Disclosure control in complex medical outputs

ELIZABETH GREEN, FELIX RITCHIE, JIM SMITH, DAVID WESTERN, PAUL WHITE

UNIVERSITY OF THE WEST OF ENGLAND
Overview

➢ Medical research and positioning
  ➢ Current practices
  ➢ Alignment with TRE standards
➢ Examples
➢ Solutions?
➢ Future considerations
Medical research

Traditionally:

- Revolves around direct informed consent and primary data collection
- Some bad examples of mismanagement data and ethics... Henrietta Lacks for example
- Easy to see direct benefit/ public good
Sharing medical data and disclosure control

- Varied practices from depositing raw data in the public domain to secure access
- Tools are being developed - uptake poor
- Unforeseen consequences encountered
- Benefit to society
Genomic Data

- Tissue sample extracted, analysis is then conducted
- Microarray data of specimen often deposited/shared in the public domain (remember knowledge advancing continually)
- Descriptive variables provided

<table>
<thead>
<tr>
<th>Participant number</th>
<th>Gender</th>
<th>Age</th>
<th>Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>48</td>
<td>Bad- chronic insomnia</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>31</td>
<td>Good</td>
</tr>
</tbody>
</table>

Data collection: The DRAGoN Hospital for Exhausted Researchers
Statistical Disclosure Control?

- Basic SDC principles- counts and thresholds
- Kaplan-Meier curves often result in low numbers, potential for low number of observations between each step.
- Not directly disclosive as alone, requires linking to contextual information (provided in the report).
- Online tools- generating a Kaplan-Meier curve with SDC inbuilt
Dermatology photographs

• Online photo repository providing examples of different skin disorders.
• Used as an aid to help experts and public identify different ailments.
• Huge public benefit
• However unintended consequence was found...
Statistical Disclosure Control? Solutions?
FMRI scan

- Huge amount of data
- Brain anatomy and structure (remember knowledge advancing continually)
- Variation in sharing online repositories open access.
- Concern for disclosure is based on rebuilding face based on structure
FMRI solutions?

Scalp the face! Remove/roughly the facial attributes
- Potential loss of data
- Disclosure elsewhere
Statistical Disclosure Control?

• SDC training and principles for medical data users!
• Unique attributes, highly detailed information, informed consent, understanding risk
• Use and implementation of tools and solutions
• Universal agreement on standards
  • What is anonymized data
  • Evaluation of data access
• Development of a network to support this work?
Thank you!

Elizabeth Green
Elizabeth7.green@uwe.ac.uk

This work is funded by UK research and Innovation [Grant Number MC_PC_23006] as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with Health Data Research UK (HDR UK) and Administrative Data Research UK (ADR UK).