THE FUTURE OF ELECTRONIC HEALTH RECORDS AND BIG DATA IN DEMENTIA RESEARCH
Disclosures

• Consultancy
  • Optum labs
  • Eisai
  • SomaLogic
  • Janssen/J&J

• Research funding/collaboration
  • AstraZeneca
  • Lilly
  • EFPIA companies through IMI

• Intellectual property
  • Patents related to biomarkers for AD held by KCL and Proteome Sciences
What is Big data?

Overwhelming

Transformative

Worrying
What is Big data?

- Volume
- Variety
- Velocity
- Complexity
• Enabling cohort and real world data for dementia research
  • European Medical Information Framework (IMI-EMIF)
  • Case Records Interactive Search (UK-CRIS)

• Using molecular and real world clinical data in drug development
  • From phase 0 to phase IV

• Looking ahead
  • Future platforms for better dementia research and care
To become the trusted European hub for health care data and insights into diseases and treatments.

**European Medical Information Framework (IMI-EMIF)**

- 14 European countries combining 58 partners
- €56 million worth of resources
- 3 projects in one
- 5 year project (2013 – 2017)

**ACADEMIC PARTNERS**

38

**SME PARTNERS**

9

**EFPIA PARTNERS**

10

**PATIENT ORGANISATION**

1
EMIF pipeline – discover, assess, use

Data Discovery
EMIF Catalogue
(Also within EPAD, DPUK as communities)

Assessment
EHR-derived data
Cohort-derived data

Access Request
Harmonisation: OMOP CDM
Harmonisation: Knowledge Objects

(Re)Use
PRREs†
tranSMART
Octopus
R Cloud
......

Ethical Code of Practice (ECoP)

† Private Remote Research Environment
Overview of Alzheimer’s disease studies in Europe

Detailed information from every study

Advanced search techniques

Compare cohorts
## Pre-competitive reutilisation of research and real-world data

### Research cohorts for dementia

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal cognition</td>
<td>31,376</td>
</tr>
<tr>
<td>Subjective cognitive complaints</td>
<td>4,369</td>
</tr>
<tr>
<td>Mild cognitive impairment</td>
<td>11,287</td>
</tr>
<tr>
<td>Probable/possible Alzheimer’s disease</td>
<td>9,754</td>
</tr>
<tr>
<td>Other dementia</td>
<td>2,453</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>~ 60,000</strong></td>
</tr>
</tbody>
</table>

### Real world datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarhus (DK)</td>
<td>2.3 million</td>
</tr>
<tr>
<td>Regional DB Tuscany (IT)</td>
<td>4.8 million</td>
</tr>
<tr>
<td>GePaRD (DE)</td>
<td>17 million</td>
</tr>
<tr>
<td>THIN (UK)</td>
<td>&gt; 11 million</td>
</tr>
<tr>
<td>IPCI (NL)</td>
<td>2.9 million</td>
</tr>
<tr>
<td>Health Service (IT)</td>
<td>&gt; 2 million</td>
</tr>
<tr>
<td>Pharmo (NL)</td>
<td>&gt; 7 million</td>
</tr>
<tr>
<td>UK cohorts (CVS, diabetes)</td>
<td>~ 500,000</td>
</tr>
<tr>
<td>EGCUT (EE)</td>
<td>52,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>~ 48 million</strong></td>
</tr>
</tbody>
</table>
EMIF-AD infrastructure: TranSMART platform

10 cohorts; > 4000 subjects; multiple phenotypes – demographics, imaging biomarkers, cognitive scores; molecular phenotypes and genotypes in process
Case Records Interactive Search (UK-CRIS)

