EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

# Analysis of clinical reports published in the context of Policy 0070

Technical Anonymisation Group (TAG) meeting, London, 29-30 November 2017
Agenda point 06

Presented by Ada Adriano
Documents Access and Publication Service

An agency of the European Union

# Acknowledgments

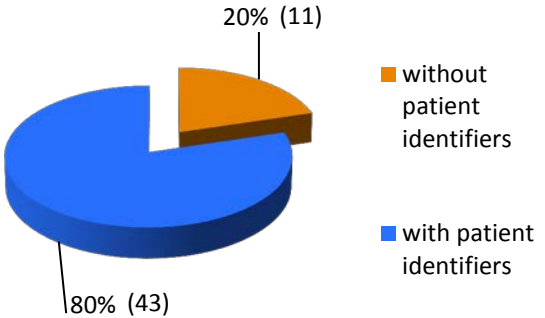
Kanako Sasaki (Visiting Expert from Japanese MHLW) and EMA Clinical Data Publication team

# Analysis of published anonymisation reports

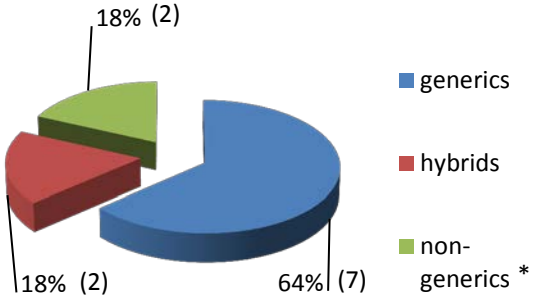
EUROPEAN MEDICINES AGENCY

- 54 anonymisation reports published (cut-off date: 06 October 2017)

Product category

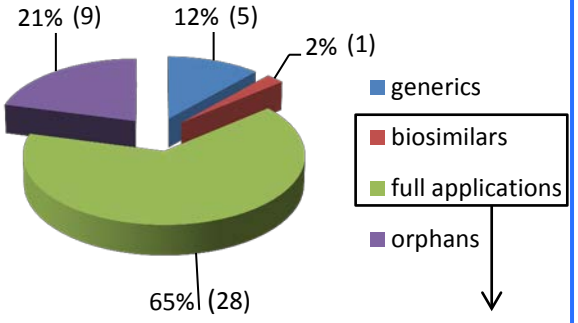



Product category
(without patient identifiers)




*(literature data)

Product category
(with patient identifiers)



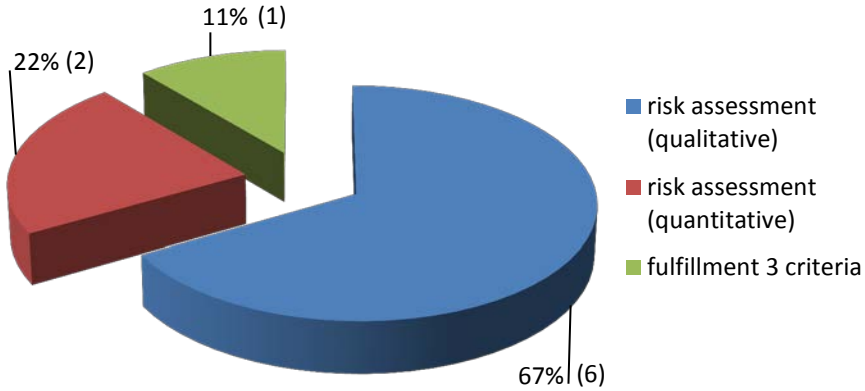
# Orphans

- Mainly small/very small study size (e.g. n=2, n=3 subjects);
- Size of study population mostly accounted for in the anonymisation process (8/9);
- Attacks envisaged linked to the type of product (e.g. gene therapy).

