



Managing Privacy Risk in Data and Document Sharing, Lessons Learned Balancing Progress and Perfection

February 2019



AstraZeneca

d-wise

Julie Holtzople
CTT Operations Director
AstraZeneca



Stephen Baker
Director, Client Manager
d-wise

Disclosure Statement

- *Opinions expressed are personal views of the presenter and are not a statement on behalf of AstraZeneca, d-wise or PhUSE*
- *This information is not legal advice.*







Discussion Topics

- ✓ The global landscape driving data and document sharing
- ✓ Consider techniques and risk in anonymizing IPLD
- ✓ Balancing Data Utility and IPLD Privacy
- ✓ A case study in risk measurement and anonymization
- ✓ How can we leverage experience to drive delivery success

Quiz Time

A stylized world map in the background, showing continents in various colors (brown for North America, green for Europe, yellow for South America, orange for Africa, and light green for Australia and New Zealand). The map is slightly faded and serves as a background for the text box.

**How many global policies exist requiring us
to publish Clinical Documents?**

Several policies are driving global disclosure of clinical documents and patient level data

There are over 90+ Registries around the world



2013

Japan (PDMA)

EMA Policy
0070

2014

2018

US FDA Pilot

Health
Canada

2019

Companies must also address data sharing needs in support of publications and other public commitments throughout the world.

Scope, approach and privacy rules vary

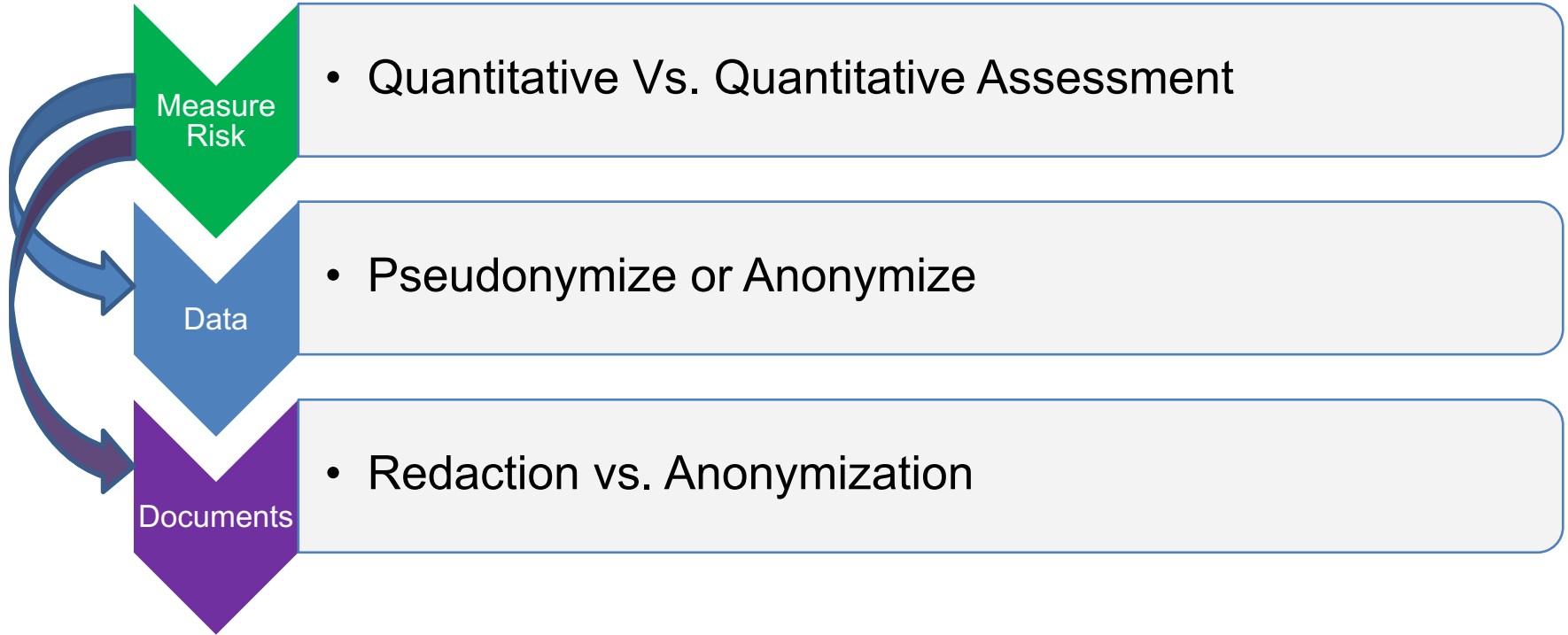
Submission Driven Clinical Data Publication				
	EMA Policy 0070	Health Canada	Japan	FDA (pilot)
Who does the PII and CCI work?	Sponsor	Sponsor	Sponsor	FDA
Who has final say on PII?	Sponsor (but must justify to EMA)	Health Canada	Japan	FDA
Who has final say on CCI?	EMA	Health Canada	Japan	FDA
Technical Requirements	Redaction accepted/ Anonymization encouraged	Anonymization, with low tolerance for Redaction	Redaction ONLY	Redaction ONLY, no transformation; seeking proactive anonymization for original authoring purposes
Individual Patient Level Details	Part 2 of Policy 0070	Yes, TBD	Yes	Not in scope, no plan to release
Which modules are made public	2.5, 2.7.1-4, Module 5 with appendices	2.5, 2.7.1-4, Module 5 with appendices	Module 1 and 2	Pivotal study Mod 5 only - no listings



