Notes from meeting 02/06/2025

👥 **Attendees**

Fan Zhang

Andreas Markoulidakis

Hayley Jones

Jaroslaw Lang

The meeting mainly focused on the following topics: project equipment updates, progress in data usage and training, introduction to datasets and background information, composition and usage of the SHIELD dataset, and data security and usage regulations.

1. Project Equipment Updates

Fan Zhang mentioned:

- An email regarding the dispatch of the laptop has been received;
- It is currently in the logistics/processing stage and has not yet been actually sent out;
- The contract and system access application process have been completed.

2. Data Usage and Training Progress

Andreas reminded that after receiving the equipment, it is necessary to:

- Set up a Public Health Scotland account;
- Complete the relevant online training for data usage.

Fan Zhang:

- Has completed a training module named “Research GDPR and Confidentiality”;
- Indicates that he will continue with other trainings after the account is activated.

Jaroslaw added:

- Completed all 8 modules on the MRC Learning System;
- Reminded to check whether a screenshot/listing of the required modules has been received and suggested sharing it in the group chat for everyone to stay synchronized.

3. Dataset Description and Background Introduction

Initial communication on datasets:

Andreas suggested providing a brief description of each major dataset to enable Fan to quickly understand the datasets he will be working with.

Jaroslaw indicated that he is drafting the relevant content, but has not yet decided whether to list the dataset names without complete explanations.

Whether to read relevant publications:

Jaroslaw mentioned that each database usually has a publication released by Public Health Scotland (PHS). It still needs to be confirmed whether Fan needs to read these publications to understand each database.

He requested Andreas and Hayley to provide guidance on expectations, considering the tight schedule, whether it is worth investing time to understand these background materials.

Hayley's suggestion:

Hayley emphasized that Fan needs to understand the sources, collection, and management of the data.

She suggested that Jaroslaw briefly introduce the SHIELD data platform as an entry point.

4. About SHIELD and Data Integration Background

- The population of Scotland is about 5 million, with a relatively small geographic area, and health data collection started early. Some data, such as hospitalization records, can date back to the 1960s.
- The types of data cover various aspects such as hospitalization, death, and receiving drug services.

Team's role:

Jaroslaw's team is responsible for collecting statistics on illegal drug use. In Scotland, there are some programs where drug users can obtain drugs for free.

- It includes data from services that provide injection equipment (to prevent the spread of diseases like HIV and hepatitis).

Diverse data sources and integration needs:

- There are multiple datasets recording information about drug users, but from different sources (hospitals, service points, death registries, etc.).
- To build a comprehensive profile, the team has started applying for permissions and conducting data linking from several similar sources after observing information about drug users from a few sources.

Data linking situation:

- The linked datasets cover the period from 2009 to 2023, integrating data from approximately 115,000 individuals.

5. Explanation of “financial Year”

Discussion on the definition of financial year:

Andreas explained that not all countries use the calendar year as the financial year. The financial year in Scotland runs from April 1st to March 31st of the following year.

Decision: To use the calendar year.

Reason: It is simpler to calculate, the calendar year is easier to manage, and it aligns better with previous work.

Study period: 2015 to 2022, based on the project prescription.

6. Data Time Range and Construction Logic

Organizing data by financial year:

- The SHIELD dataset is constructed by fiscal year, from April 1, 2009, to March 31, 2023 (from the 2009/2010 financial year to the 2022/2023 financial year).

Background and objectives of the SHIELD dataset construction:

- The dataset serves as the basis for estimating the prevalence of drug use, including previous related studies and reports.
- The dataset integrates data from multiple sources to form a more comprehensive understanding of drug users.

7. Composition of the SHIELD Dataset (Five Sub-datasets in Total)

SDMD (Scottish Drugs Misuse Database):

- Data on individuals who proactively seek help from drug service institutions.
- From 2015 to the end of March 2021.
- Includes questionnaire information, treatment plans (such as OAT: Opioid Agonist Treatment), etc.
- Up to five types of drugs can be recorded, each with its corresponding route of administration.

DAISy (Drugs and Alcohol Information System):

- The successor system to SDMD, starting from April 2021.
- Collects similar content to SDMD, but the structure and variables may be slightly different.
- Fan needs to pay attention to the differences between the two systems when using the data.
- Up to ten types of drugs can be recorded, with a separate flag for the main drug.