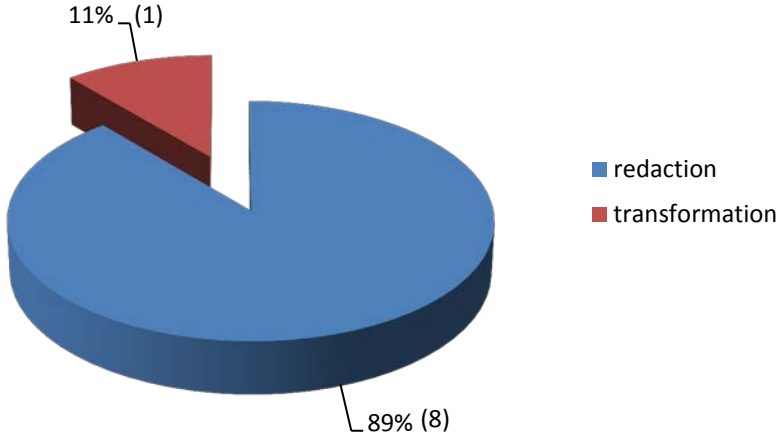
EUROPEAN MEDICINES AGENCY

# Methodology applied

## Anonymisation assessment




## Anonymisation technique

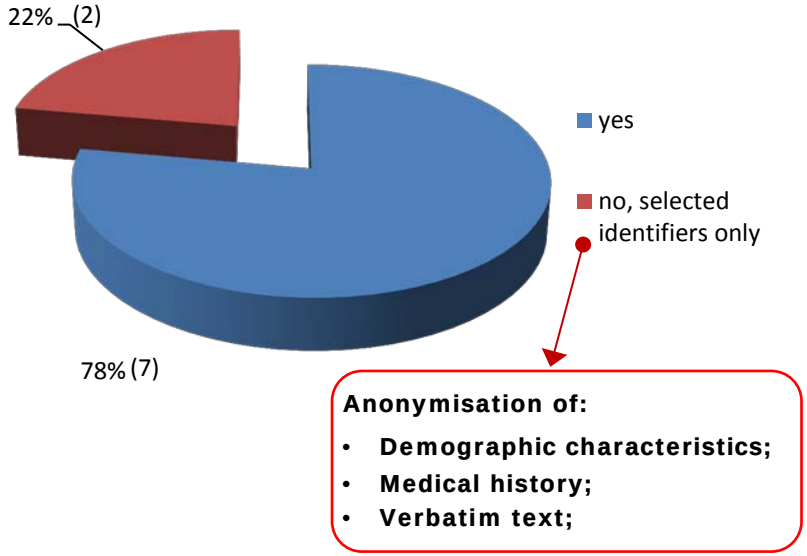



N(orphans) = 9

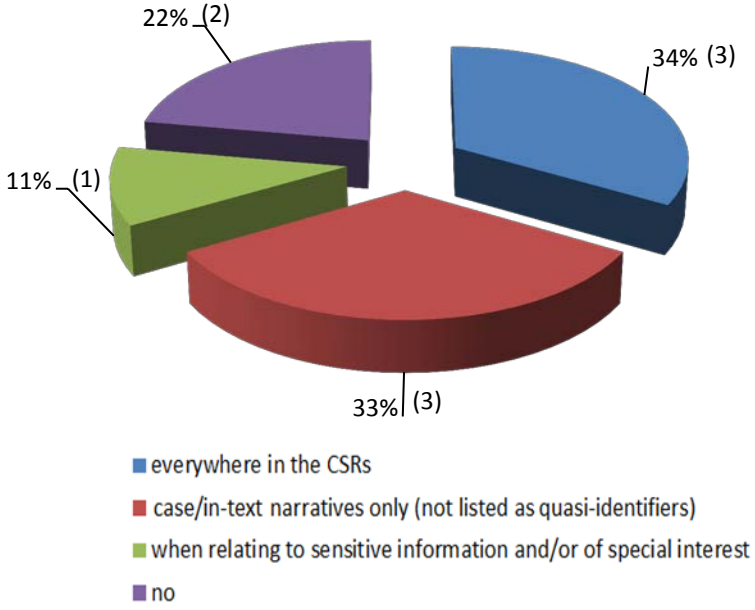
# Anonymisation applied

- Redaction of medical history and demographic characteristics throughout CSRs (8/9);

Full redaction of case narratives




Redaction of adverse events



EUROPEAN MEDICINES AGENCY

## Examples of quantitative approaches

### Alprolix:

- Redaction of quasi-identifiers to remove unique combinations of quasi-identifiers;
- Full redaction of narratives performed;
- For subgroups ≤11*, median, minimum and maximum values redacted.

### Darzalex:

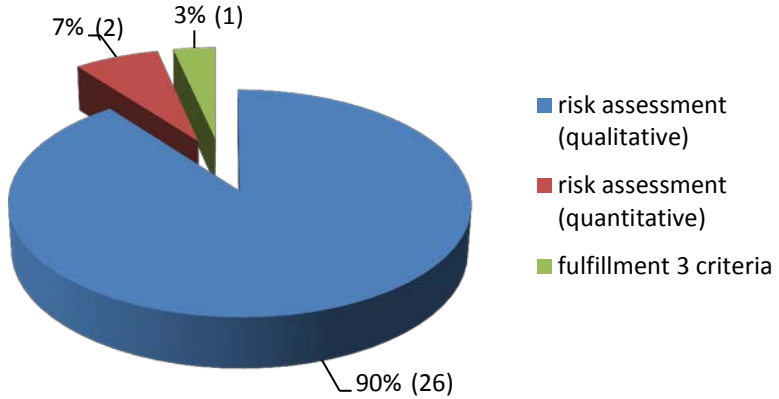
- Same approach used for non-orphan/non-generic product (i.e. Afinitor);
- Case narratives NOT fully redacted!