Discussion Topics

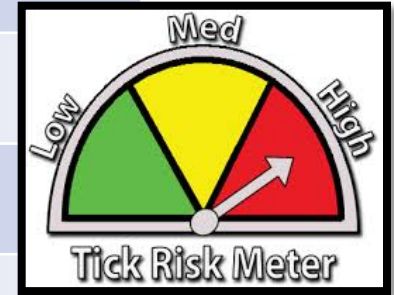
- ✓ The global landscape driving data and document sharing
- ✓ Consider techniques and risk in anonymizing IPLD
- ✓ Balancing Data Utility and IPLD Privacy
- ✓ A case study in risk measurement and anonymization
- ✓ How can we leverage experience to drive delivery success

How do you anonymization IPLD



Qualitative risk assessment relies on the perception of a risk occurring based likelihood and sensitivity

How Identifying	How Sensitive	Overall Risk
Unlikely	Unsensitive	Low
Somewhat	Unsensitive	Medium
Highly	Unsensitive	High
Unlikely	Sensitive	Medium
Somewhat	Sensitive	High
Highly	Sensitive	High



Quantitative risk assessment relies on mathematical calculation of the risk of reidentifying 1 person

- Some calculations rely on the study population and others look more broadly across the disease population
- Emerging standard accepted risk is .09
 - *This means that one individual should not stand alone in a group of less than 11 individuals with the same characteristics as them*
 - *ie. A group of 11 males, ages 30-40 from Poland*



Redaction vs. Anonymization

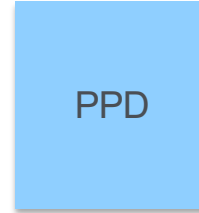
There is different utility in the different ways we mask the data in both data and documents.



Keep



Anonymize



Redact

Higher utility

No Utility

Increased Risk

Lower Risk

Redaction eliminates all sensitive and identifying information fully

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient [PPD] is a [PPD] [PPD] from [PPD] aged [PPD] [PPD] entered the trial on [PPD] [PPD] was hospitalised on [PPD] due to fatigue.

Anonymization beyond redaction attempts to keep some information utility, but eliminate the sensitive details. Example, Scrambling

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient XH37DKWEUWKDO₂ is a PPD PPD from PPD aged PP_D PP_D entered the trial on PPD PP_D was hospitalised on PPD due to fatigue.

Anonymization technique:

Replace Dates with Study day

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient XH37DKWEUWKD is a PPD PPD from PPD aged PP D PP D entered the trial on Day 19 . PP D was hospitalised on Day 54 due to fatigue.