OAT Prescription Data:

- Records prescription information for individuals receiving Opioid Agonist Treatment (OAT).
- Sourced from Scotland's National Prescription Information System.
- Records OAT prescribed by doctors and dispensed by pharmacies.
- Each row of data represents a specific prescription, containing the payment/reimbursement date.
- Infer whether a person is undergoing OAT during a certain period based on these dates.

Hospitalization Data:

- Contains only hospitalization events caused by drug use, not other hospitalization records.
- An individual may have multiple hospitalization events.
- Each record includes:
  - Admission date
  - Discharge date
  - Whether death occurred upon discharge
  - Reason (drug overdose? Cardiovascular event? Other?)
- Analysis challenge: Diverse reasons for hospitalization, need to classify events. The cause analysis of hospitalization data may be overly complex?

Drug-Related Deaths:

- Contains records of deaths related to drug use.
- One record per person, including:
  - Date of death

    - Coding: Primary and secondary causes (e.g., death from stroke caused by drug use)
- Need to determine in the analysis whether to use the primary cause, secondary cause, or both to define “drug-related death.”

➡ The data is linked to form an integrated structure, with each row corresponding to an individual. An individual may appear in all five datasets or only in one, such as in the death records.

8. How to Use These Data for Analysis

Building a study cohort:

- You will receive a data extraction file containing several columns selected from the SHIELD dataset.
- Establish a cohort of “opioid-dependent individuals” based on the OAT prescription data, i.e., all individuals receiving OAT.
- Track events and assess risks for each person:
    - Whether dependent on cocaine (from DAISy)
    - Whether hospitalization occurs (hospital data) to study hospitalization due to drug overdose or other adverse events (such as cardiovascular diseases) and explore how these health events are related to the course and risks of OAT recipients.
    - Further link the death data with individuals in the OAT cohort to analyze the occurrence and patterns of death events within the cohort.

9. Data Security and Usage Regulations

High sensitivity of data (personal health data):

- Involves real individuals who may be identifiable.
- Access and analysis can only be conducted within the Public Health Scotland (PHS) environment:

    - Data cannot be exported to personal computers, university computers, or any non-PHS devices.
    - Even when using the PHS-issued laptop, data cannot be stored on the local hard drive.
    - All operations must be completed through PHS's server remote file system.
- Jaroslaw can assist:
    - Once you obtain the PHS laptop, he will guide you through the technical operations of logging into the server and accessing files.

10. Research Question Structure (PECO Style)

Component	Example for Your Study
Population	Individuals in Scotland who received Opioid Agonist Treatment (OAT) between 2015 and 2022 (i.e., identified as opioid-dependent). This population is derived from the OAT prescription data integrated in the SHIELD dataset.
Exposure	Concurrent use of Cocaine while receiving OAT, i.e., constituting C&O (Cocaine and Opioid) polydrug use. Cocaine use is mainly recorded in the DAISy and SDMD databases.
Comparator	OAT recipients who only use opioids (not using cocaine). That is, individuals who have not been exposed to cocaine use during or after OAT treatment.
Outcome	Drug overdose events (fatal/non-fatal), mortality, OAT retention, and other health

11. Next step plan

- Meeting time arrangement:

  It's best to open it on Friday for one hour every two weeks.

  It's best to have a face-to-face meeting once a month.

  It would be very useful to have the other executive chairman, Matt Hickman, attend as many meetings as possible.
- Clare Jarvis (HPRU in Behavioural Science and Evaluation Admin Mailbox admin-hprubse@bristol.ac.uk) can help inform Matt Hickman the available time.

- Poster deadline and schedule planning:

  Please submit the poster before 12:00 noon on July 18th.

  Invite Andreas and others to the poster conference on the morning of Tuesday, July 22nd.

Order	Time	Name	Programme
intro	9:30	Unit leads	
1	9:40	Amy Stapleton	Medical Statistics and Health Data Science
2	9:50	Fan Zhang	Medical Statistics and Health Data Science

- Start background reading (Andreas and Matthew)

- Wait for your laptop to come from PHS

💻Jaroslaw will provide logistical support for laptops.