. | Maybe yes (either anonymized but better coded) |
| 3.6 | No data collection | KUL, NVS | CHUT, UNIBO, UNIGE, UOSL, BioNotus, Teva | Human | Yes (via WP4) | Maybe yes (either anonymized but better coded) |
| 3.7 | No data collection (generation of protocol) | UNIBO, Covance | CHUT, UNIGE, KUL, Lilly, NVS, Teva, Ellegaard | Animal | No | No |
| 3.8 – 3.9 | No data collection | UNIGE, NVS | UOSL, CHUT, KUL, Teva | None | No | No |

Please comment……..

**3. Please describe where the collected data will stored for each task and type of data collection**

| Task | Type of data to be collected | Where (at which physical location) is primary (original) data stored? | What software is used for storage of data? | What format and type of data standards will you use to store the data? | How will you make the primary data accessible to other partners in consortium. Are there restrictions? |
|---|---|---|---|---|---|
| 3.1 | Literature lactation data | At partners location – the cloud (Sharepoint) | WP3 common space on ConcePTION Website *[Endnote or equivalent (software should support easy exchange between partners)]* | Text in PDF, DOC or XLS | WP3 common space on ConcePTION Website |
| | Literature data from existing in vitro/in silico models | At partners location – the cloud (Sharepoint) | WP3 common space on ConcePTION Website *[Endnote or equivalent (software should support easy exchange between partners)]* | Text in PDF, DOC or XLS | WP3 common space on ConcePTION Website |
| 3.2 | In vitro animal and human cell culture data | At KUL and UNIBO location – the cloud (Sharepoint) | Microsoft Excel<br>R<br>Graphpad Prism | XLS<br>CSV<br>PZF | WP3 common space on ConcePTION platform |
| 3.3 | In vivo experimental animal data | • At UNIBO location – the cloud (Sharepoint)<br>• At Covance location | • Microsoft Excel<br>R<br>Graphpad Prism?<br>• GLP commercial data collection system | • XLS<br>CSV<br>PZF<br>• Word, PDF | WP3 common space on ConcePTION platform |
| 3.4 | In silico data generation | At KUL and UNIBO location – the cloud | Microsoft Excel<br>R | XLS, CSV, TXT and software-specific | WP3 common space on ConcePTION platform |

| Task | Type of data to be collected | Where (at which physical location) is primary (original) data stored? | What software is used for storage of data? | What format and type of data standards will you use to store the data? | How will you make the primary data accessible to other partners in consortium. Are there restrictions? |
|---|---|---|---|---|---|
| | | (Sharepoint) | Simcyp<br>PK-SIM<br>NONMEM | formats | |
| | Clinical data from literature, related external studies and/or WP4 | At partners location – the cloud (Sharepoint)<br>Storage on common space on ConcePTION Website | WP3 common space on ConcePTION Website<br>*[Endnote or equivalent (software should support easy exchange between partners)]* | TXT, CSV, DOC | WP3 common space on ConcePTION website |
| 3.5 | Literature data on predicted infant drug exposure | At partners location – the cloud (Sharepoint) | WP3 common space on ConcePTION Website<br>*[Endnote or equivalent (software should support easy exchange between partners)]* | Text in PDF, DOC or XLS | WP3 common space on ConcePTION Website |
| | No infant data collection in WP3/WP4. | NA | NA | NA | NA |
| 3.6 | No data collection | NA | NA | NA | NA |
| 3.7 | No data collection (generation of protocol) | NA | NA | NA | NA |
| 3.8-3.9 | No data collection | NA | NA | NA | NA |

