Novartis journey in implementing CDP including performing a motivated intruder attack

Janice Branson, VP Biostatistics Head, Novartis Pharma AG
Amsterdam
June, 2018
Disclaimer

• This presentation is based on publicly available information (including data relating to non-Novartis products or approaches)

• The views presented are the views of the presenter, not necessarily those of Novartis

• These slides are intended for experience sharing purposes only and for the personal use of the audience. These slides are not intended for wider distribution outside the intended purpose without presenter approval

• The content of this slide deck is accurate to the best of the presenter’s knowledge at the time of production
Agenda

• Original approach to implementing CDP Policy

• Performing a motivated intruder attack
  – Plan and methodology
  – Assumptions
  – Example
  – Results

• Conclusions
Original experience with implementation of Policy 0070

Overarching objective: Balance the risk of reidentification against data utility

- As we evolve our quantitative risk assessment methodology, we want to gather the empirical evidence specific to clinical trial data to support the methodology assumptions.

- For first dossiers in scope of Policy 0070, Novartis balanced the lack of direct experience concerning risk assessment with data utility until further understanding is gained. This balance was key to protecting patient privacy.

- Novartis used first in scope dossiers to test assumptions, understand challenges, assess level of QC needed and gain experience with the methodology and inform Novartis’ future approach and planned a motivated intruder attack.

- Specifically Novartis employed a hybrid approach until more experience is gained:
  - Redaction/removal of Narratives in section 14.3.3 and per patient per visit line listing
  - Anonymization as basis

- Followed this approach for the first 5 dossiers.
Interim Novartis process for EMA Policy 0070

Create Awareness

Identify documents in scope of Policy 70

Assess for any potential CCI */ sensitive information (based on the available public information from publications, conferences etc.)

Novartis Internal Board meeting – reviews team’s proposal for endorsement

Share document with vendor for PPD** anonymization / redaction*** + report generation

Mark CCI (if any)

Submit to EMA redaction proposal documents

Rework package and submit

*CCI – commercially confidential information

**PPD- protected personal data

***At this time Novartis redacts patient narratives and per patient per visit listings

Prepared by Novartis

Prepared by Vendor

---

5 PSI 5th June 2018 - Confidential - Novartis copyright 2018
Deidentification Cycle in Policy 0070

1. Set Risk Threshold
Based on the characteristics of the data and precedents, a quantitative risk threshold is set. Policy 0070 recommended 0.09*

2. Measure Risk
Appropriate metrics are selected and used to measure reidentification risk from the data.

3. Evaluate Risk
Compare the measured risk against the threshold to determine if it is above or below it.

4. Apply Transformations
If the measured risk does not meet the threshold, specific transformations are applied to reduce the risk.

*Novartis may decide in certain cases to use a lower threshold
Objective of motivated Intruder Attack

• To empirically test whether the probability of re-identification that was computed during the anonymization process was indeed as low as assumed:
  ➢ Quantitative re-identification risk calculations that are typically used are based on statistical models, and these models make assumptions. The assumptions that we make tend to be conservative, which means that the true re-identification risk might be underestimated

• The motivated intruder test provided the following outcomes to the overall risk mitigation approach:
  ➢ Provides an additional data point to improve the anonymization practices for a particular data release or type of data release
  ➢ Helps adjust the assumptions that have been made in the re-identification risk measurement (for example, about what are the plausible direct and indirect identifiers that can be used in an attack)
  ➢ Ensures that Novartis has an updated understanding of the real risks in some data recipient environments
Motivated Intruder Attack Plan

• Good Research, an independent third party security and privacy consulting firm based in California, were provided with the re-identification protocol, anonymized CSR

• The CSR was anonymized using the risk-based methodology applied by Privacy Analytics for EMA Policy 0070 document releases

• The motivated intruder test was conducted against 500 US patients in the study

• The re-identification attack followed the motivated intruder test approach described by the UK Information Commissioner’s Office

• A comprehensive protocol was developed that detailed the steps and the measurements that must be taken during such an attack, as well as the principles that must be followed
Types of attack

• The motivated intruder test attempted these two types of attacks on the data:

  ➢ Population to sample attack - The adversary knows someone in the population and attempts to find that person's records in the data. In the context of a motivated intruder test, this could be a famous person or an acquaintance of the intruder.

