TransCelerate BIOPHARMA INC.

A Model Approach to Data Anonymization

26 January 2018

AGENDA

- Setting the scene: TransCelerate

- Balancing Clinical Utility and Participant Privacy

- Pillars of Security

- Clinical Data Transparency Workstream
  - *Clinical Documents*
  - *Patient Level Data*

- Model Approach:
  - *Anonymization of PLD*
  - *Scope & Definitions*
  - *De-identification Steps*
  - *Example*
  - *QC & Recommendations*
  - *Conclusion*

TransCelerate BIOPHARMA INC.

TransCelerate is a not for profit entity created to drive collaboration

Our vision

To improve the health of people around the world by **accelerating and simplifying** the research and development of innovative new therapies.

Our mission

To collaborate across the global research and development community to identify, prioritize, design and facilitate implementation of solutions designed to drive the efficient, effective and high quality delivery of new medicines.

Founded in 2012 by 10 Members

abbvie
AstraZeneca
Boehringer Ingelheim
Bristol-Myers Squibb
gsk GlaxoSmithKline
SANOFI
Lilly
Pfizer
Roche
Johnson & Johnson

TransCelerate BIOPHARMA INC.

Currently, 19 of the World's Most Successful Biopharmaceutical Companies Comprise TransCelerate Membership

2012

TransCelerate Founded

10 Founding Members
5 Active Initiatives

2014

17 Member Companies
12 Active Initiatives
1 Exploratory Initiative

2016

BioCelerate

- TransCelerate launched a subsidiary to focus on preclinical
- Currently 6 members

2017

- ***19 Member Companies***
- ***Launch of 2 Pharmacovigilance Initiatives***
- ***Evolving Clinical Portfolio***
  - *12 Initiatives Delivering Solutions*
  - *7 Initiatives Focused on Facilitating Adoption*
- ***25+ Solutions delivered to industry***

Growth of Global Impact:

- Country Network of 22 Countries
- Engagement with 14+ Global Regulatory Authorities

TransCelerate BIOPHARMA INC.

Portfolio Overview

Active Portfolio
Design, Develop, & Deploy Phase

Patients

1. Clinical Research Awareness
2. Clinical Research Access & Info Exchange
3. eConsent
4. eLabels
5. Patient Experience
6. Patient Technology