14
NHS Mental Health Trusts across the UK

2.5m+
De-identified electronic patient records

1. Avon and Wiltshire Mental Health Partnership NHS Trust
2. Cambridgeshire and Peterborough NHS Foundation Trust
3. Camden and Islington NHS Foundation Trust
4. Devon Partnership NHS Trust
5. Kent and Medway NHS and Social Care Partnership Trust
6. Mersey Care NHS Foundation Trust
7. North East London Foundation Trust
8. Nottinghamshire Healthcare NHS Foundation Trust
9. Northumberland, Tyne and Wear NHS Foundation Trust
10. Oxford Health NHS Foundation Trust
11. Southern Health NHS Foundation Trust
12. South London and Maudsley NHS Foundation Trust
13. South West London and St George’s NHS Foundation Trust
14. West London Mental Health NHS Trust
UK-CRIS: safe, secure and complete

Trust network

- Input files
- Output files
- NHS Trust automated extraction
- EHR Backup and Data Dictionary
- Extracted NHS Trust database
- Data Dictionary

Output files sent via secure FTP to UK-CRIS. Data is encrypted

UK-CRIS datacentre and secure network

- UK-CRIS receives encrypted files.
- Data de-identified and uploaded
- NHS Trust secure VPN tunnel
- Elastic index (local)
- Elastic index (OMOP)
- PostgreSQL (local)
- PostgreSQL (OMOP)

Private network

- Access outside Trust (Authorised staff)
- NHS Trust secure VPN tunnel

NHS Trust Controlled Environment

- sFTP client
- Secure VPN tunnel
- sFTP server
- Web
- SQL
- NHS computer access

Web
- EHR Backup and Data Dictionary
- Extracted NHS Trust database
- Data Dictionary

SQL
- Elastic index (local)
- Elastic index (OMOP)
- PostgreSQL (local)
- PostgreSQL (OMOP)

NHS Trust secure
VPN tunnel
• EMIF publications
  • [http://www.emif.eu/results](http://www.emif.eu/results)
    • or search EMIF EU references
  • > 85 papers 2012-2017

• CRIS publications
    • or search CRIS BRC references
  • > 65 papers 2009-2017
• Target nomination and proof of concept
• Biomarkers for clinical trials
• Participant identification and recruitment
• Phase IV
• Regulatory / payer approval
• Generate pathways from complete GWAS datasets and perform clustering analysis for shared pathways

• Correlate pathway load per disease with risk relationship between disease and Alzheimer’s *from real world data*

• Perform proof of concept in human samples *using EMIF catalogue*
• Generate disease signature using differential transcriptomics with validation in vitro and in vivo using open access and repurposed datasets

• Identify compounds from 90k screen and in silico using Broad cMAP; confirm hits

• Proof of concept for compound hit class using real world data
Biomarkers for clinical trials

- Finding ‘impossible’ cohorts
  - **EMIF 500 and EMIF 1000**
  - discovery of blood based correlates of CSF and PET measures of amyloid load
  - Replicating markers with PPV 0.86

### Clinical study

<table>
<thead>
<tr>
<th>Clinical study</th>
<th>‘n’</th>
<th>Cost of Aβ PET screen failure</th>
<th>Cost of Aβ/tau PET screen failure</th>
<th>Screen failure saving with a blood biomarker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep and Frequent trial</td>
<td>250</td>
<td>£450k</td>
<td>£1.2m</td>
<td>£318k / £848k</td>
</tr>
</tbody>
</table>

Assumptions: screen failure rate ~40%; cost per amyloid PET £3k; cost per dual PET £8k; PPV of blood marker 0.86
Participant identification and recruitment

Readiness cohort → EPAD Registry → integration → Cohort A
Readiness cohort → EPAD Registry → integration → Cohort B
Readiness cohort → EPAD Registry → integration → Cohort C

Selection criteria

Trial cohort

Placebo → Adaptation by change in intermediate phenotype
Rx 1 → Adaptation on cognition outcomes
Rx 2 → Adaptation on cognition outcomes
Rx ... n → Adaptation on cognition outcomes

EPAD Registry

Trial cohort

Selection criteria

Adaptation by change in intermediate phenotype

Adaptation on cognition outcomes
Potential to deliver post-marketing data at scale and in real-world contexts