*Please make a data localization and format organogram*

**4. Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| 3.1 | Literature lactation data | As defined by the process of systematic literature review: removal of irrelevant data, duplicates, standardization | 3.1 participants | NA | NA | NA | NA |
| | Literature data from existing in vitro/in silico models | Removal of irrelevant data, duplicates, standardization | 3.1 participants | NA | NA | NA | NA |
| 3.2 | In vitro animal and human cell culture data | Exclusion of data from experiments not passing quality indicators | KUL, UNIBO, BioNotus | Calculation of rates, clearances; dealing with BLQ values; statistical analysis, modelling | Excel, R, Graphpad Prism | KUL, UNIBO , BioNotus | NA |
| 3.3 | In vivo experimental animal data | Exclusion of data from experiments not passing | UNIBO | Calculation of rates, clearances; dealing with BLQ values; | Excel, R, Graphpad Prism | KUL | NA |

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| | | quality indicators | | statistical analysis, modelling | | | |
| 3.4 | In silico data | Data transformation as part of data processing | KUL, BioNotus | Simulation-based outputs | Modelling and simulation software (R, Simcyp, PK-SIM, NONMEML) | KUL, BioNotus | NA |
| | Clinical data from literature, related external studies and/or WP4 | removal of irrelevant data, duplicates, standardization | KUL | NA | NA | NA | NA |
| 3.5 | Literature data on predicted infant drug exposure | removal of irrelevant data, duplicates, standardization | KUL | NA | NA | NA | NA |
| | No data collection (data from WP4) | Data transformation | KUL | NA | NA | NA | NA |
| 3.6 | No data collection | NA | NA | NA | NA | NA | NA |
| 3.7 | No data collection (generation of | NA | NA | NA | NA | NA | NA |

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| | protocol) | | | | | | |
| 3.8-3.9 | No data collection | NA | NA | NA | NA | NA | NA |

Please explain if you wish……

**5. How will you conduct quality control?**

| Task | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| 3.1 | Literature lactation data | Yes | NA | NA |
| | Literature data from existing in vitro/in silico models | Yes | NA | NA |
| 3.2 | In vitro animal and human cell culture data | Yes | KUL | The QC will be done by an independent scientist |
| 3.3 | In vivo experimental animal data | Yes | UNIBO, Ellegaard | The QC will be done by an independent scientist |
| 3.4 | In silico data | Yes | KUL | The QC will be done by an independent scientist. The outputs of independently developed models (2 scientists) will be compared |
| | Clinical data from literature, | Yes | KUL | NA |

| Task | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| | related external studies and/or WP4 | | | |
| 3.5 | Literature data on predicted infant drug exposure | Yes | KUL | NA |
| | Infant data may become available via related studies (outside ConcePTION).<br>No infant data collection is planned in WP3/WP4. | Yes | KUL | The QC will be done by an independent scientist |
| 3.6 | No data collection | NA | NA | NA |
| 3.7 | No data collection | NA | NA | NA |
| 3.8-3.9 | No data collection | NA | NA | NA |

## 6. Governance please complete for each participating partner

| Partner | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| **KUL** | ?? (In vitro cell culture and<br>PBPK modelling) | KUL Server with Sharepoint security protected (2-factor authentication) | KUL server (behind university firewall)<br>ConcePTION common website | Toon Boon |
| **UNIBO** | ?? (In vitro cell culture and in vivo data) | | | |
| TEVA | NA | | | |
| Novartis | NA | | | |
| Lilly | NA | | | |
| **Covance** | In vivo data: GLP-like | | | |

| Partner | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| | status | | | |
| Ellegaard | NA | | | |
| **BioNotus** | ?? BA assays | | | |
| UOSL | NA | | | |
| CHUT | NA | | | |
| UNIGE | NA | | | |

**Please feel free to provide comments**

# Annex 4: Data management and strategy survey WP4

## 1. Data summary

**What is the purpose of the data collection/generation and its relation to the objectives of the project?**

Assessment of drug (and active metabolites) concentration values in milk of breastfeeding women as part of the overall objective of ConcePTION to demonstrate feasibility of performing research with samples collected at the BBMRI breast milk biobank and bioanalytical centre.