Sites

1. Investigator Registry
2. Shared Investigator Platform

Sponsors

1. eSource
2. Quality Management System

Info Sharing Harmonization

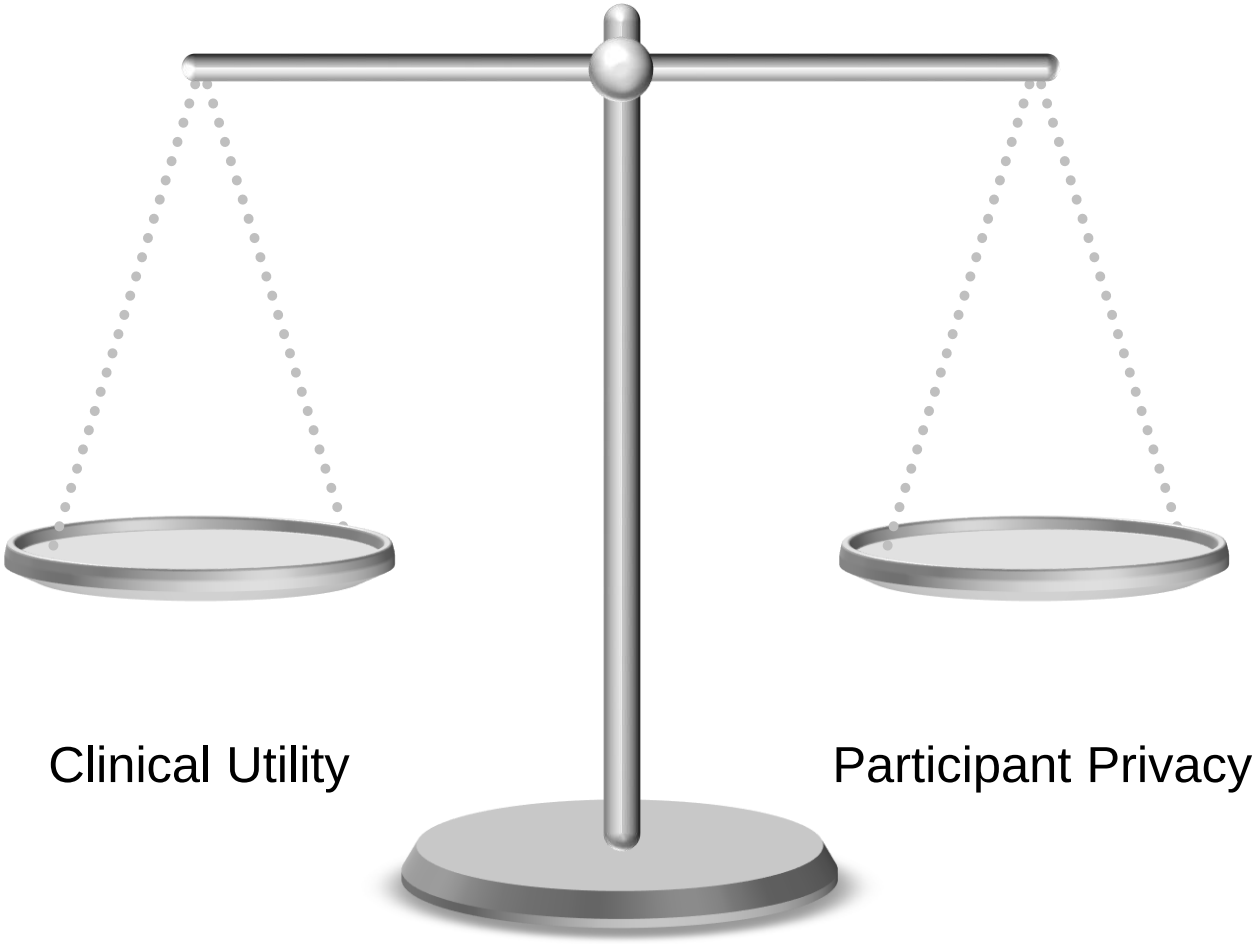
1. Common Protocol Template
2. Data Standards
3. Placebo / Standard of Care

Realization & Governance

1. Clinical Data Transparency
2. Clinical Trial Diversification
3. Comparator Network
4. Pediatric Trial Efficiencies
5. Risk Based Monitoring
6. Site Qualification & Training

TransCelerate BIOPHARMA INC.

