Principles to maintain patient confidentiality

Katherine Tucker, Roche

13 November 2014

Agenda

1. Why do we need to protect patient confidentiality?
2. Some general thoughts and assumptions
3. Example
4. Outline of draft principles from EFSPI/PSI DS WG
5. Conclusions

Why do we need to protect patient confidentiality?

- Recent drives to share CT data more openly with researchers, for benefit of public health and further scientific research
- Highlighted importance of appropriate protection of patient identities and personal information of patients who participate in our trials
- Specifically we need to consider:
  - **what** to share: including level of de-identification / anonymization steps to apply
  - **how** to share: what type of controlled access & legal agreements
- Need to consider data privacy laws around the world, regulations, guidances and other policies

Some general thoughts and assumptions

- The level of de-identification / anonymization (DA) you need to apply is related to how you will share
- Balance between data utility vs. risk of patient re-identification

Some general thoughts and assumptions

Lets assume you want to minimise risk of patient re-identification, where to set the dial on level of control/security and subsequent DA steps?

Place data openly on internet, no protection or control on access

Place data in password protected secure system with controlled access to named individuals, legal agreement signed by researchers

Highest level of DA required to minimise risk of patient re-identification *

* "Data can either be useful or perfectly anonymous, but never both"
Law Professor Paul Ohm 2009

More pragmatic approach considered, relatively high level of DA still required, but retaining reasonable data utility

Some general thoughts and assumptions

Lets assume you are sharing in a reasonably controlled and secure manner, where to set the dial on level of DA steps vs. data utility?

Apply high level of DA steps

Lower data utility *

Lower risk of patient re-identification

Apply low(er) level of DA steps

Higher data utility

Higher risk of patient re-identification

* Implications on reproducibility of original analyses

Example

QUALIANCE

Trusted Partner since 2010

5 Years

Data Transparency – Data De-Identification and The Role of the Programmer

PhUSE 2014, London

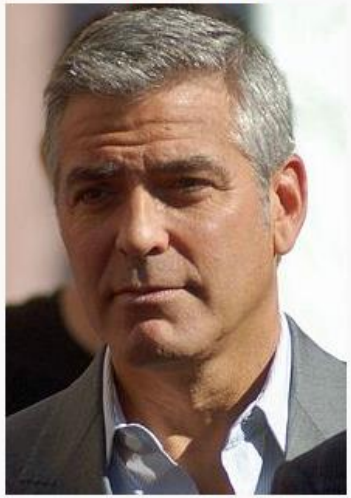
Jean-Marc Ferran

Consultant & Owner

Who is Georges?

Patient ID	DoB	Age	G		Country	Partner Age
1	12APR1963	51	M		nada	48
2	28MAY1974	40	M		ance	41
3	06MAY1961	53	M		ited States	36
4	28MAY1954	60	Fe		ain	65
5	14JUL1969	45	M		azil	41
6	13AUG1964	50	Fe		gentina	45
7	18MAR1961	53	M		ited States	48
8	22JAN1961	53	M		ited States	37
9	27SEP1924	90	M		nada	73
10	07FEB1956	58	Male	White	Canada	62

George Clooney



Clooney at a ceremony for John Wells to receive a star on the Hollywood Walk of Fame in January 2012

Born	George Timothy Clooney May 6, 1961 (age 53) Lexington, Kentucky, U.S.
Occupation	Actor, filmmaker
Years active	1978–present
Spouse(s)	Talia Balsam (m. 1989–93) Amal Alamuddin (m. 2014)
Parents	Nick Clooney Nina Warren
Relatives	Rosemary Clooney (aunt) Jose Ferrer (uncle) Miguel Ferrer (cousin) Rafael Ferrer (cousin)

QUALIANCE

Who is Georges?

Patient ID	Age Category	Age	Gender	Race	Country	Partner Age
1	<89	51	Male	White	Canada	
2	<89	40	Male	Asian	France	
3	<89	53	Male	White	United States	
4	<89	60	Female	Black	Spain	
5	<89	45	Male	Black	Brazil	
6	<89	50	Female	White	Argentina	
7	<89	53	Male	White	United States	
8	<89	53	Male	White	United States	
9	≥89	.	Male	White	Canada	
10	<89	58	Male	White	Canada	

QUALIANCE

Who is Georges?

