Maximizing Research Impact through Clinical Trial Transparency: *Leveraging the EFPIA Anonymization Gradient*

HC
Holtzople
Consulting

Julie Holtzople
Independent Consultant

February 2025

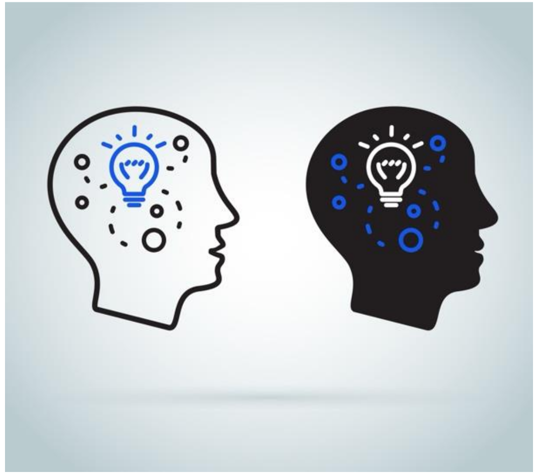
Objectives

- Explore How Clinical Trial Transparency Policies Can Foster Value for the Health Care Ecosystem
- Introduce the Anonymization Gradient and how to use it to enable tailored anonymization for different types of data sharing.
- Highlight the importance and possibilities for customizing datasets to support secondary research success

Table of Contents

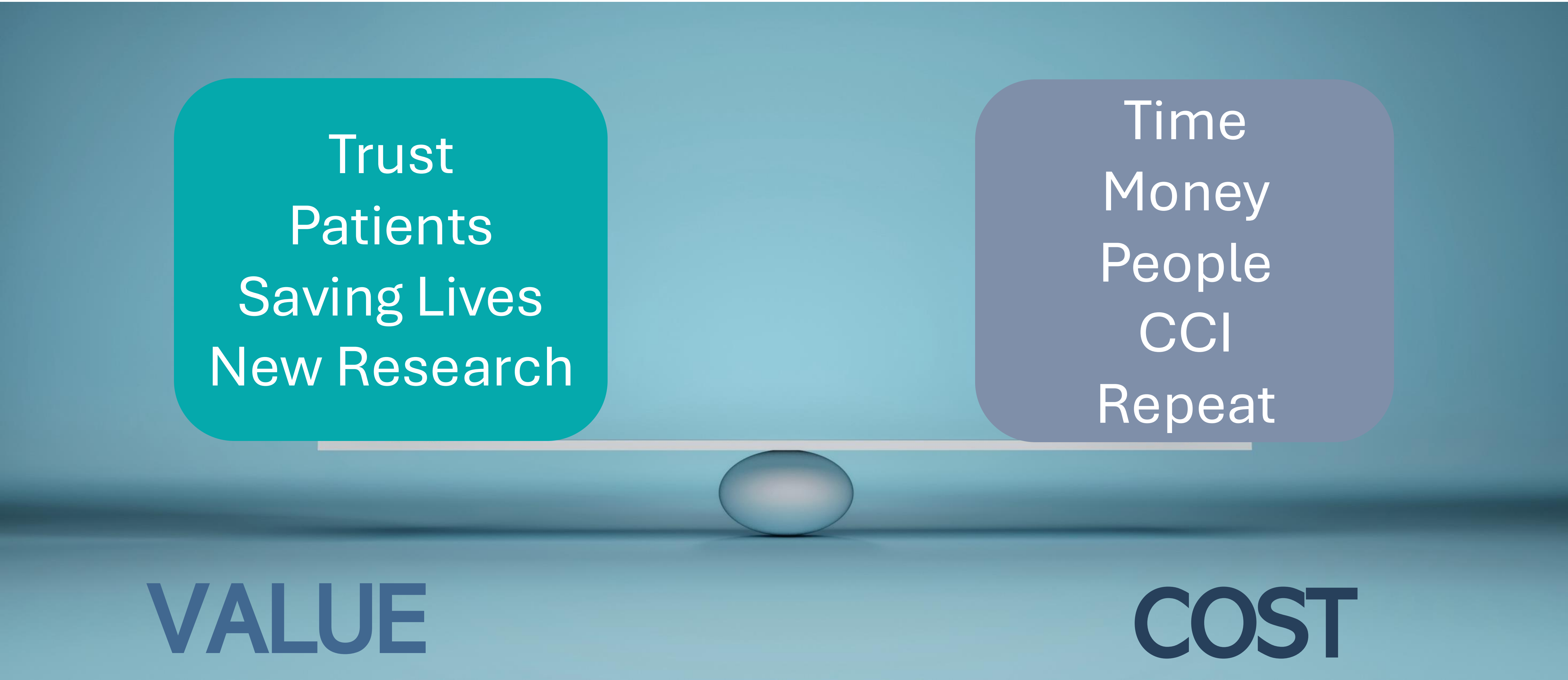
1. Deriving Value from Clinical Trial Transparency Policies, Processes and Standards
2. What is the EFPIA Anonymization Gradient?
3. Applying the Anonymization Gradient to support various data sharing scenarios
4. Conclusions