*N patients=1/0.09

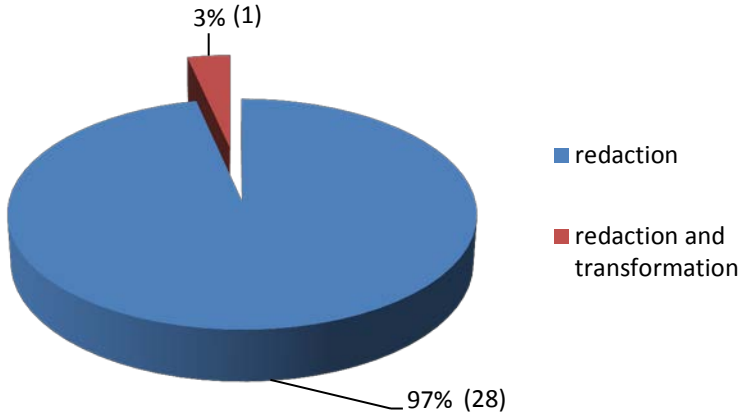
# Non-orphans / non-generics

- Usually large studies (i.e. >100 subjects);
- Few studies with <100 subjects (e.g. Phase I studies);

Anonymisation assessment




Anonymisation technique




N(non-orphans/non-generics)=29

# **Qualitative approach** (non-orphans/non-generics)

EUROPEAN MEDICINES AGENCY

- Qualitative risk threshold to be set (e.g. low, very low);
- No calculation of re-identification risk;
- Risk assessment based on subjective evaluation;
- Analytical approach?
- Redaction as preferred technique;
- Study categorisation driven by sample size (12/26): what is small/big?
- Heterogeneity in the anonymisation performed.

# Qualitative approach (non-orphans/non-generics)

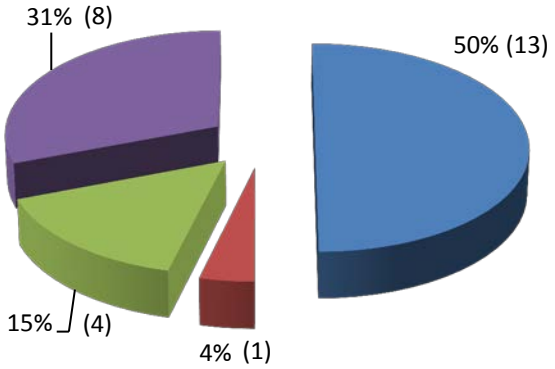
EUROPEAN MEDICINES AGENCY

## Anonymisation applied

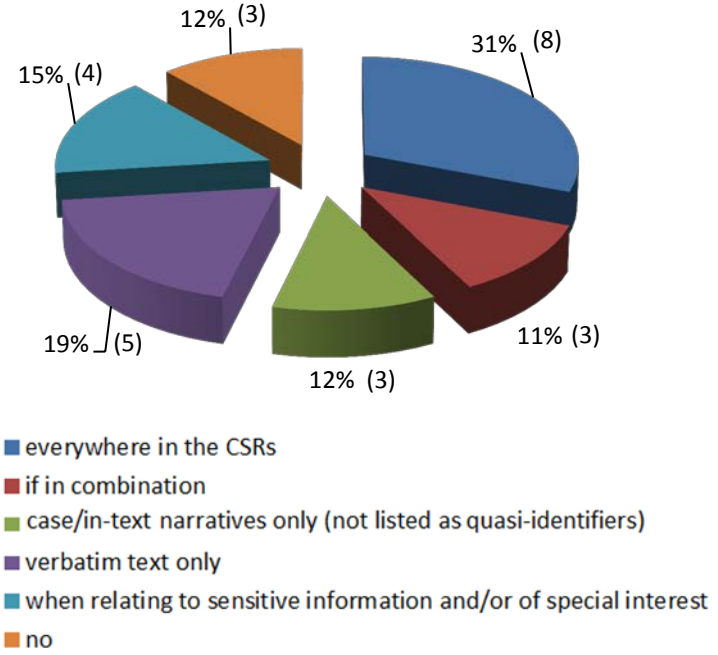
Redaction

- Demographic characteristics (26/26)
- Medical history (20/26)
- In-text narratives (14/26)

