Piloting the European Unified Patient Identity Management (EUPID) Concept to Facilitate Secondary Use of Neuroblastoma Data from Clinical Trials and Biobanking

Ebner Hubert
Introduction

• Neuroblastoma most common cancer in infancy
• Rare disease – about 650 Incidents p.a. USA [1]
• Data collection during clinical trials needs to be done over large networks and many institutions
Secondary Use

• Re-Use of existing datasets from clinical trials
• Combine datasets from different sources
Secondary Use

- Re-Use of existing datasets from clinical trials
- Combine datasets from different sources

Aims

- Provide new findings
- Improve future treatment
Secondary Use

- Re-Use of existing datasets from clinical trials
- Combine datasets from different sources

Aims

- Provide new findings
- Improve future treatment

Hurdles

- Regulatory barriers
- Patient consent

31.05.2016
Preparations for Secondary Use

• Pseudonymisation of Patients - Removing of patient identifiers from datasets

• Provide linkage patients over different trials

• Provide linkage of the accompanying datasets
Pilot Test Setting

Context Clinical Trial (CT)  

Context Bio Bank (BB)

European Unresectable Neuroblastoma (EUNB) trial  
160 Patients

Results from array comparative genomic hybridisation (CGH) analyses from corresponding tumour samples  
48 Patients

31.05.2016
Needs for Secondary Use in the Pilot Test

• Creation of Pseudonyms
  – Pseudonymisation of patients in Clinical Trial (CT)
  – Pseudonymisation of patients in Bio Banking (BB)

• Linkage of patients
• Linkage of accompanying datasets
European Unified Patient Identity Management
EUPID Concept

Context 1 (e.g. Clinical Trial)

First Name, Last Name, Date of Birth

+ Encryption Hashing

De-Identification

PID₁

IDATCRYPT

IDATHASH

• No double Registration
• Re-Identification

Pseudonymisation

PSN₁

EUPID

Same Patient
New Context (e.g. Biobank)

New Pseudonym
for Context 2

PID₂

PSN₂

Context 2

Pseudonym for Context 1

Input (GUI)
accessible (Output)
inaccessible (internal)
The patient has been registered successfully and a new EUPID was generated. No match with already registered patients was found.

Pseudonym: 1099D11A

* Mandatory in this context
EUPID - Client

- Webapp on local client
- Plain text IDAT never leaves client
- Encrypting and phonetic hashing of IDAT done locally

- Only Encrypted and Hashed Values are transferred
EUPID - Pseudonymisation

Context Clinical Trial (CT)

Context Bio Bank (BB)

Patient IDAT

Patient IDAT

PSN-1_Ct
PSN-2_Ct
PSN-3_Ct
PSN-4_Ct
PSN-5_Ct

PSN-1_BB
PSN-2_BB
PSN-3_BB
PSN-4_BB

IDAT...Identification Data
PSN... Pseudonyms
## EUPID – Linkage

**Reference Table**

<table>
<thead>
<tr>
<th>PSN&lt;sub&gt;CT&lt;/sub&gt;</th>
<th>PSN&lt;sub&gt;BB&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSN-1&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>PSN-1&lt;sub&gt;BB&lt;/sub&gt;</td>
</tr>
<tr>
<td>PSN-2&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>PSN-2&lt;sub&gt;BB&lt;/sub&gt;</td>
</tr>
<tr>
<td>PSN-3&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>PSN-3&lt;sub&gt;BB&lt;/sub&gt;</td>
</tr>
<tr>
<td>PSN-4&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>PSN-4&lt;sub&gt;BB&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

PSN... Pseudonyms
Needs for Secondary Use in the Pilot

• Creation of Pseudonyms
  – Pseudonymisation of patients in Clinical Trial (CT)
  – Pseudonymisation of patients in Bio Banking (BB)

• Linkage of Patients (if there is a link)

• Linkage of accompanying data
ABCD-4-E pipeline for data production

- EUPID Patient Registration
- Standardization of Data
  (Operational Data Model [14], XML-based)
- PSN-Exchange
- Data upload to Repository
EUPID - Patient Registration

- CLINICAL TRIAL (160 patients)
- BIOBANK (48 patients)

IDAT (SRN, DateOfBirth)

PSN ... Pseudonym
ODM ... Operational Data Model

IDAT ... Identification Data

31.05.2016
CSV2ODM – Standardization of Data

- CLINICAL TRIAL (160 patients)
- BIOBANK (48 patients)

IDAT ... Identification Data
PSN ... Pseudonym
ODM ... Operational Data Model
ODM2CoreDataset - PSN-Exchange and Data upload

- CLINICAL TRIAL (160 patients)
- BIOBANK (48 patients)

IDAT (SRN, DateOfBirth) Rawdata

ODM ODMPSN

BIOBANK (48 patients)
CLINICAL TRIAL (160 patients)

IDAT … Identification Data
PSN   … Pseudonym
ODM … Operational Data Model

a) b)
c

PSN ODM

ODM2CoreDataset

31.05.2016
Results

- EUPID was able to link each of the 48 patients from the biobank (BB) to their corresponding pseudonym from the clinical trial (CT).
- Joined dataset containing linked pseudonyms for 48 patients
- AIT took over the role of the trusted third party to access the Reference Table.
Discussion

- EUPID overcomes barriers of regulatory framework and patient consent

- Re-identification possible by Trusted Third Party

- Strategies to avoid Re-identification by comparing datasets
  - Randomizing order of datasets rows
  - Coarsening of typical individual information
  - k-anonymity
  - l-diversity
References


Thank you

Ebner Hubert
hubert.ebner.fl@ait.ac.at