DERIVING VALUE FROM TRANSPARENCY POLICIES



- Transparency **regulations** and best practices have been evolving and growing across the globe since the first registries appeared in the early 2000s.
  - EUDRACT in 2004
  - Clinicaltrials.gov in 2008
- Transparency can be a **strategic value** for clinical research sponsors, or it can be a compliance exercise.
- As **citizens and patients**, we should all advocate for driving strategic value from the transparency activities we solution and deliver along side the exciting field of clinical research.

Purposeful Clinical Trial Transparency delivers value to the health care ecosystem and manages costs.



Trust
Patients
Saving Lives
New Research

Time
Money
People
CCI
Repeat

VALUE

COST

Sponsors should be Purposeful with their policies and processes to ensure clinical research benefits everyone

- How can sponsors support secondary research sharing their clinical trial data safely and successfully?

- Is anonymization always done the same way, regardless of who we share with and where we are sharing the data?

- Which of your transparency deliveries is providing the greatest value to society?

Consider anonymization is contextual. It is more then how you transform data.

Contractual Controls

- Are users named?
- Are users legally liable to protect privacy?
- Are users held accountable with legal consequence?
- How long do they have access to the data?

Organizational controls

- Environment secure, monitored, etc.?
- Can data be combined with other data?
- Are user actions traceable?

Technical Controls

- How is the data Anonymized?
- Measured risk or not?
- Is all data accounted for in the risk measurement such as adverse events and medical history?

Do all 3 controls always need to be High in all sharing scenarios?

HC
Holtzople
Consulting

Table of Contents

1. Deriving Value from Clinical Trial Transparency Policies, Processes and Standards
2. What is the EFPIA Anonymization Gradient?
3. Applying the Anonymization Gradient to support various data sharing scenarios
4. Conclusions

Why did EFPIA develop the Anonymization Gradient?

Why does the Pharmaceutical Industry Anonymize Data ?

To facilitate the the sharing of research data, which contributes to the development of more effective and safer medicines;

To achieve a high level of patient protection by minimising the potential harms from personal data being used

To preserve trust of patients and the research community

There is no clear and consistent approach to data anonymization that is sufficiently robust for all stakeholders.

A clear technical and organizational standard remains elusive.

Available guidance and documentation is complex and fragmented.

Anonymization should be context specific.