Balancing Clinical Utility and Participant Privacy

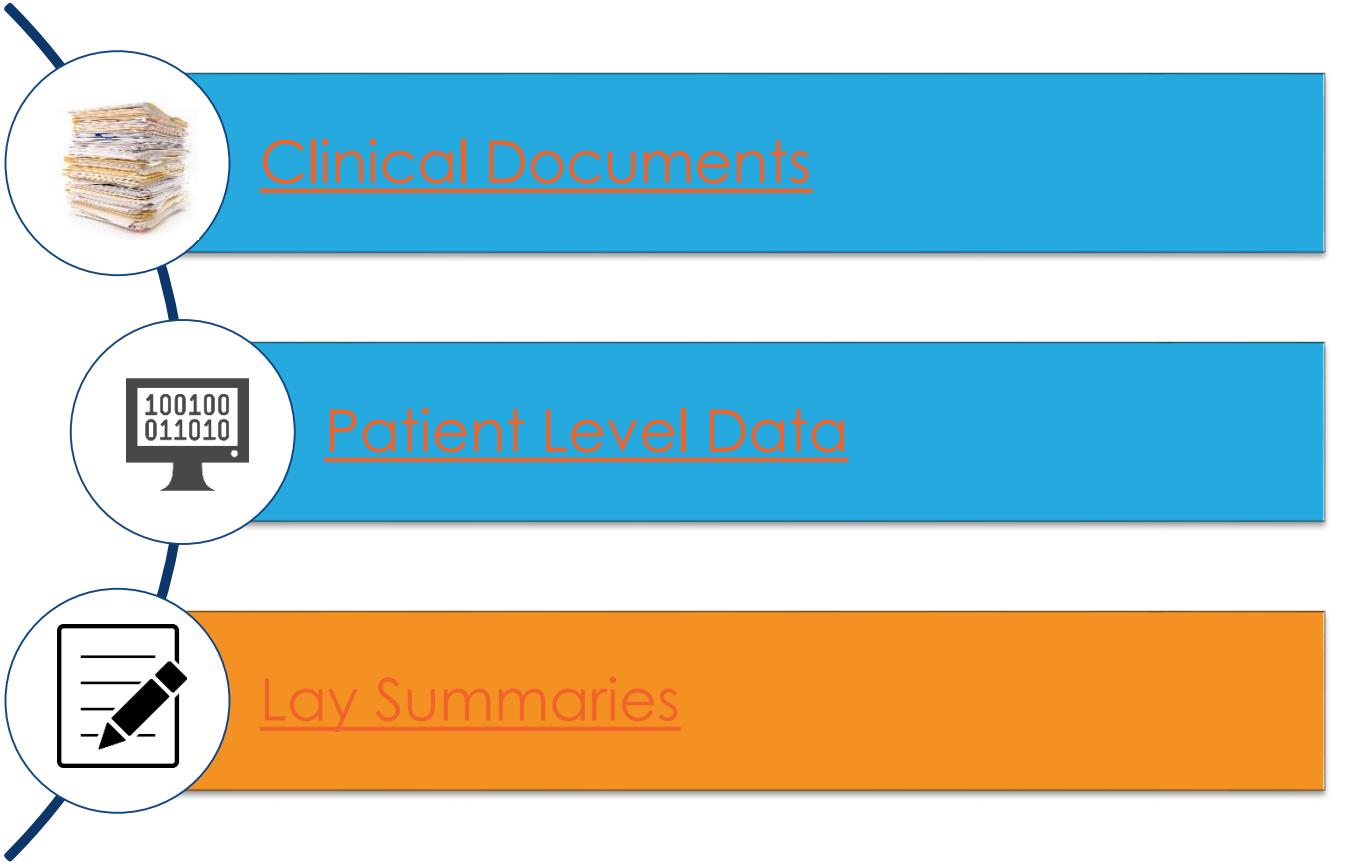


TransCelerate BIOPHARMA INC.

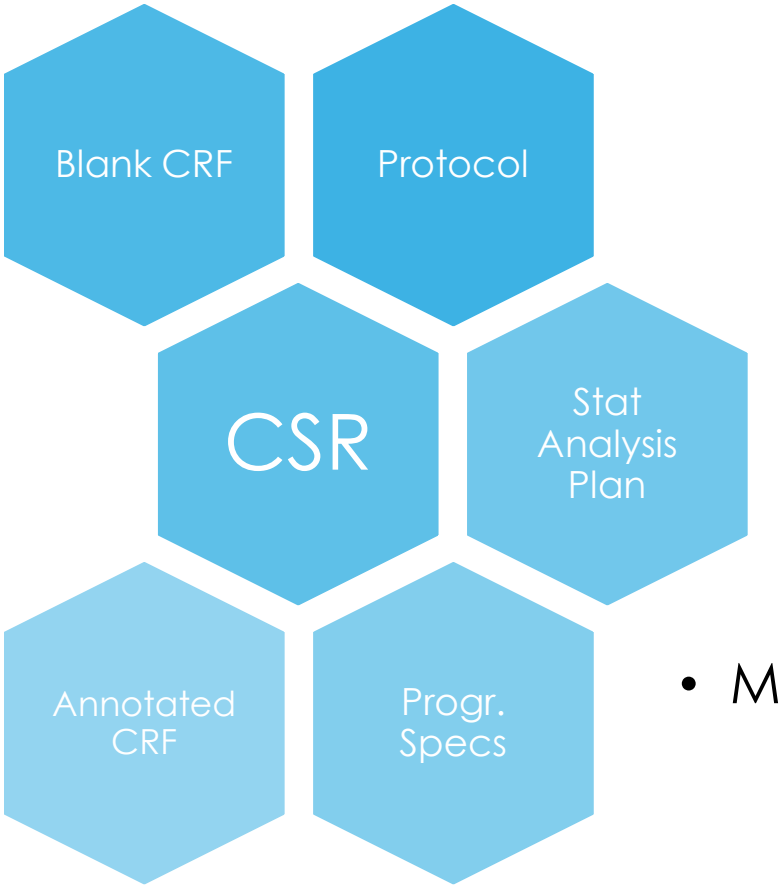
Pillars of Security

TransCelerate BIOPHARMA INC.

Clinical Data Transparency Workstream



Clinical Documents



- Voluntary Sharing
  - Public websites eg, transparency sites or company sites
  - Clinical Study Data Request/YODA
- Mandatory Activities
  - EMA P0070
  - Canada
  - More to follow!