**What types and formats of data will the project generate/collect?**

Bioanalysis data. Rawdata are reported in form of text file (.txt) and as excel file (.xls). The processing data will be in format of MassLynx TargetLynx or any other software used for processing of mass spectrometry Rawdata (alternative in Thermo instrument).
The popPK modelling data and models will be processed in specific software (NONMEM, MonoLix or any other).
The reports will be in word format (.docx), and final versions is Acrobat Reader format (.pdf file).

Please specify if the data will be in SAS datasets
Please specify if the preferred software used for PK analyses and modeling
Please indicate number of users for the ConcePTION database UPPS – 8.

Master protocol
Informed Consent Form Master Template

Clinical trial data from collection and analysis of blood and breastmilk from women – data from each clinical sites conducting breast milk sampling (stored at UMCU, UPPS at Biobank)

- Study sites information
- Study personnel information

- Protocol
- Informed Consent
- Case report form
- Ethics Board Approvals
- Screening information
- Inclusion/exclusion data
- Woman's disease or condition
- Medication dose, frequency and exposure
- Concomitant medications
- Adverse events
- Subject disposition
- Protocol deviations
- Demographic data
- Date of birth of the child
- Gestational age of the child
- (Weight and height of mother
- Weight of baby
- Time of last dose of studied medication
- Date and time of sampling for milk and blood
- Volume of breast milk from which sample was taken
- Collection Method

Samples stored (samples shipped to and stored at UPPS Biobank)

- Name of site
- Type of sample (plasma/breast milk)
- Date of sample collection
- Date of shipment to biobank
- Date of receipt to biobank
- Noticeable problems at reception
- Specifications met for shipment (yes/no according to IATA recommendations)
- Notification that samples are frozen when received
- Storage temperature of samples
- Processing time
- Deviations from protocol
- Storage temperature at site

- Freeze-Thaw-Cycles
- Storage time before analysis/withdrawal
- Processing times
- Records of measuring instruments
- Records of methods validation
- Records of the maintenance of the equipment

Data on storage at UPPS (data stored at UPPS Biobank))

- Date and time of reception
- Freezer location
- Storage temperature (also with temperature alarms on freezers)
- Freeze-Thaw-Cycles
- Date and time of of retrieval
- Withdrawals for analysis
- LIMS data – all biobank data will be in LIMS and traceability of samples and shipments will be managed in the LIMS

Bioanalytical data (Data stored at UPPS)

- Drug concentration value in breast milk
- Drug concentration value in plasma from mother
- Drug concentration value in plasma/blood from infant
- Analytical methods validation results
- Quality data

PK data (Data stored in ConcePTION database)

- $AUC_{\tau}$: Area under the curve over a dosing interval
- Cav: Average concentration over a dosing interval, equal to $Cav=AUC_{\tau}/\tau$
- Cmax: Maximum observed drug concentration
- tmax:Time of the maximum observed concentration
- $\lambda z$: First-order terminal elimination rate constant, calculated from a semi-log plot of the milk (plasma) concentration vs time curve
- t1/2: First-order terminal elimination half-life, calculated as $0.693/\lambda z$

Population PK modeling data (Data stored in ConcePTION database)

| |
|---|
| Calculated infant dose and relative infant dose (Data stored in ConcePTION database) |
| **What is the origin of the data?**<br><br>Pregnant and breastfeeding women from 5 studies of different drugs |
| **Will you re-use any existing data and how?**<br><br>Data from medical records will collected at clinical sites. |
| **What is the expected size of the data?**<br><br>25 terabytes |
| **To whom might it be useful ('data utility') in the project and after wards?**<br><br>For monitoring purposes regarding pharmacovigilance, for academic research and for drug companies.<br><br>For physicians and patients to understand medication exposures through breast milk. |