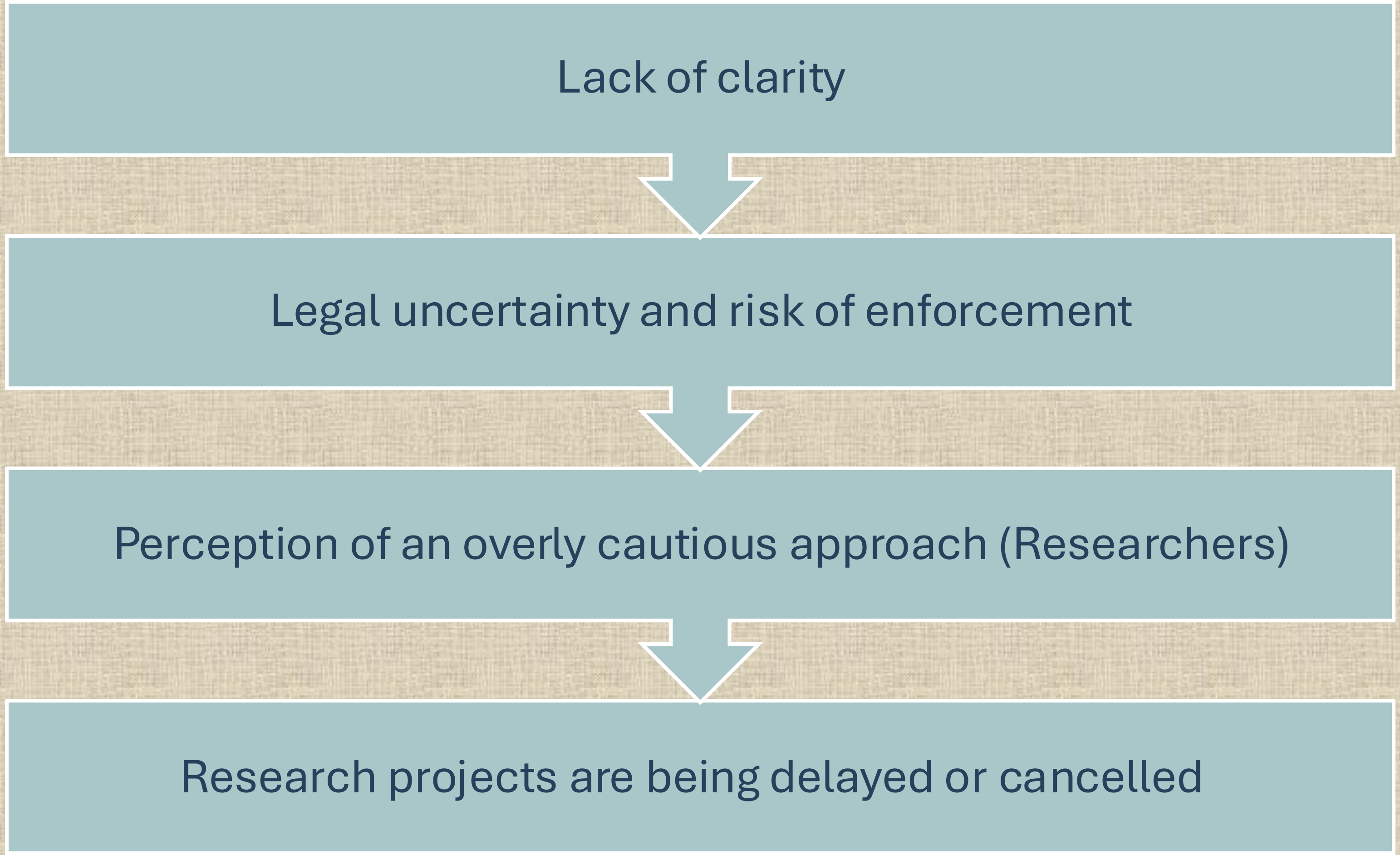
The current requirements around anonymization can be restrictive and lack options for an approach based on the absence of a reasonable reidentification.

The regulatory landscape continues to be complex with the introduction of new rules on “non personal data.”