  ➢ Sample to population attack - The adversary starts from the data and tries to match records with real people in the population. The population can be represented by a registry (say a voter registration list). The registry may be pre-existing or may be created by conducting searches on the web. No commercial databases will be considered for population information.
Methodology

• The re-identification attack followed the motivated intruder test approach described by the UK Information Commissioner’s Office

• A confidence score based on a five-point scale was used to assess the potentially re-identifiable candidates where scores in the 4 or 5 range would be needed to have any kind of reasonable assurance that the matches are correct

• No actual attempt is made to identify patients – all information is collected and compared to the data and the confidence score is calculated

• At the end of a motivated intruder test 2 numbers were generated:
  ➢ The percentage of individuals in the dataset that have been correctly re-identified
  ➢ The effort to re-identify an individual
Assumptions

• Characteristics of the motivated intruder tester:
  ➢ Reasonably competent to perform a re-identification attack
  ➢ Has access to resources, such as the Internet, libraries, and all public documents. No commercial databases were purchased
  ➢ Employs investigative techniques such as making enquiries of people who may have additional knowledge of the identity of a data subject or advertising to anyone with information to come forward
  ➢ Does not have access to specialist equipment
  ➢ Does not resort to criminality, such as burglary, to gain access to data that is kept securely
Example from original Policy 0070 submission (Redaction)

- Original

Table 14-127  Listing of Treatment Emergent Serious Adverse Events (TEAE) (Safety Analysis Set)

<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Age(Y)/Sex</th>
<th>Treatment</th>
<th>Preferred Term</th>
<th>Part</th>
<th>Onset Day</th>
<th>AE Duration</th>
<th>Severity</th>
<th>Outcome</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>257</td>
<td>50/Male</td>
<td>Vehicle</td>
<td>Myocardial infarction</td>
<td>25</td>
<td>21 Days</td>
<td>Severe</td>
<td>resw</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>257</td>
<td>50/Male</td>
<td>Vehicle</td>
<td>Sepsis</td>
<td>25</td>
<td></td>
<td>Severe</td>
<td>ongw</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>258</td>
<td>50/Female</td>
<td>Vehicle</td>
<td>Cardiac failure congestive</td>
<td>6</td>
<td></td>
<td>Moderate</td>
<td>ongw</td>
<td>NR</td>
<td></td>
</tr>
</tbody>
</table>

- Redacted

PPD
Example from original Policy 0070 submission (Anonymization)

- Original

**Table 14-127**

<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Age(Y)/Sex</th>
<th>Treatment</th>
<th>Preferred Term</th>
<th>Part</th>
<th>Onset Day</th>
<th>AE Duration</th>
<th>Severity</th>
<th>Outcome</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>257</td>
<td>50/Male</td>
<td>Vehicle</td>
<td>Myocardial infarction</td>
<td>25</td>
<td>21 Days</td>
<td></td>
<td>Severe</td>
<td>resw</td>
<td>NR</td>
</tr>
<tr>
<td>257</td>
<td>50/Male</td>
<td>Vehicle</td>
<td>Sepsis</td>
<td>25</td>
<td></td>
<td></td>
<td>Severe</td>
<td>ongw</td>
<td>NR</td>
</tr>
<tr>
<td>258</td>
<td>50/Female</td>
<td>Vehicle</td>
<td>Cardiac failure congestive</td>
<td>6</td>
<td></td>
<td></td>
<td>Moderate</td>
<td>ongw</td>
<td>NR</td>
</tr>
</tbody>
</table>

