PhD studentship at the Asthma UK Centre for Applied Research (AUKCAR):

Are anonymised databases truly anonymous?

An Introduction

Ref: AUKCAR-17-01a

This work is funded by CMVM/UoE. This work is carried out with the support of the Asthma UK Centre for Applied Research [AUK-AC-2012-01].

By Aryelly Rodriguez
22 NOV 2017
Background

Demands from funders, regulators and/or publishers to share clinical trials data to help with:

• Out of study scope questions
• Further exploratory analysis of outcomes
• Validation and transparency
• Analysis of Individual Participant Data (IPD) in meta-analysis
• Development of new methodologies
• Reducing costs and increase efficiency of health care
• Minimise the bias of positive reporting
Importance

• The importance and significant benefits of clinical data sharing have been well documented by many researchers in such as Gøtzsche(1), Packer(2), Al-Shahi et al.(3), Pisani et al. (4) and Bertagnolli et al. (5) just to cite a few:
Importance (cont.)

“interests of the patients must override commercial interests”(1)

“A moral imperative”(1)

“Respect for trial participants who often run a personal and unknown risk by participating in trials requires that they - and therefore also the society at large that they represent - be seen as the ultimate owners of trial data”(1)

“National and supranational legislation is needed to make data sharing happen as guidelines and other voluntary agreements do not work”(1)

“Now look at the medical literature in the 21st century. We no longer publish our data; instead, we present truncated summaries in the hope that readers will believe our conclusions without seeing the raw observations...It is no wonder that many clinicians have stopped reading the medical literature”(2)
Current situation

• “Concerns that patient confidentiality and consent may be breached are often cited by researchers as a reason for not sharing data” (4)

• “Threat of data parasites” (4)

• Presence of overly restrictive internal governance (timelines and quality of data)

• Fear of the public’s opinion
Current situation (cont.)


Ten oldest verified cases ever

Main article: List of the verified oldest people
Main article: List of the verified oldest women

Systematic verification of longevity has only been practiced in recent decades and only in certain parts of the world.

Note: All ten oldest people ever are women.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Sex</th>
<th>Birth date</th>
<th>Death date</th>
<th>Age</th>
<th>Place of death or residence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jeanne Calment</td>
<td>F</td>
<td>21 February 1875</td>
<td>4 August 1997</td>
<td>122 years, 164 days</td>
<td>France</td>
</tr>
<tr>
<td>2</td>
<td>Sarah Knauss</td>
<td>F</td>
<td>24 September 1880</td>
<td>30 December 1999</td>
<td>119 years, 97 days</td>
<td>United States</td>
</tr>
<tr>
<td>3</td>
<td>Lucy Harinar</td>
<td>F</td>
<td>16 July 1875</td>
<td>21 March 1993</td>
<td>117 years, 246 days</td>
<td>United States</td>
</tr>
<tr>
<td>4</td>
<td>Marie-Louise Meilleur</td>
<td>F</td>
<td>29 August 1880</td>
<td>16 April 1990</td>
<td>117 years, 130 days</td>
<td>Canada</td>
</tr>
<tr>
<td>5</td>
<td>Violet Brown</td>
<td>F</td>
<td>10 March 1900</td>
<td>15 September 2017</td>
<td>117 years, 189 days</td>
<td>Jamaica</td>
</tr>
<tr>
<td>6</td>
<td>Emma Morano</td>
<td>F</td>
<td>29 November 1899</td>
<td>15 April 2017</td>
<td>117 years, 137 days</td>
<td>Italy</td>
</tr>
<tr>
<td>7</td>
<td>Nabi Tajima</td>
<td>F</td>
<td>4 August 1899</td>
<td>Living</td>
<td>117 years, 52 days</td>
<td>Japan</td>
</tr>
<tr>
<td>8</td>
<td>Misao Okawa</td>
<td>F</td>
<td>5 March 1890</td>
<td>1 April 2015</td>
<td>117 years, 27 days</td>
<td>Japan</td>
</tr>
<tr>
<td>9</td>
<td>Maria Capurro</td>
<td>F</td>
<td>14 September 1889</td>
<td>27 August 2006</td>
<td>116 years, 347 days</td>
<td>Ecuador</td>
</tr>
<tr>
<td>10</td>
<td>Susannah Mushatt Jones</td>
<td>F</td>
<td>6 July 1899</td>
<td>12 May 2016</td>
<td>116 years, 311 days</td>
<td>United States</td>
</tr>
</tbody>
</table>

Men

Main article: List of the verified oldest men

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Birth date</th>
<th>Death date</th>
<th>Age</th>
<th>Place of death or residence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kiyomori Kimura</td>
<td>19 April 1897</td>
<td>12 June 2013</td>
<td>116 years, 54 days</td>
<td>Japan</td>
</tr>
<tr>
<td>2</td>
<td>Christian Mortensen</td>
<td>16 August 1882</td>
<td>25 April 1998</td>
<td>115 years, 256 days</td>
<td>United States</td>
</tr>
<tr>
<td>3</td>
<td>Emmeline Mercado del Toro</td>
<td>21 August 1891</td>
<td>24 January 2007</td>
<td>115 years, 165 days</td>
<td>Puerto Rico</td>
</tr>
<tr>
<td>4</td>
<td>Matthew Beard</td>
<td>9 July 1870</td>
<td>16 February 1895</td>
<td>114 years, 222 days</td>
<td>United States</td>
</tr>
<tr>
<td>5</td>
<td>Walter Boulton</td>
<td>21 September 1896</td>
<td>14 April 2011</td>
<td>114 years, 205 days</td>
<td>United States</td>
</tr>
<tr>
<td>6</td>
<td>Yuichiro Chigara</td>
<td>23 March 1889</td>
<td>28 September 2003</td>
<td>114 years, 169 days</td>
<td>Japan</td>
</tr>
<tr>
<td>7</td>
<td>Joan Rudawetz</td>
<td>15 December 1889</td>
<td>5 March 2004</td>
<td>114 years, 61 days</td>
<td>Spain</td>
</tr>
</tbody>
</table>
Current situation (cont.)