HC
Holtzople
Consulting

efpia
European Federation of Pharmaceutical Industries and Associations

Consequences of the current landscape impact value and progress in research



Lack of clarity

Legal uncertainty and risk of enforcement

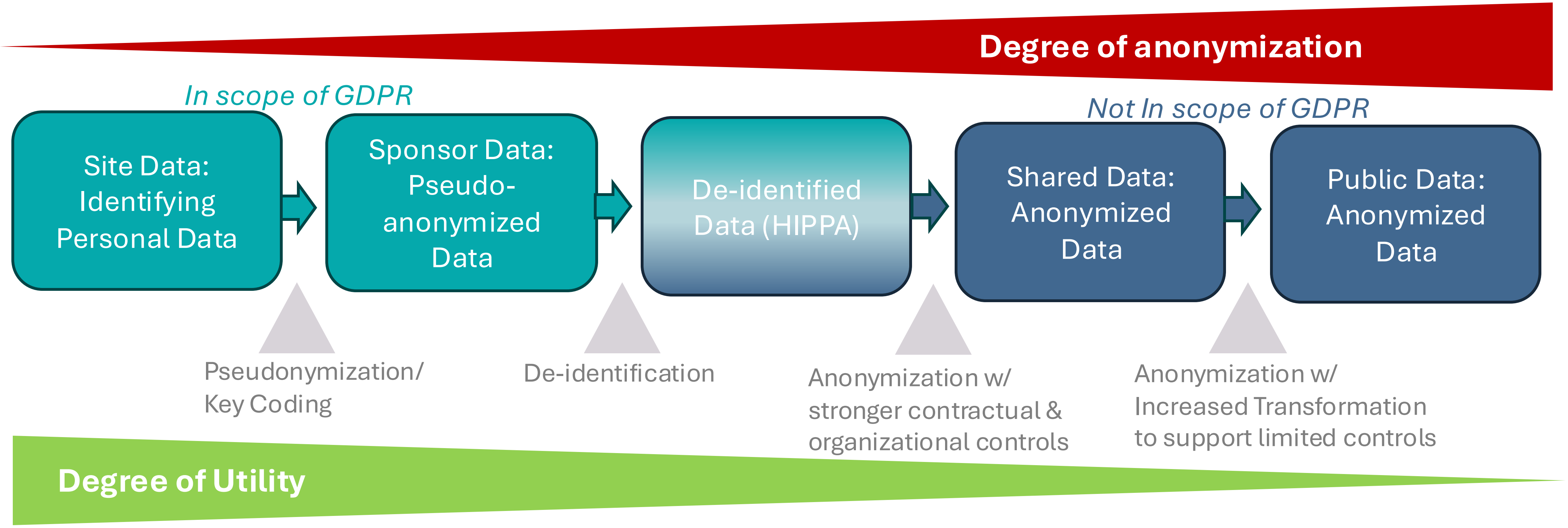
Perception of an overly cautious approach (Researchers)

Research projects are being delayed or cancelled

Data Anonymization: Balancing privacy and progress

A visual aid exploring the trade-offs researchers face in protecting patients' identities and sharing data to advance science.

- EFPIA finds the objectives best served by **a risk-based, context-specific approach to anonymisation.**
- Thus, the Anonymization Gradient was developed as such a tool


Degree of anonymization
In scope of GDPR
Not In scope of GDPR
Site Data: Identifying Personal Data
Sponsor Data: Pseudo-anonymized Data
De-identified Data (HIPPA)
Shared Data: Anonymized Data
Public Data: Anonymized Data
Pseudonymization/ Key Coding
De-identification
Anonymization w/ stronger contractual & organizational controls
Anonymization w/ Increased Transformation to support limited controls
Degree of Utility