Patient ID	Age Category 2	Age	Gender	Race	Continent	Partner Age
1	50-59		Male	White	North America	
2	40-49		Male	Asian	Europe	
3	50-59		Male	White	North America	
4	60-69		Female	Black	Europe	
5	40-49		Male	Black	South America	
6	50-59		Female	White	South America	
7	50-59		Male	White	North America	
8	50-59		Male	White	North America	
9	≥89		Male	White	North America	
10	50-59		Male	White	North America	

QUALIANCE

Who is Georges?

	Patient ID	DoB	Age	Gender	Race	Country	Partner Age
	1						
	2						
	3						
	4						
	5						
	6						
	7						
	8						
	9						
	10						

QUALIANCE

Outline of draft principles from EFSPI/PSI DS WG SG5

- Subgroup on ensuring patient data confidentiality:
  - Good rules for DA
  - Role of controlled access
  - Role of legally binding agreements

- WHY?
  - consistent approach across pharma => better for researchers
  - develop standard rules & code => ↑ efficiencies
  - proposal for discussion with EMA (policy 0070 2nd phase consultation)

DA subteam

Data de-ident WG

- Liaison with other groups (TransCelerate/PhUSE)
- High level summary of SG5 draft recommendations

EFSPI/PSI Data Transparency WG SG5 recommendations/principles on data redaction

Assumptions:

- Aim to balance maintaining pt confidentiality vs. data utility
- Data shared in secure & controlled manner (min password protected data access) with legal safeguards on misuse of data.
- Sponsors may provide all datasets for a study or just those required for the research proposal.
- EFPIA/PhRMA principles should be followed as a minimum.
- Recommendations apply to both raw and analysis-ready datasets.
- Genetic data are out of scope for this discussion.

EFSPI/PSI Data Transparency WG SG5 recommendations/principles on data redaction

- Steps applied should be in line with HIPAA & Hrynaszkiewicz principles :
  - Grouping of categories such as age or race should be considered in the presence of uncommon values.
  - Studies involving rare diseases and in small populations need to be assessed on a case by case basis as to whether sufficient steps can be taken to adequately maintain patient confidentiality
- Recommend general principles:

PII = personally identifiable information

  - Removing PII. Includes recoding identifiers (replace original subject id with a new id), remove free text verbatim terms, replace DOB with year of birth or age (recommend age) and replace all dates with study day or shifting using a random offset.
  - Destroy subject id link (code key) between the dataset provided & original dataset.

EFSPI/PSI Data Transparency WG SG5 recommendations/principles on data redaction

- The following indirect identifiers from Hrynaszkiewicz, should be considered for **retention** as removing them may limit data utility:
  - Anthropometry measures (e.g. weight, height) since these are frequently key covariates for dosing (e.g. mg/kg) or exposure.
  - Sex and race (generally mapped according to FDA recommendations) as these are often important factors in understanding disease progression and/or drug effects.
  - Events/endpoints with low frequency since removal of these would generally limit the ability of a researcher to conduct meaningful analyses, particularly in the case of adverse event reporting.

EFSPI/PSI Data Transparency WG SG5 recommendations/principles on data redaction

- Operational aspects to data redaction:
  - QC steps should be taken in order to review and agree on handling of any datapoints which may be in a ‘grey area’ for removal or retention and to check that planned DA steps have been correctly applied
  - Documentation should be provided to researchers outlining the DA steps taken, so researchers are able to understand the limitations.
  - Consider applying DA steps for a study on all datasets at the same time. This ensures that new random subject identifiers are synchronised across all datasets.

Conclusions

- Consider **how** data will be shared as well as **what** will be shared
- Balance data utility vs. risk of patient re-identification
- Guidance already exists as to basic DA steps to apply BUT clear that each study needs to be assessed at some level on a case-by-case basis
- Look out for recommendations from EFSPI/PSI plus PhUSE and TransCelerate
- Further consultation coming with EMA on optimal approach to minimise re-identification of patients and share data

Thanks to:

EFSPI/PSI subgroup 5

Jean-Marc Ferran

George Clooney

Questions?