Anonymization technique:

Elevate Countries to Continents

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient XH37DKWEUWKD is a PPD PPD from Europe aged PPD PPD entered the trial on Day 19 . PPD D was hospitalised on Day 54 due to fatigue.

Anonymization technique:

Numeric Band

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient XH37DKWEUWKD is a PPD PPD from Europe aged 30-39 PPD entered the trial on Day 19 . PPD D was hospitalised on Day 54 due to fatigue.

Anonymization technique:

Hiding In Plain Sight in Text

Patient ABCD-123-12345 is a white female from Spain aged 36. She entered the trial on October 14, 2016. She was hospitalised on 18/11/1996 due to fatigue.

Patient ZYXW-098-09876 is a PPD PPD from Europe aged 33 PPD entered the trial on October 26, 2016 . PPD was hospitalised on 30/11/1996 due to fatigue.



Discussion Topics

- ✓ The global landscape driving data and document sharing
- ✓ Consider techniques and risk in anonymizing IPLD
- ✓ **Balancing Data Utility and IPLD Privacy**
- ✓ A case study in risk measurement and anonymization
- ✓ How can we leverage experience to drive delivery success

Which one would you prefer?



- Utility is in the eye of the beholder
- Can we measure utility of the data we anonymize?

Yes. Consider a comparison of the **precision** compared to the **damage** of the data?

How **Precise** is the data?

Variable	Rule / Damage	Precision / Measure
Age	Keep unchanged	1.00
	5 Year age bands	0.75
	10 Year age bands	0.50
	Adult Child Age Bands	0.25
	Drop	0.00

Multivariable precision is also measured



Variable	Rule	Precision	Variable	Rule	Precision	Overall
Age	Keep	1.00	Sex	Keep	1.00	1.00
	5 Year age band	0.75		Keep	1.00	0.875
	10 Year age band	0.50		Keep	1.00	0.75
	Adult Child Age Band	0.25		Keep	1.00	0.625
	Drop	0.00		Keep	1.00	0.5
	Keep	1.00		Drop	0.00	0.5
	5 Year age band	0.75		Drop	0.00	0.375
	10 Year age band	0.50		Drop	0.00	0.25
	Adult Child Age Band	0.25		Drop	0.00	0.125
	Drop	0.00		Drop	0.00	0.00

Weights can be added for researcher priorities

Variable	Rule	Precision	Variable	Rule	Precision	Overall
Age (1)	Keep	1.00	Sex (2)	Keep	1.00	1.00
	5 Year age band	0.75		Keep	1.00	0.91666
	10 Year age band	0.50		Keep	1.00	0.833
	Adult Child Age Band	0.25		Keep	1.00	0.75
	Drop	0.00		Keep	1.00	0.667
	Keep	1.00		Drop	0.00	0.333
	5 Year age band	0.75		Drop	0.00	0.25
	10 Year age band	0.50		Drop	0.00	0.1667
	Adult Child Age Band	0.25		Drop	0.00	0.0833



Discussion Topics

- ✓ The global landscape driving data and document sharing
- ✓ Consider techniques and risk in anonymizing IPLD
- ✓ Balancing Data Utility and IPLD Privacy
- ✓ A case study in risk measurement and anonymization
- ✓ How can we leverage experience to drive delivery success

Case Study:

Example of how measured risk and anonymization can increase the story within the documents