HC Holtzople Consulting

efpia European Federation of Pharmaceutical Industries and Associations

Considerations for each data sharing scenarios

Consider various data sharing scenarios

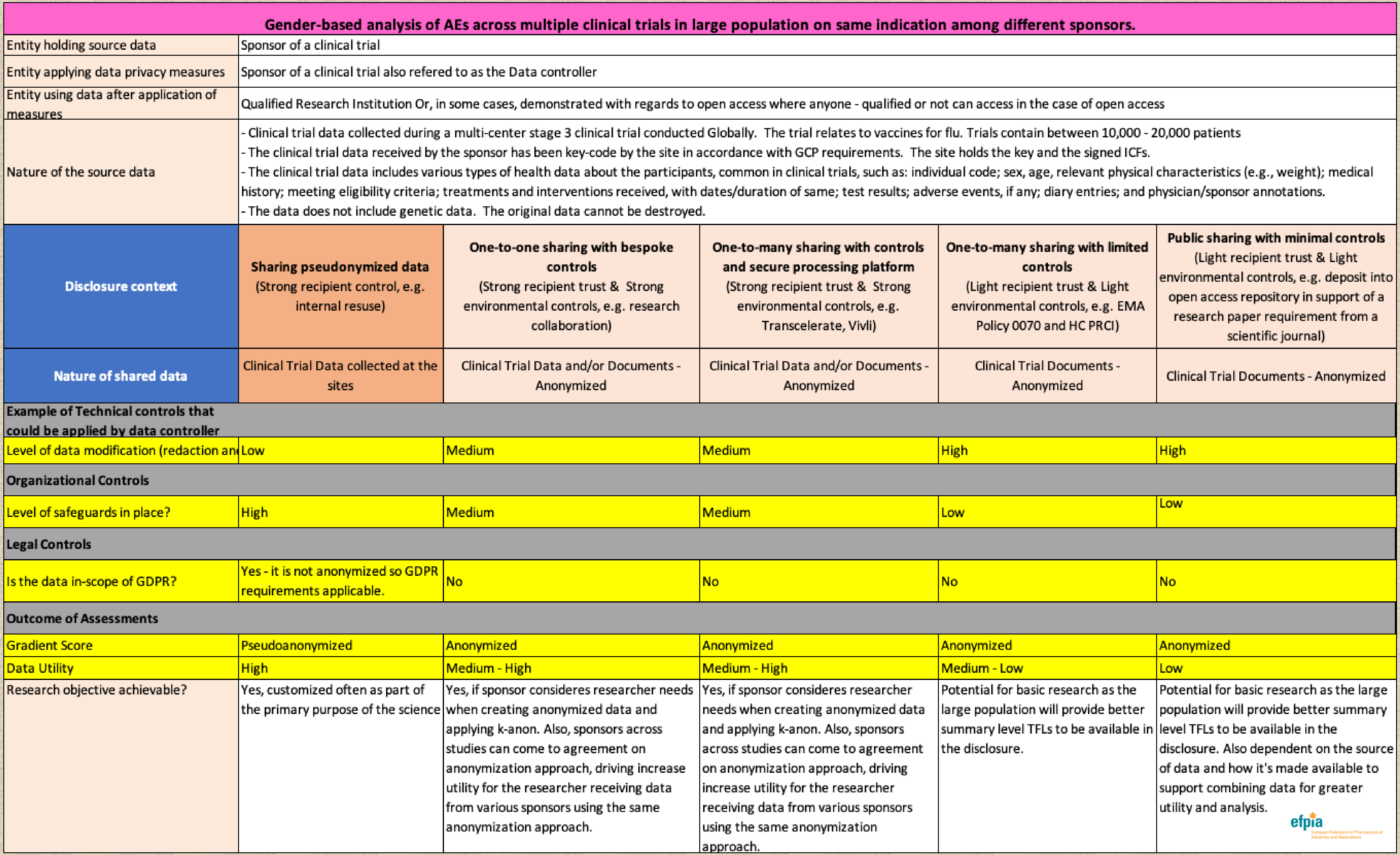
- **Small** Population (50-100 patients), small number of sites
- **Medium** Population size (2,000 patients), common disease, multi-site in EU, Phase 3
- **Large** population size (20,000 patients), Phase 3, Multi-site global

Consider the various Controls available to Sponsors

- **Technical Controls** - Base Transformation, Transformation based on quantitative risk assessment, Transformation based on qualitative Risk assessment, and Alternatives
- **Organization Controls** - Individual vs. Entity, Environmental controls, Duration of access, Can you add ore data or not? Can you export the data?
- **Legal Controls** - In scope of GDPR, Contractual control strength, Vetting of requestor

Consider outcomes of the data use based on:

- **Gradient score** – Pseudonymized or Anonymized
- **Data Utility** – High, Medium or Low
- Can the Research objectives be achieved easily? – Yes, No, Maybe



