



CASE STUDY – CARFILZOMIB MAA

CHRISTINE FLETCHER

EXECUTIVE DIRECTOR BIOSTATISTICS, AMGEN LTD



DISCLAIMER

I am an employee of Amgen Inc. The views expressed herein represent those of the presenter and do not necessarily represent the views or practices of the presenter's employer or any other party.

CHARACTERISTICS OF THE CARFILZOMIB MAA

- **The development program**
 - **Multiple myeloma (Orphan Designation)**
 - **19 clinical studies**
 - **8 in US only**
 - **5 in US + Canada**
 - **6 multiregional**
 - **N = 11 to 929 subjects**
- **The dossier**
 - **> 75,000 pp clinical documents in scope**
 - **Most were written before Policy 070 came into effect**

SOME FACTORS WE CONSIDERED

- **Consent**
- **Potential harm to a subject who is re-identified**
- **Orphan disease population**
- **Small studies**
- **Possibility of a deliberate attempt to re-identify subjects**
- **Impact of big data and social media**
- **Time to implementation**
- **Alternative mechanisms to share data**

CHOICES WE MADE

- Qualitative approach
- Redaction
- Re-identification scenarios based on prosecutor risk (an attacker is aware that the target is represented in the data)
- “Maximum risk” concept – consider the data subjects who are at highest risk of re-identification
- Defined rules with risk stratification by
 - study characteristics (number of subjects, geographic area)
 - data presentation (granularity of data, how many data points presented for 1 subject?)

VARIABLES

- Direct identifiers (redact all)
 - subject identification numbers
 - safety case numbers
 - names of individuals
 - signatures
 - addresses of individuals
 - email addresses of individuals
 - phone numbers of individuals
- Quasi identifiers (redact all)
 - calendar dates
 - geographic locations
 - ages above 89 years
 - individual genotype
- Quasi identifiers (see risk matrix)
 - age
 - race and/or ethnicity
 - sex
 - height, weight, body mass index
 - medical history and prior treatments
 - categorised genetic data

HIGHER RISK = FULL REDACTION (REMOVE)

MODERATE RISK = REDACT QUASI-IDENTIFIERS (OR PARTS OF TABLES WITH LOW COUNTS)

LOW RISK = NO REDACTION



Data presentation ^a

Study Characteristics

< 100 subjects or single center

100 to <1000 subjects or single country

Direct identifiers

Full narratives

“Sensitive” individual data

Brief narratives

Listings, brief text

Subgroup data for small groups

Text with 1 quasi-identifier

Demographic data for small groups

Summary data without quasi-identifiers

Individual data without quasi-identifiers

^a Operational definitions were created for each presentation type

CHALLENGES OF SOCIAL MEDIA

“[Username]. I was diagnosed [day, month, year]... While I am ISS-X and DS-X my cytogenetic profile classifies me as [risk class] MM. I have [list of 5 specific genetic markers]. Despite this genomic profile I had no symptoms & the bone marrow biopsy (X% plasma cells) report said [verbatim text]. Only my [imaging procedure] was indicative of myeloma... On [day, month, year], I began care at [study center] in a carfilzomib clinical trial. X cycles of Carfilzomib [dose] with lenalidamide [dose] and lo-dose dex, followed by 1 yr of maintenance with Len [dose]. In [month, year], [test] after X cycles, indicated [outcome].... My spouse and I are [specific university] alum and we have [number] [sex of children]. Education: [scientific field]”