**2. Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Task | Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|---|
| 4.1 | Protocol for collection and storage | UPPS Biobank (storage) and ULAUS (collection) | BBMRI, UCB, NVS | Protocol | None | No |
| 4.2 | Protocol and SOP, bioanalytical methods data, PK profiles, popPK models | UPPS | ULAUS, CHUT, UOSL, UCB | Protocol, popPK models, reports | None | No |
| 4.3 | Patient data, data of breast milk and blood, sample values, standard curves, sample collection details | ULAUS | UPPS, UOSL, CHUT | Patient data, data from PK analysis | Ethical review for sampling at each collecting site. Ethical review for biobanking and analyses | Yes |
| 4.4 | 1. Standard documents (e.g. process descriptions, SOP, | BBMRI-ERIC | Uo Leipzig, Wroclaw MU, UPPS, ULAUS, CHUT, UOSL | 1. process descriptions, SOP, checklists, forms, questionnaire | 1.None<br>2.None<br>3.None | 1.No<br>2. Yes, anonymised needed<br>3. No |

| | | | | | | |
|---|---|---|---|---|---|---|
| | checklists, forms, questionnaire)<br>2. Verification documents (e.g. records)<br>3. QMS documents | | | 2. records<br>3. documented information about QM Structure preferably in a digital format | | |

Please comment……..

**3. Please describe where the collected data will stored for each task and type of data collection**

| Task | Type of data to be collected | Where (at which physical location) is primary (original) data stored? | What software is used for storage of data? | What format and type of data standards will you use to store the data? | How will you make the primary data accessible to other partners in consortium. Are there restrictions? |
|---|---|---|---|---|---|
| 4.1 | See above | Uppsala | LabWare, LIMS, Socrates, NONMEM | | Only pseudonymized data will be shared |
| 4.2 | See above | Uppsala | Text (.txt), excel (.xls), Acrobat Reader (.pdf) files, NONMEM (Monolix?) | Rawdata, Processed analytical data, PK parameters, popPK models, Reports | ConePTION consortium members will have access to the data |
| 4.4 | See 2. / task 4.4 | Uppsala / BBMRI-ERIC | MS Office Software, software of the new database? | | |

*Please make a data localization and format organogram*

4. **Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| 4.2 | Bioanalysis of breast milk and plasma samples, Population PK | | | | | | |

Please explain if you wish……

5. **How will you conduct quality control?**

| Task | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate verifying data transformation & analysis? |
|---|---|---|---|---|
| | | Yes | Study PI | |
| 4.2 | Bioanalysis of breast milk | Yes, regulatory guidance | Study PI | |

| | | | | |
|---|---|---|---|---|
| | and plasma samples, Population PK | (FDA and EMA), best practice | | |
| 4.4 | Monitoring and updating the standard documents | Yes | Version control by BBMRI-ERIC | Version control done by BBMRI-ERIC, outdated versions will be abrogated by BBMRI-ERIC |

## 6. Governance - please complete for each participating partner

| Partner | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| All | ConcePTION CoC | | | |

**Please feel free to provide comments**

# Annex 5: Data management and strategy survey WP5

1. **Data summary**

| |
|---|
| **What is the purpose of the data collection/generation and its relation to the objectives of the project?**<br>Get end users input and insights on needs, experience and preferences for information about drug use during pregnancy and lactation |
| **What types and formats of data will the project generate/collect?**<br>1. Ad hoc surveys and focus groups, literature reviews<br>2. Knowledge bank |
| **What is the origin of the data?**<br>1. Newly collected data through surveys or focus groups<br>2. Literature references and synthesis of evidence in knowledge databases developed by local TIS |
| **Will you re-use any existing data and how?**<br>TBD (information from existing knowledge databases) |
| **What is the expected size of the data?**<br>1.Unclear but expected to be limited (Surveys in thousands of end users, focus groups of limited size)<br>2. Unclear (note: check with existing KB e.g. Lareb) |
| **To whom might it be useful ('data utility') in the project and after wards?**<br>2. Accessible harmonized repository on evidence regarding drug use in pregnancy and breastfeeding will be useful for all stakeholders wanting to give advice about drug use in women |

2. **Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Task | Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|---|
| 5.1.1 | surveys | Newcastle | EFPIA, ENTIS, ? | | ? | Yes? |
| 5.1.3 | surveys | Synergist | EFPIA, ? | | ? | Yes? |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5.1.3 | Focus group | Lareb | EFPIA, ? | | Local ERB | Yes? |
| 5.2 | Literature, labelling, regulatory documents, study results, etc | Lareb | Orcion, ENTIS, | | none | no |

Please comment……..