Gender-based analysis of AEs across multiple clinical trials in large population on same indication among different sponsors.					
Entity holding source data	Sponsor of a clinical trial				
Entity applying data privacy measures	Sponsor of a clinical trial also refered to as the Data controller				
Entity using data after application of measures	Qualified Research Institution Or, in some cases, demonstrated with regards to open access where anyone - qualified or not can access in the case of open access				
Nature of the source data	- Clinical trial data collected during a multi-center stage 3 clinical trial conducted Globally. The trial relates to vaccines for flu. Trials contain between 10,000 - 20,000 patients - The clinical trial data received by the sponsor has been key-code by the site in accordance with GCP requirements. The site holds the key and the signed ICFs. - The clinical trial data includes various types of health data about the participants, common in clinical trials, such as: individual code; sex, age, relevant physical characteristics (e.g., weight); medical history; meeting eligibility criteria; treatments and interventions received, with dates/duration of same; test results; adverse events, if any; diary entries; and physician/sponsor annotations. - The data does not include genetic data. The original data cannot be destroyed.				
Disclosure context	**Sharing pseudonymized data** (Strong recipient control, e.g. internal resuse)	**One-to-one sharing with bespoke controls** (Strong recipient trust & Strong environmental controls, e.g. research collaboration)	**One-to-many sharing with controls and secure processing platform** (Strong recipient trust & Strong environmental controls, e.g. Transcelerate, Vivli)	**One-to-many sharing with limited controls** (Light recipient trust & Light environmental controls, e.g. EMA Policy 0070 and HC PRCI)	**Public sharing with minimal controls** (Light recipient trust & Light environmental controls, e.g. deposit into open access repository in support of a research paper requirement from a scientific journal)
Nature of shared data	Clinical Trial Data collected at the sites	Clinical Trial Data and/or Documents - Anonymized	Clinical Trial Data and/or Documents - Anonymized	Clinical Trial Documents - Anonymized	Clinical Trial Documents - Anonymized
Example of Technical controls that could be applied by data controller					
Level of data modification (redaction an	Low	Medium	Medium	High	High
Organizational Controls					
Level of safeguards in place?	High	Medium	Medium	Low	Low
Legal Controls					
Is the data in-scope of GDPR?	Yes - it is not anonymized so GDPR requirements applicable.	No	No	No	No
Outcome of Assessments					
Gradient Score	Pseudoanonymized	Anonymized	Anonymized	Anonymized	Anonymized
Data Utility	High	Medium - High	Medium - High	Medium - Low	Low
Research objective achievable?	Yes, customized often as part of the primary purpose of the science	Yes, if sponsor considered researcher needs when creating anonymized data and applying k-anon. Also, sponsors across studies can come to agreement on anonymization approach, driving increase utility for the researcher receiving data from various sponsors using the same anonymization approach.	Yes, if sponsor consideres researcher needs when creating anonymized data and applying k-anon. Also, sponsors across studies can come to agreement on anonymization approach, driving increase utility for the researcher receiving data from various sponsors using the same anonymization approach.	Potential for basic research as the large population will provide better summary level TFLs to be available in the disclosure.	Potential for basic research as the large population will provide better summary level TFLs to be available in the disclosure. Also dependent on the source of data and how it's made available to support combining data for greater utility and analysis.

HC
Holtzople
Consulting

efpia
European Federation of Pharmaceutical Industries and Associations

Opportunities for further considerations

1. Handling of **Medical Information** for individual patients needs to be considered carefully given the sharing scenario
    - Much safer to share in a controlled scenario with a DSA that indicates researcher is not trying to reidentify anyone
2. **Reference vs. Study Population** risk measurement options
    - Reference Population can enable greater utility and should be considered for more secure data sharing scenarios such as Vivli. DSA enables this.
3. When is Pseudo-anonymized data sharing going to be allowed?
    - Consider new EDPO position paper
4. Building out scenarios that enable organizations to consider how to shape their policies and **evolve from one size fits all solutions.**

Table of Contents

1. Deriving Value from Clinical Trial Transparency Policies, Processes and Standards
2. What is the EFPIA Anonymization Gradient?
3. Applying the Anonymization Gradient to support various data sharing scenarios
4. Conclusions