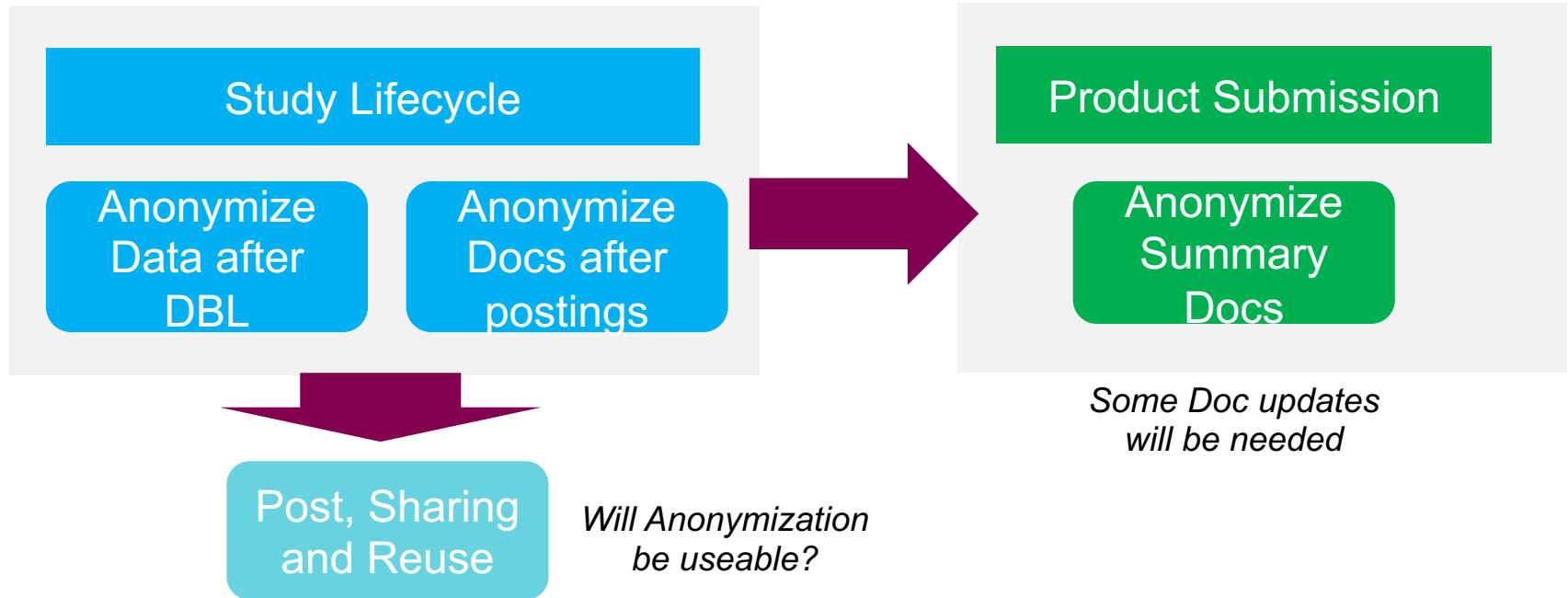
- *Tagrisso – study 9 – published on the EMA Policy 0070 Portal in 2017*
 - *Scramble IDs, Retain Gender, band ages, offset dates*
 - *Look at the story of 1 study participant throughout the document*
 - *Narrative example with date offset*
- *Tagrisso – study 10 – published on the EMA Policy 0070 Portal in 2017*
 - *Scramble IDs, even distribution band for age, retain gender and country*
 - *Look at SAEs, Narrative and Safety Analysis*



Discussion Topics

- ✓ The global landscape driving data and document sharing
- ✓ Consider techniques and risk in anonymizing IPLD
- ✓ Balancing Data Utility and IPLD Privacy
- ✓ A case study in risk measurement and anonymization
- ✓ How can we leverage experience to drive delivery success

A potential future state for anonymization



Global Alignment of standards could make this possible.

What can we do for success in this journey?



- Sponsors should be able to create the package once and submit it to multiple agencies.
- Data Anonymization beyond redaction is a key for usability.
- Measuring risk helps sponsors and individuals.
- Utilize standards where possible.
- What's the best source to use to measure risk?
- How do you scale anonymization to meet global needs?
- How do you deal with the multiple places data is being disclosed and the increased risk with each disclosure?
- Will agencies recognize each others decisions/submissions in full?
- Will guidelines continue to changed?



Discussion



Julie Holtzople
CTT Operations Director
AstraZeneca

d-wise

Stephen Baker
Director, Client Manager
d-wise