3. **Please describe where the collected data will stored for each task and type of data collection**

| Task | Type of data to be collected | Where (at which physical location) is primary (original) data stored? | What software is used for storage of data? | What format and type of data standards will you use to store the data? | How will you make the primary data accessible to other partners in consortium. Are there restrictions? |
|---|---|---|---|---|---|
| 5.1.1 | Survey results | ? | ? | ? | |
| 5.1.3 | Survey results | ? | | | |
| 5.1.3 | Focus groups summary | ? | | | |
| 5.2 | Summary of evidence + list of references | ? | | | |

*Please make a data localization and format organogram*

4. **Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

**Not Applicable**

| Task | Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

Please explain if you wish……

5. **How will you conduct quality control? (NA)**

| Task | Type of analysis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| | | | | |

6. **Governance please complete for each participating partner**

| Partner | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| 5.1.1 | | | | |
| 5.1.3 | | | | |

**Please feel free to provide comments**

# Annex 6: Data management and strategy survey WP7

1. **Data summar**

| |
|---|
| **What is the purpose of the data collection/generation and its relation to the objectives of the project?**<br>To collect information on ethical hurdles to share data from pregnant women and data access providers, to assess governance issues and information, to characterize datasources and to assess whether they are fit for purpose to provide evidence on drug safety in pregnancy |
| **What types and formats of data will WP7 generate/collect?**<br>**Task 7.1**: **Reward models**: ethico legal documents from data access providers, and interview data from DAPs participating in the ConcePTION project<br>Format: mp3, .docx<br>**Task 7.2 Governance**: codes of conduct for analysis of different types of data that are used in ConcePTION<br>Format: .docx<br>**Task 7.3: Ethics:** interview data from pregnant women who consent to participate<br>Format: mp3, .docx<br>**Task 7.4: Catalogue:** non-curated publicly available information on datasources (EURoMediSafe), curated data on organizations, responsible persons and datasources<br>Format: .Xls; Catalogue (Molgenis)<br>**Task 7.5: Common data models**: common data models for use by WP1 and 2<br>Format: .docx, .xls<br>**Task 7.6: Data characterization:** approaches all the DAPs for WP1 and 2 that agree to participate in data characterization, data collected will comprise indicators of the data (aggegrate). Event definition templates to find codes to extract events /data/ Simulated data to develop programs<br>Format: .csv .docx<br>**Task 7.7: Algorithm validation:** Will assess impact of the use of different algorithms to extract events and evaludate novel validation manners<br>Format: .csv .docx<br>**Task 7.8: Double coding of scripts: in this task codes for distributed analyses will be created and hosted.** |

Simulated data for programming
Format: .csv ; .txt; R, GitHUB repository
**Task 7.9: Reprotox data**: collection of reprotoxicology data from rats abd rabbits extracted from regulatory submission
Format: .csv; .docx
**Task 7.10:** Ethics issues: collection/repository of documents
Format: .docx; .pdf

**What is the origin of the data?**
Task 7.1: Data access providers in the consortium as partners or third parties
Task 7.2: Existing codes of conduct that are publicly available or owned by consortium members
Task 7.3: Pregnant women
Task 7.4: EuroMEdiSAfe (data from internet) (non-curated), survey data with DAPs partners in consortium (curated)
Task 7.5: Common data model development by WP7, 1 and 2, based on prior CDM and those in Sentinel, OMOP,Eurocat
Task 7.6: Data access providers that participate in the data characterization & algorithm validation/
Task 7.7: Data access providers that participate in the data characterization & algorithm validation
Task 7.8: Tools written by the WP7 participants
Task 7.9: Data collected by CBG and RIVM
Task 7.10: Documents generated by different WP and organizations

