Utilizing Finnish nationwide RWD to create an external control arm to a clinical trial

FUTURE CLINICAL TRIALS – FROM TOMORROW TO 2030 – WHY CHOOSE THE NORDICS

Jussi Leinonen
Future Clinical Trials Project
Expected duration: 2020 - 2024

// We develop unique
// patient centric and
// data driven solutions
// to challenges in clinical trials today.

// With a potential to become global innovations

// Our core project team members work in global roles of Bayer’s
// Data Science & Analytics
// Clinical Development Operations
// Oncology Development Operations
// R&D IT
// Integrated Evidence Generation in Medical Affairs
The Main Objectives of the Project (Data Science)

- **Trial 1**: DATA INTEGRATION
  - Building capabilities for utilizing external RWD and internal legacy RCT data

- **Trial 2**: PATTERN RECOGNITION
  - Design and implement data science methods to predict and manage the risks in clinical trials

- **Trial 3**: RWE USE-CASES / EXTERNAL CONTROLS
  - Methodology for providing virtual or synthetic controls
External Control Arm

// Process Flowchart
  // Data Access & Management
  // Data Sources
  // Data Curation & Analysis

// Summary
**Process Flowchart**

**Step 1: Feasibility assessment**
- Assessment of availability of suitable patients and the required variables.

**Step 2: Study preparation**
- Preparation of the study plan in collaboration with the study team

**Step 3: Study permit procedures**
- Preparing and submitting data permit application to Findata

**Step 4: Analysis of the pseudonymized data**
- Data curation
- Matching with RCT data
- Preparation of TLFs defined in the protocol

**Step 5: Analysis of the anonymized data**
- Replication of TLFs and comparison to Step 4 analysis
Data Access & Management

Restricted environment
- Register holders
  - National registers
    - Drug use
    - Health care
    - Deaths
  - Data lakes
    - EMR

Findata environment with limited access
- Bayer / collaborators
  - Pseudonymized raw data
  - Data check and clean
  - Cleaned and checked pseudonymised data
  - Processed pseudonymous data

Bayer environment
- Bayer / collaborators
  - Processed anonymized data
  - Anonymization
  - Processing
  - Merging with RCT data

Results
- Pseudonymized RWD and RCT
- Anonymized RWD and RCT
Data Sources
Electronic health records + National registries

Statistics Finland, Causes of death, 1966–2025

Statistics Finland is a Finnish public authority that collects, combines, and stores data on society, and produces and releases statistics on a wide range of topics.

Statistics Finland – home

Data for FinRegistry: Time and causes of death.

- Examples of variables or variable types:
  - Date of death
  - The basic cause of death
  - The immediate cause of death
  - The contributing cause of death

- Data dictionary

https://www.finregistry.fi/finnish-registry-data
Data Curation & Analysis

Findata Secure Environment (remote connection)

- **RWD Sources**: Real-world data
  - Clinical trial data
  - Clinical database

- **RWD Cohort (SDTM-)**

- **Analysis dataset**

- **TLFs**

Selected:
1. Demographic Information
2. Medical History (predefined terms in focus)
3. Concomitant Medications
4. Laboratory Results & Vital Signs
Mapping into SDTM-

Data model concept

SDTM- datasets consist of 3 set of variables:

• **Main variables** – can be harmonized across all raw datasets, e.g. MHTERM, MHSTDTC.

• **Source identification variables** – variables used to identify source data for the record, e.g. SRCDOM, SRCVAR, SRCID.

• **Supplemental variables** – other variables which are different across source datasets, e.g. additional visit information, type of service. Can be any kind of variables.

Hybrid of a relational (main variables + SRC variables) and document-oriented database (supplemental variables as JSON) which can still be stored as a plain SAS dataset.
Dataflow, broad picture
## Example MH dataset

<table>
<thead>
<tr>
<th>USUBJID</th>
<th>MHTERM</th>
<th>MHSTDTC</th>
<th>...</th>
<th>SRCDOM</th>
<th>SRC VAR</th>
<th>SRC ID</th>
<th>SUPP JSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA-1234567</td>
<td>I12</td>
<td>2017-04-12</td>
<td></td>
<td>HILMO_1996_2020_VISIT</td>
<td>PDG0</td>
<td>17236</td>
<td>{ “serviceprovider”: “R20”, “typeOfContact”: “inpatient”, … }</td>
</tr>
<tr>
<td>AAA-1234567</td>
<td>I34</td>
<td>2019-12-31</td>
<td></td>
<td>AURIA_DIAGNOSES</td>
<td>DIAGNOSIS</td>
<td>213</td>
<td>…</td>
</tr>
<tr>
<td>AAA-1234567</td>
<td>I45</td>
<td>2016-17-12</td>
<td></td>
<td>HILMO_AVOHILMO</td>
<td>ICPC2_1</td>
<td>8136</td>
<td>{ “geoRegion”: 73, “gender”: “Male” }</td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Summary**

- Augmenting clinical trials with external controls can lead to greater patient diversity and a shorter duration of clinical trials with considerable potential savings involved.
  - Estimated 10-20% cost & time savings

- Finland has implemented secondary use of health data legislation and biobank legislation to enable extensive data sharing with an industry friendly approach.

- Nordic national registries, independently or combined with EHRs by unique personal IDs, can be used for reliable long-term tracking of outcome events and medication compliance.

---

**Investigated treatment:** Investigational drug and control drug

**Participants:** 8,255 patients (RWD external control arm)

**RWD Sources:**
- Regional hospital data lake of Southwest Finland via Auria Clinical Informatics (study population identification and data collection); the nationwide healthcare registers — Care Registers for Healthcare (Hilimo and AvoHilmo) by Finnish Institute for Health and Welfare (THL) (data collection); the nationwide cause of death register by Statistics Finland (data collection), and the Prescription Centre and Drug Prescription Registry by Social Insurance Institution of Finland (Kela) (data collection).

**Time span:** 11/2020–

**Funding:** Bayer and Business Finland.

**Partners:** Bayer, MedEngine, Veil.AI

[Link](https://mediabank.businessfinland.fi/l/8DcLn5LqTKWP)