TransCelerate BIOPHARMA INC.

Patient Level Data

- Voluntary Sharing
  - Clinical Study Data Request/YODA

SDTM and ADaM formats

- Mandatory Activities
  - ICMJE Requirements
  - Ct.g
  - Regulations?

Model Approach: Anonymization of PLD

- Background
- Introduction
- Assumptions and Considerations
- Defining Protected Information
- Scope
- Key Topics
- De-identification Steps
- Quality Checks
- Process Recommendations
- Conclusion
- References
- App1: Defining Protected Information
- App2: Summary of Approach
- App3: Ex – removal of PII

Model Approach – Scope & Definitions

- **Scope** – 'all' datasets (not genetic or imaging?); secure sharing

- **Definitions**
- **Data de-identification** - the process by which a dataset is derived in such a way that the data subject is no longer identifiable, ensuring that the risk of re-identifying a participant in a clinical trial, by all reasonably likely means to be used, is small.
- **Anonymization** - a step subsequent to de-identification that involves irreversibly destroying all links between the deidentified datasets and the original datasets. This includes destroying the key code that was used to generate the new identification code numbers from the originals, and destroying the deltas if dates were de-identified using the offset method

Model Approach: De-identification Steps

- **Direct Identifiers**
- **Quasi-identifiers**
- **Dates**
  - Offset: 02APR2017 → 29JUN2017 (+70 days)
  - Relative Day: 02APR2017 → Day 8
- **DOB/Age**
  - Consider EudraCT pediatric age categories
  - HIPAA: Ages >89
- **Dictionaries / Questionnaires**
- **Free-text fields**
- **Sensitive Information / Low Frequency**

Model Approach: Example

Center ID	Investigator ID	Investigator Name	Subject ID	Unique subject ID	Age (years)		AE Start Date	AE End Date	Verbatim Term
00123	279344	Dr Smith	5	TJF00123.0005	57		29DEC2010	27JAN2011	Headache
00123	279344	Dr Smith	2	TJF00123.0002	72		10JAN2011	06APR2011	Nausea
00123	279344	Dr Smith	1	TJF00123.0001	91		25MAR2011	12AUG2011	Cold
05678	333721	Dr Jones	19	TJF05678.0019	85		14OCT2010	20OCT2011	Cold
05678	333721	Dr Jones	4	TJF05678.0004	53		24MAY2011	.	Headache
05678	333721	Dr Jones	23	TJF05678.0023	76		01MAR2011	15MAR2011	Pain

New center ID	New Investigator ID	Remove	New Subject ID	New Unique Subject ID	Remove Ages >89	Create new age category	Add offset to each date	Add offset to each date	Remove

Center ID	Investigator ID	Investigator Name	Subject ID	Unique subject ID	Age (years)	Age Category (years)	AE Start Date	AE End Date	Verbatim Term
03145	148227		8754	EHR03145.8754	57	<=89	02FEB2011	03MAR2011	
03145	148227		5681	EHR03145.5681	72	<=89	09NOV2010	03FEB2011	
03145	148227		1475	EHR03145.1475	.	>89	03JUL2011	20NOV2011	
90876	687208		1457	EHR90876.1457	85	<=89	06JUL2010	12JUL2011	
90876	687208		2214	EHR90876.2214	53	<=89	03MAY2011		
90876	687208		2236	EHR90876.2236	76	<=89	08MAR2011	22MAR2011	

TransCelerate BIOPHARMA INC.

Model Approach: QC and Recommendations

- This is not an activity where you can blindly just press the 'Anonymize' button! Manual review is very important.

- Where are you planning to share the anonymized data (and associated redacted/anonymized documents)?
- Review of Requests → IRP?
- Data Sharing Agreements
- Which Data & Docs Provided?
- How much to anonymize vs how much to share?

- Regular Review of Procedures

TransCelerate BIOPHARMA INC.

Model Approach: Conclusion

- A common approach to anonymization increases utility for researchers (multi-sponsor data sharing sites)
- As technology advances, and more data sources are made publicly available, anon processes will need regular review
- The changing Regulatory landscape means that anonymization should not be considered a stand-alone task. Strategies are needed to automate (as far as possible) and integrate within standard study reporting processes.
- A Data Sharing Plan for each study to connect processes:
  - anonymizing datasets
  - registration & results disclosure (ct.g fields)
  - Publication planning (ICMJE statement)