**Will you re-use any existing data and how?**
Yes, for the data characterization we will access data that is available at the sites of the data access providers. This will be done in a distributed manner. Data will stay local and data characterization scripts will be sent to the DAPs, these scripts will generate results that will be sent to the secure platform

**What is the expected size of the data?**
2) WP7 anticipates that the data that is shared as results will be small in terms of bytes
3) Simulated data will be hosted on the platform to develop scripts

**To whom might it be useful ('data utility') in the project and after wards?**
6) WP7 will support methods development and conduct of demonstration studies by WP1 and 2
7) WP7 will also systematically characterize datasources and assess whether they are fit for purpose, this will increase the transparency and potentially take away criticisms about quality of RWE

**2. Please describe which partners will be collecting some type of data in your WP, and link to the tasks. Please be specific for each type of data collection**

| Type of data to be collected | Responsible partner | Collaborating partners | Type of species (specify) | What type of ethical review is needed? | Do you collect identifiable information? |
|---|---|---|---|---|---|
| 7.1 Interviews | UMCU | All DAPs in consortium as partners or TP | Human | None | Yes, DAP name |
| 7.2 Code of conducts/guidance | iHD | UPPS, GSK UMCU…. | NA | NA | no |
| 7.3 Ethics interviews | UMCU | EIWH, EFGCP | Human | Ethics board, UMCU | Yes (consented) |
| 7.4 Catalogue details | UMCG | BBMRI, all DAPs, J&J, GSK | Human / datasources | None | Yes, DAP name, contact details and privacy officer names |
| 7.5 Results of analyses on health data using common data model | UMCU | ARS, UMCG, BBMRI, GSK, J&J All DAPs and WP1/2 partners | Human (aggregated data) | Governance review at each site of protocol | No (only aggregated results, coded/anonymized) |
| 7.6 Data characterization | UMCU | ARS, GSK, J&J All DAPs | Human | Governance review at sites in study team | No (only aggregated results, coded/anonymized) |
| 7.7 Algorithm validation | ARS | UMCU, GSK, J&J, DAPs, WP1 | Human | Governance review at sites in study team | No (only aggregated results, coded/anonimized) |
| 7.8 Double coding | UMCU | ARS, J&J, | Human/simulated data | None | NA |
| 7.9 Reprotoxicology | RIVM | CBG, NVS, GSK, UMCU | Rabbit & Rats | NA | Re-use of data |

**3. Please describe where the collected data will stored for each task and type of data collection**

| Type of data to be collected | Responsible partner | Collaborating partners | Storage of data |
|---|---|---|---|
| 7.1 Interviews | UMCU | All DAPs in consortium as partners or TP | Original data Local network UMCU, summary results ConcePTION shared website |
| 7.2 Code of conducts/guidance | iHD | UPPS, GSK UMCU…. | ConcePTION shared website member area |
| 7.3 Ethics interviews | UMCU | EIWH, EFGCP | Original data local network UMCU, summary results ConcePTION shared website |
| 7.4 Catalogue details | UMCG | BBMRI, all DAPs, J&J, GSK | BBMRI server, Catalogue is publicly available except negotation part |
| 7.5 Results of analyses on health data in DP | UMCU | ARS, UMCG, BBMRI, GSK, J&J All DAPs and WP1/2 partners | Aggregate outputs on ConcePTION Platform |
| 7.6 Data characterization | UMCU | ARS, GSK, J&J All DAPs | Aggregate outputs on ConcePTION Platform |
| 7.7 Algorithm validation | ARS | UMCU, GSK, J&J, DAPs, WP1 | Aggregate outputs on ConcePTION Platform |
| 7.8 Double coding | UMCU | ARS, J&J, | GitHUB |
| 7.9 Reprotoxicology | RIVM | CBG, NVS, GSK, UMCU | CBG/RIVM local computers summary data on ConcePTION shared website |
| 7.10 Ethical follow-up | UMCU | All WPs | ConcePTION shared website |