https://www.thesun.co.uk/news/2917810/elephant-man-drug-testing-trial-tgn1412

‘Nothing is really safe’: a focus group study on the processes of anonymizing and sharing of health data for research purposes

Gill Haddow PhD,1 Ann Bruce BSc MSc,2 Shiva Sathanandam MBBS MPH3 and Jeremy C. Wyatt DM(Oxon) FRCP(London)4

A UK-based clustered randomised controlled trial that set out to evaluate the safety and efficacy of an investigational drug for the treatment of idiopathic pulmonary fibrosis (IPF) became notorious as the ‘Elephant Man’ drug trial because of the horrific side-effects they endured.

Potential side-effects TGN1412 left the men writhing in agony and projectile vomiting before their immune systems crashed and they suffered multiple organ failure.

What happened next for the Elephant Man trial patients?

The worst affected was trainee plumber Ryan Wilson, then 21, who almost died after a devastating immune reaction left him with heart, liver and kidney failure.

When he woke from his coma two and a half weeks later, doctors told him: “You should be dead.”

He spent four months in hospital with pneumonia, sepsis and respiratory distress. He was told his first task to be learnt was to eat and drink

Close
Current situation (cont.)

https://www.patientslikeme.com/

(c) 2005-2017 PatientsLikeMe. All Rights Reserved. Information on PatientsLikeMe.com is reported by our members and is not medical advice.
“The Issue” for this project

Are patients at risk of being identified under the current methods / guidelines used for anonymisation at CTUs?
The PhD project aims and objectives are:

a. Investigate whether individual participants can be identified from a range of datasets that have been anonymised and made available for sharing

b. Identify factors that could increase the risk of re-identification of a dataset

c. Develop evidence-based recommendations on anonymization techniques and data security

d. Further explore researchers, patients and public perception of the sharing of clinical research data under the scenarios identified by this research
Literature Review

A systematic literature review is currently being executed with the goal to investigate:

- The current trends and practices among researchers/organisations regarding anonymization techniques and policies for data sharing
- The potential re-identification methods that can be used on anonymised IPD
- The strategies that are being used to protect anonymised and published data against re-identification

The literature review is including academic literature, reviews/reports in national media (UK) and social media such as blogs, Twitter and Facebook
Quantitative phase

Data collection: A selection of datasets that have been anonymised and made available for sharing will be obtained from the Edinburgh Clinical Trial Unit, Asthma UK, other Clinical Trial Units registered with UK-CRC, private sponsors such as GSK and from peer reviewed medical journals.

Analysis: All collected datasets will be crosschecked against any relevant re-identifying source using publicly available information (including study publications and web resources) to investigate whether any small groups or individuals can be identified. Also, the datasets will be assessed to determine how usable they are for further research. Finally, characteristics of the datasets (e.g. size of the study, public vs privately funded, rare vs common conditions) will be recorded to assess the potential risk of re-identification.
Qualitative phase

Canvas academic community, patients and public opinion about:

• The requirement for sharing IPD, including recognition of possible motives for an individual or organisation to identify individual participants
• Strategies to protect anonymity of patients if their study becomes newsworthy

This would be delivered by using questionnaires, focus groups and interviews
# Overall Schedule for key milestones

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Literature Review</th>
<th>Quantitative phase</th>
<th>Qualitative phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>0-10</td>
<td>Perform review</td>
<td>Issue requests and collect datasets</td>
<td>Identification of stakeholders and tool development for data collection</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Publish results</td>
<td></td>
<td>Send out questionnaires</td>
</tr>
<tr>
<td></td>
<td>14-22</td>
<td></td>
<td></td>
<td>Collect questionnaires, hold focus groups and interviews</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>26-44</td>
<td></td>
<td>Summarise progress</td>
<td>Code and Analyse responses</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td></td>
<td>Test datasets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>38-46</td>
<td></td>
<td>Generate and Publish results</td>
<td></td>
</tr>
<tr>
<td></td>
<td>48</td>
<td></td>
<td>Characterise datasets and determine risk factors</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>50-58</td>
<td></td>
<td>Generate results</td>
<td>Generate and Publish results</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td></td>
<td>Write PhD dissertation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62-72</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Many thanks to my funders

Asthma UK Centre for Applied Research

THE UNIVERSITY OF EDINBURGH

And supervisors:

Prof S Lewis (University of Edinburgh) steff.lewis@ed.ac.uk
Prof C Weir (University of Edinburgh) Christopher.Weir@ed.ac.uk
Prof S Eldridge (Queen Mary University of London) s.eldridge@qmul.ac.uk
References

(1) Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. Trials. 2011;12:249. pmid:22112900 Why we need easy access to all data from all clinical trials and how to accomplish it


(4) Pisani Elizabeth, Aaby Peter, Breugelmans J Gabrielle, Carr David, Groves Trish, Helsinki Michelle et al. Beyond open data: realising the health benefits of sharing data
BMJ 2016; 355 :i5295

DOI: 10.1056/NEJMs1702054

Thank you!

Any questions?