Some use cases to consider

	Vivli Sharing (Secondary Use)	HC PRCI Publication (Secondary Use)	Internal Data Reuse (Secondary Use)	Research Collaboration (Secondary Use)	Research partner during the study (Primary Use)
Contractual Controls	• High	• Very Low	• Very High	• Very High	• Very High
Organizational Controls	• High	• Very Low	• Very High	• Medium (when using platform allowing data export)	• High (shared database inside Pharma using top notch security protocols)
Technical Controls	• Minimal Transformation • Reference Population	• High Transformation and/or Masking • Study Population	• Minimal Transformation • Reference Population	• Medium Transformation • Reference Populations	• Site Coded Data • No transformation because data is being analyzed as per the ICF to support study objectives
Output	Anonymized & High Utility	Anonymized & Low Utility	Anonymized & High Utility	Anonymized & High Utility	Pseudo-anonymized & High utility

HC
Holtzople
Consulting

USE CASE 1

Sponsor is asked to prepare a dataset for sharing with researcher at a University

Researcher would like to conduct research on the dataset to test his hypothesis that younger patients respond better to the treatment they underwent during the clinical trial. (Age required)

- The Pharma company sponsored a clinical trial and collected data from 2000 participants
- The clinical trial data includes:
  - individual code; sex, age, relevant physical characteristics (e.g., weight); medical history; meeting eligibility criteria; treatments and interventions received, with dates/duration of same; test results; adverse events, if any; diary entries; and physician/sponsor annotations.
  - The data does not include genetic data.
- All study participants have diabetes.
- Research sites were in the EU.
- Secondary research to be conducted in South Africa.

Strong contractual controls, Strong Organizational controls

- One Recipient at the University will access the data.
  - The data will be shared via a private portal that has high security controls and is maintained to highest standards.
  - The data will be accessible for a limited period of 12 months.
  - The data cannot be exported.
  - Data cannot be combined with other data.
- There is a DSA in place with the university.
- There's a history of successful collaboration.

USE STUDY 1

Sponsor is asked to prepare a dataset for sharing with researcher at a University

Potential approach to anonymise the dataset to achieve the objective:

Study Sponsor policy allows for less transformation when there are high contract and organizational controls in place. Thus, the follow technical controls are applied:

- Scramble the patient IDs
- Offset dates
- Retain medical information except sensitive and/or identifying comments
- Apply a k-anon risk measurement on secondary identifiers with a Risk Threshold of .50 prioritizing age to be retained in the risk assessment
- Do not transform data on the age and treatment outcome

USE STUDY 1

Sponsor is asked to prepare a dataset for sharing with researcher at a University

What is the result of this data sharing scenario?

OUTCOMES:

- Researcher receives high utility data and completes their analysis in 12 months.
- Patient privacy is retained due to all controls in place.
- Public Health is advanced when the researcher publishes in scientific journal to share their findings with the medical community.

USE CASE 2

Sponsor wants to deposit data in an open access portal to support publication

The purpose of future research is unknown. The data is being deposited upon request of the journal to achieve publication requirements.

- The academic sponsor conducted a clinical trial that included approximately 2000 participants .
- The clinical trial data includes:
  - individual code; sex, age, relevant physical characteristics (e.g., weight); medical history; meeting eligibility criteria; treatments and interventions received, with dates/duration of same; test results; adverse events, if any; diary entries; and physician/sponsor annotations.
  - The data does not include genetic data.
- All study participants have lung cancer.
- Research sites are in the EU.
- Publication portal hosted in the US.

Weak contractual controls, Weak Organizational controls

- Unknown who will access the data.
  - The data will be shared via open access portal on the internet.
  - The data will be accessible forever.
  - The data can be exported.
  - Data can be combined with other data.
- There is a terms of use checkbox to access the data.
- There is no verification of identity or background check for anyone using the data.

HC
Holtzople
Consulting

USE CASE 2

Sponsor wants to deposit data in an open access portal to support publication

Potential approach to anonymise the dataset to achieve the objective?

Study Sponsor policy requires high transformation controls due to weak contractual and organizational controls. Thus, the follow technical controls are applied:

- Scramble the patient IDs
- Offset dates
- Redact medical information at the individual level
- Retain all summary level/aggregated data
- Apply a k-anon risk measurement on secondary identifiers with a Risk Threshold of .09 prioritizing gender to be retained if possible.
- Ensure all sensitive information is redacted.