- South London & Maudsley NHS FT CRIS data
- n=2460; dementia treatment with AChEIs
- MMSE derived from coded and uncoded data
- Improvement by 4.2 units per year in first 6 months
- Predictors of response:
  - Better early response in non-white patients
  - Worse early response in vascular dementia

Beyond research data modeling - using CRIS for real-world, individual level costs

- Destination from uncoded data and linkage
- N=3075 (5624 6m windows)
  - 25% alone; 52% ADL problems
  - 37% physical illness; 45% moderate severity
- Mean costs of severe dementia >2x that for mild
- Increased care costs associated:
  
  *with* severity, functional problems, agitation, living alone
  
  *but not* physical illness, depression or gender

Big data – velocity missing
Deep & Frequent Phenotyping

- **Deep phenotyping**
  - Amyloid and tau PET
  - Blood, CSF, MRI, MEG, EEG, retinal, cognition, gait, connected devices for activity and function

- **Frequent phenotyping**
  - ~2-3 month repeat measures
  - Near continuous measures possible

- **Open science**
  - data sharing planned from outset
**Information governance**

**Trust network**
- Input files
- Output files
- NHS Trust automated extraction
- EHR Backup and Data Dictionary
- Extracted NHS Trust database
- Data Dictionary

**Private network**
- Access outside Trust (Authorised staff)
- NHS Trust secure VPN tunnel

**UK-CRIS datacentre and secure network**
- Output files sent via secure FTP to UK-CRIS.
- Data is encrypted
- UK-CRIS receives encrypted files.
- Data de-identified and uploaded
- Elastic index (local)
- Elastic index (OMOP)
- PostgreSQL (local)
- PostgreSQL (OMOP)

**NHS Trust Controlled Environment**
- Secure VPN tunnel
- Audit log
- Managed by Stakeholder-led oversight committee

**Data is encrypted**

**Managed by Stakeholder-led oversight committee**
CRIS consent for contact model

Assessment of capacity and consent for re-contact

Patient record re-identified in EMR

CRIS ID matched to EMR ID by Trusted Third Party

Researcher identifies eligible patient

Researcher contacts patient and informs clinical team

74% agreement

20,000 consents and samples in 3 years

Platforms for dementia research

- Target identification
- Proof of concept clinical trials
- Efficacy trials – feasibility to recruitment
- Regulatory and payer approval
- Phase IV

Open source molecular and clinical data
Cohort data with samples and re-contact
Real world data with re-contact

Real world data with re-contact
Cohort data with samples and re-contact
Open source molecular and clinical data

- AChEi
- CRIS Network
- EMA/Demed
- GAAIN
- EMIF
- EPAD
- Dementia Platform UK
- Global Alzheimer's Platform Foundation
- AAIC 17
acknowledgments

- CRIS
  - Mike Denis (Oxford)
  - Rob Stewart (KCL)
  - Matthew Broadbent (KCL/SLaM)
  - Tanya Smith (Oxford)
  - Felicity Callard (KCL)

- EMIF
  - Bart Vannieuwenhuyse (J&J)
  - Johannes Streffer (J&J)
  - Pieter Jelle Visser (VuMC)
  - Jose Luis Oliveira (Aveiro)
  - Michel Van Speybroeck (J&J)

- EPAD
  - Serge Van Der Geyten (J&J)
  - Craig Ritchie (Edinburgh)
  - Jose Luis Molinuevo (BBBRC)

- Oxford (targets & markers)
  - Alejo Nevado-Holgado
  - Chi-Hun Kim
  - Danielle Newby
  - Laura Winchester
  - Tim Johanssen
  - John Davis
  - Andrew Judge
  - Alison Baird
  - Sarah Westwood
  - Jennifer Lawson
  - Chris Hinds
  - Ivan Koychev
  - Lynn Rochester (Newcastle)
# DataParasites

# DataPhilanthropists

# DataSavesLives