Full Redaction of case narratives

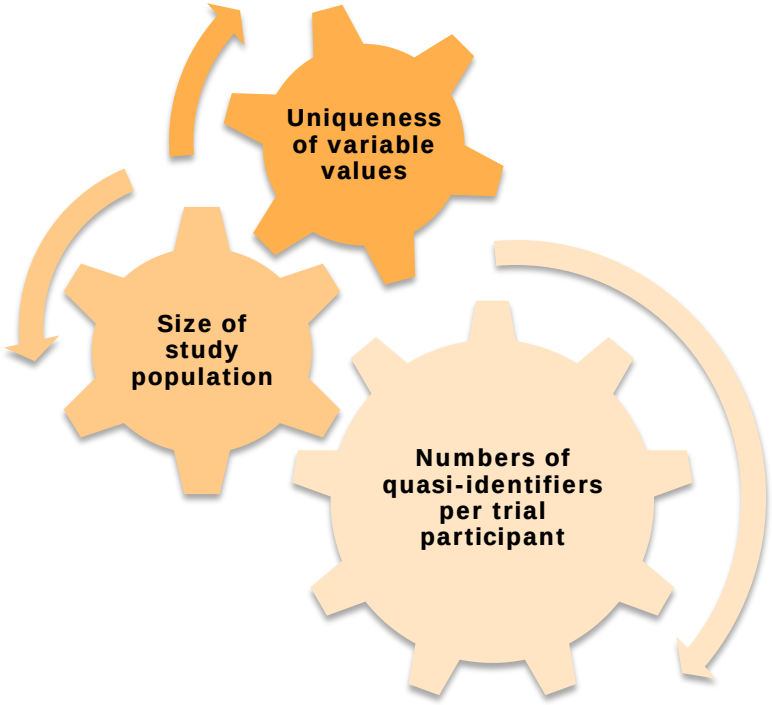



Redaction of adverse events



# **Qualitative approach** (non-orphans/non-generics)

EUROPEAN MEDICINES AGENCY




**Uniqueness of variable values (11/26):**

- Criterion for identifiers selection;
- Redaction of specific variable values;
- Non-uniqueness considered.

**Numbers of quasi identifiers per trial participant (18/26):**

- Combination of variables considered.

**Size of study population (18/26):**

- Study categorisation based on study characteristics;
- Lack of harmonisation in the identifiers/sections redacted.

# **Quantitative approach** (non-orphans/non-generics)



- Quantitative risk threshold to be set (0.09);
- Calculation of re-identification risk;
- Transformation as additional technique (e.g. pseudo-anonymisation, offset dates, randomisation, generalisation of medical history to MedDRA HLT, HLGT and SOC);
- Less conservative assumptions (data set considered, attacker knowledge);
- Different methodologies applied.

# **Quantitative approach** (non-orphans/non-generics)

EUROPEAN MEDICINES AGENCY

## Zinbryta:

- **Full combined population of all studies** used in the analysis;
- Subjects grouped into equivalence classes (minimum equivalent class size= 12);
- Verbatim terms and sensitive data not included in the risk assessment;
- **Redaction** as anonymisation technique.

- **No full** redaction of **case narratives** (subject ID, dates, age);
- Adverse events redacted when in **combination** and/or **unique**;
- Redaction selected frequencies in table summarizing adverse events by body weight.

## Afinitor:

- **Population in similar trials** used in the analysis;
- Quasi-identifiers that are caught and those missed accounted for in the risk calculation;
- Local recoding: different transformation based on the level of risk;
- **Transformation** as anonymisation technique (dates, age, medical history).

- **Suppression** applied to some identifiers (e.g. race);
- Subject IDs pseudo-anonymised;
- **Full** redaction of **case narratives** prior to risk assessment;
- Serious adverse events redacted in narratives.

# Data utility

- Not integrated in the risk assessment;
- Linked to aggregated data only;
- Expectations of end users not clearly addressed;
- Impact of full redaction of narratives not always addressed.

# Conclusions

- Disease and/or study population driving the anonymisation process;
- Limited experience (public release, potential adversaries, unstructured text);
- Limited confidence with the assumptions (threshold, data set, type of attacks).

# Any questions?

**European Medicines Agency**
30 Churchill Place • Canary Wharf • London E14 5EU • United Kingdom
**Telephone** +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
**Send a question via our website** www.ema.europa.eu/contact

Follow us on **@EMA_News**