**4. Please describe the process of cleansing, transforming and analysis of the data. Please be specific for each type of data**

| Type of data to be collected | What type of data cleaning is needed? | Who is responsible for data cleaning? | What type of data transformation/ analysis do you anticipate | What software will be used for cleaning, transformation and analysis? | Where/by whom will the analysis be conducted? | What standards will you use for code development / access and re-use? |
|---|---|---|---|---|---|---|
| 7.1 Interviews | Recording and transcribing | UMCU | Summarize and reflect based on ethical framework | Word | UMCU other 7.1 partners | None |
| 7.2 Code of conducts/guidance | None | NA | NA | NA | iHD, GSK | Existing standards/guidances |
| 7.3 Ethics interviews | Recording and transcribing | UMCU | Summarize and reflect based on ethical framework | Word | UMCU | None |
| 7.4 Catalogue details | Review and uploading of information and documents | DAPs, WP1 partners, WP7 | Uploading of data collected in questionnaire format into the catalogue | Excel | UMCG | Open source /Molgenis |
| 7.5 Results of analyses on health data in DP | Quality indicators of data to support Demonstration projects | DAP | Transformation of original data into CDM, transformation of data in CDM into analysis tables And pooling & final analysis of results | Locally into CDM: SAS, SQL, Stata, R<br><br>From CDM into analysis table: R<br><br>From analysis table into final results: WP1/2 partners | Locally ETL from original to CDM: DAP<br>From CDM into analysis table: UMCU, ARS, J&J<br>On platform: Analysis: | ETL will be available on catalogue<br><br>Data transformation script: open source<br><br>Final analysis script: double coded |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | WP1/2 | |
| 7.6 Data characterization | Quality indicators of data | DAP | Transformation of original data into CDM, transformation of data in CDM into analysis tables And pooling & final analysis of results | Locally into CDM: SAS, SQL, Stata, R<br><br>From CDM into analysis table: R<br><br>From analysis table into final results | Locally ETL from original to CDM: DAP<br>From CDM into analysis table: UMCU, ARS, J&J<br><br>On platform: Analysis: UMCU, ARS, DAP | ETL will be available on catalogue<br><br>Data transformation script: open source |
| 7.7 Algorithm validation | Algorithm development, verification of impact | DAP | Analysis of different combinations of components to define event | From CDM into analysis table: R<br><br>From analysis table into final results | From CDM To platform: Analysis: UMCU, ARS, DAP | Data transformation script: open source |
| 7.8 Double coding | None | NA | Double coding of scripts | R | UMCU/ARS, J&J | ENCePP |
| 7.9 Reprotoxicology | Extraction from files | CBG | Extracting and transforming into Excel | Excel | CBG,RIVM | GLP |
| 7.10 Ethical follow-up | None | None | None | NA | NA | NA |

### 5. How will you conduct quality control?

| Type of data | Type of analyis | Will you work according to specific protocol? | Who will create the statistical analysis plan? | How do you anticipate to verify data transformation & analysis? |
|---|---|---|---|---|
| Real world Health data | Quality indicators | Data characterization | UMCU | Double coding |

### 6. Governance please complete for each type of data

| Type of data | What code of conduct will you use in each task | What levels of data security do you have locally? | What level of security will the primary data that you use in ConcePTION be? Please list all | Who are the data privacy officers in each of the participating organizations (e-mails) |
|---|---|---|---|---|
| Real world health data | ENCePP | Secure local and central platform | Authentication to access | Are being identified as part of the interview on ethico-legal issues |