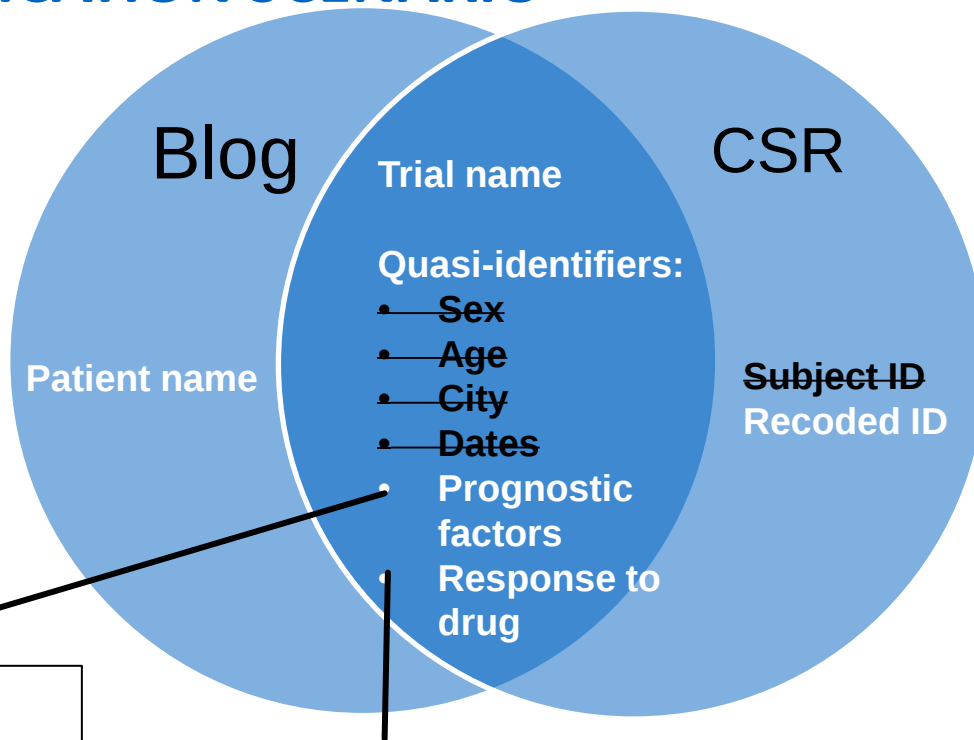
Some premises:

- Patients with a serious illness may be motivated to share information about their clinical trial experience
- Self-identifying as a trial participant increases the risk of re-identification
- It is difficult to model what information a patient is “likely” to share
- Voluntary sharing of some information does not imply--
 - consent to disclose additional information
 - absence of harm if additional information were disclosed

THOUGHTS ON NARRATIVES

- multiple quasi-identifiers for the same subject, which effectively reduces cell size to 1
- difficult to support assumptions about what variables an attacker could know
 - serious adverse events but not non-serious events
- verbatim (non-coded) text which can be highly unpredictable, hard to distinguish, hard to model
 - the “1-armed lorry driver”
- possibility for inference
 - prior medications -- medical history– baseline laboratory values
- identifying information is also important for case interpretation
 - marginal risk > marginal utility

SOCIAL MEDIA RE-IDENTIFICATION SCENARIO



Listing of baseline disease characteristics

Subject 12345

- ISS-**X** and DS-**X**
- cytogenetic profile [**risk class**] MM.
- [**list of 5 specific genetic markers**]
- no symptoms
- bone marrow biopsy (**X**% plasma cells)
- [**imaging procedure**] indicative of myeloma"

Listing of efficacy response data

Subject 12345

"[**test**] after **X** cycles, indicated [**outcome**]"

POTENTIAL IMPACT

Although the patient has self-reported some information, re-identification might reveal new information that they did not plan to share

This could range from a trivial to a substantial amount – for example, if all of the patient's records are linked by the same ID number

Listing of baseline disease characteristics

Subject 12345

- ISS-**X** and DS-**X**
- cytogenetic profile [**risk class**] MM.
- [**list of 5 specific genetic markers**]
- no symptoms
- bone marrow biopsy (**X%** plasma cells)
- [**imaging procedure**] indicative of myeloma"
- **Prognostic factor X**
- **Prognostic factor Y**

Listing of efficacy response data

Subject 12345

"[**test**] after **X** cycles, indicated [**outcome**]"

"[**test**] after **X+1** cycles, indicated response

"[**test**] after **X+2** cycles, indicated response

"[**test**] after **X+3** cycles, indicated progression

Additional records

Subject 12345

- **Listing of adverse events**
- **Listing of medical history**
- **Listing of laboratory results**
- **Safety narrative**

RECOMMENDATIONS

- **Think about how changing social media norms may disrupt standard assumptions about**
 - **what external data sources are readily available**
 - **prevalent population size**
 - **what variables, and how many variables, an intruder may know**
 - **the most effective ways to mitigate risk**

PHILOSOPHICAL

- Clinical reports are complex & multidimensional. It is not trivial to fit these into anonymization frameworks that were developed based on structured data sets
- The context for *clinical trials* and *clinical trial participants* is different than for routine medical practice, in ways that substantially impact risk
- In extending existing anonymization frameworks to clinical reports, we should
 - pressure-test assumptions built into these frameworks
 - actively seek disconfirming information
 - gather empirical evidence about their fitness in real world use