- Anonymized

**Table 14-127**

<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Age(Y)/Sex</th>
<th>Treatment</th>
<th>Preferred Term</th>
<th>Part</th>
<th>Onset Day</th>
<th>AE Duration</th>
<th>Severity</th>
<th>Outcome</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>263</td>
<td>41/Male</td>
<td>Vehicle</td>
<td>Myocardial infarction</td>
<td>25</td>
<td>21 Days</td>
<td></td>
<td>Severe</td>
<td>resw</td>
<td>NR</td>
</tr>
<tr>
<td>263</td>
<td>41/Male</td>
<td>Vehicle</td>
<td>Sepsis</td>
<td>25</td>
<td></td>
<td></td>
<td>Severe</td>
<td>ongw</td>
<td>NR</td>
</tr>
<tr>
<td>294</td>
<td>43/Female</td>
<td>Vehicle</td>
<td>Cardiac failure congestive</td>
<td>6</td>
<td></td>
<td></td>
<td>Moderate</td>
<td>ongw</td>
<td>NR</td>
</tr>
</tbody>
</table>
Results of Motivated Intruder Attack

• The investigation highlighted 6 subjects from the targeted 500, that could potentially be matched with external sources

• Each match has a confidence score associated with it and the 6 confidence scores obtained in this test were 1 and 2 (Notes scores of 4 and 5 are required to have reasonable confidence of a likely match)

• The external information that was used to match for these 6 cases came from public death records and social media
  ➢ Information from regulators (the FDA and EMA) that was possible to obtain in the timeframe of the study was not useful for this re-identification attack, and also information from other public trials registries were not useful for the attack
Results of Motivated Intruder Attack

Estimated effort per subject ~24 hours

Entire effort (170 hours) excluding 25 hours of project management tasks (e.g. writing the report, project meetings) for a total of 145 hours. 145 hours / 6 candidates = ~24.2h per candidate.
# Suspected Identified Subjects

<table>
<thead>
<tr>
<th>Approach</th>
<th>External Source</th>
<th>Conf. Score</th>
<th>Conf. Group</th>
<th>Suspected Case# (Target Report)</th>
<th>Reason for Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-Social Media</td>
<td>Facebook</td>
<td>2</td>
<td>Low</td>
<td></td>
<td>Date of surgery + Location + Symptoms + Diabetic (inferred)</td>
</tr>
<tr>
<td>5-Social Media</td>
<td>Reddit</td>
<td>1</td>
<td>Low</td>
<td></td>
<td>Date of surgery + Age + Gender + Diabetic</td>
</tr>
<tr>
<td>2-Death Records</td>
<td>Ancestry.com</td>
<td>1</td>
<td>Low</td>
<td>Case 2108</td>
<td>Age + Date of death + Ethnicity (inferred) + Unknown diabetic status</td>
</tr>
<tr>
<td>2-Death Records</td>
<td>Ancestry.com</td>
<td>1</td>
<td>Low</td>
<td>Case 2108</td>
<td>Age + Date of death + Ethnicity (inferred) + Unknown diabetic status</td>
</tr>
<tr>
<td>2-Death Records</td>
<td>Ancestry.com</td>
<td>1</td>
<td>Low</td>
<td>Case 2108</td>
<td>Age + Date of death + Ethnicity (inferred) + Unknown diabetic status</td>
</tr>
<tr>
<td>2-Death Records</td>
<td>Ancestry.com</td>
<td>2</td>
<td>Low</td>
<td>Case 8710</td>
<td>Age + Diabetic (inferred) + Details of Death + Location</td>
</tr>
</tbody>
</table>

Table 3 - List of suspected identified subjects for each approach

Subject 1 exact date of surgery and similar symptoms
Subject 6 death details matched
Conclusions

• No high confidence of potential matches from this investigation

• Even if the confidence scores were ignored, the match rate was ~0.01, which is much lower than the EMA threshold of 0.09

• This re-identification attack provides empirical evidence that a motivated intruder attack on this anonymized document did not yield any potential subject re-identifications

• To the extent that these results can be generalized across other CSRs, this demonstrates the adequacy of the proposed EMA anonymization methodology against contemporary attacks

• Novartis plans to update our approach and anonymize the documents including narratives and is committed to continue the assessment of anonymization methodology specifically we do not stop our efforts after this intruder attack