USE CASE 2

Sponsor wants to deposit data in an open access portal to support publication

What is the result of this data sharing scenario?

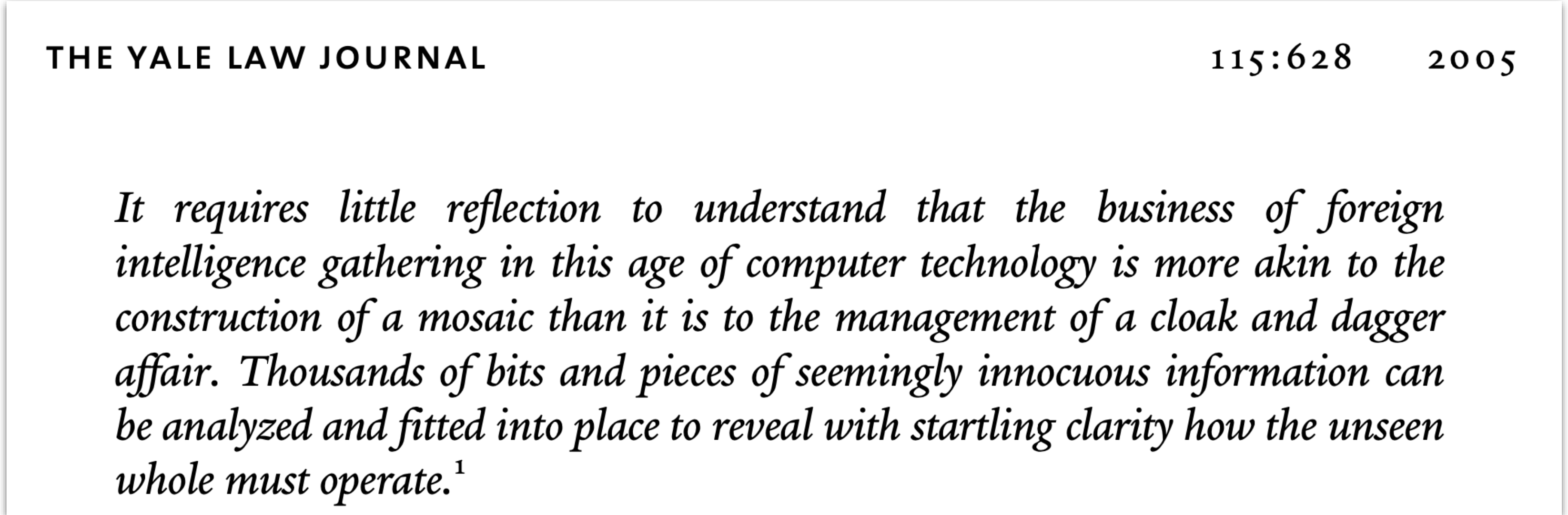
OUTCOMES:

- Data is published with low utility.
- Patient privacy risks are higher as the ability to combine this data with additional data remains the life of this database and any downloaded copies
- Consider the future for these patients
  - Mosaic Theory
  - Changing nature of technology and data access

HC
Holtzople
Consulting

”The Mosaic Theory”

“The “mosaic theory” describes a basic precept of intelligence gathering: Disparate items of information, though individually of limited or no utility to their possessor, can take on added significance when combined with other items of information. Combining the items illuminates their interrelationships and breeds analytic synergies, so that the resulting mosaic of information is worth more than the sum of its parts. “



THE YALE LAW JOURNAL 115:628 2005

> *It requires little reflection to understand that the business of foreign intelligence gathering in this age of computer technology is more akin to the construction of a mosaic than it is to the management of a cloak and dagger affair. Thousands of bits and pieces of seemingly innocuous information can be analyzed and fitted into place to reveal with startling clarity how the unseen whole must operate.*[1]

Reference: https://www.yalelawjournal.org/pdf/358_fto38tb4.pdf

HC
Holtzople
Consulting

Technology and Data Access continue to grow at lightening pace

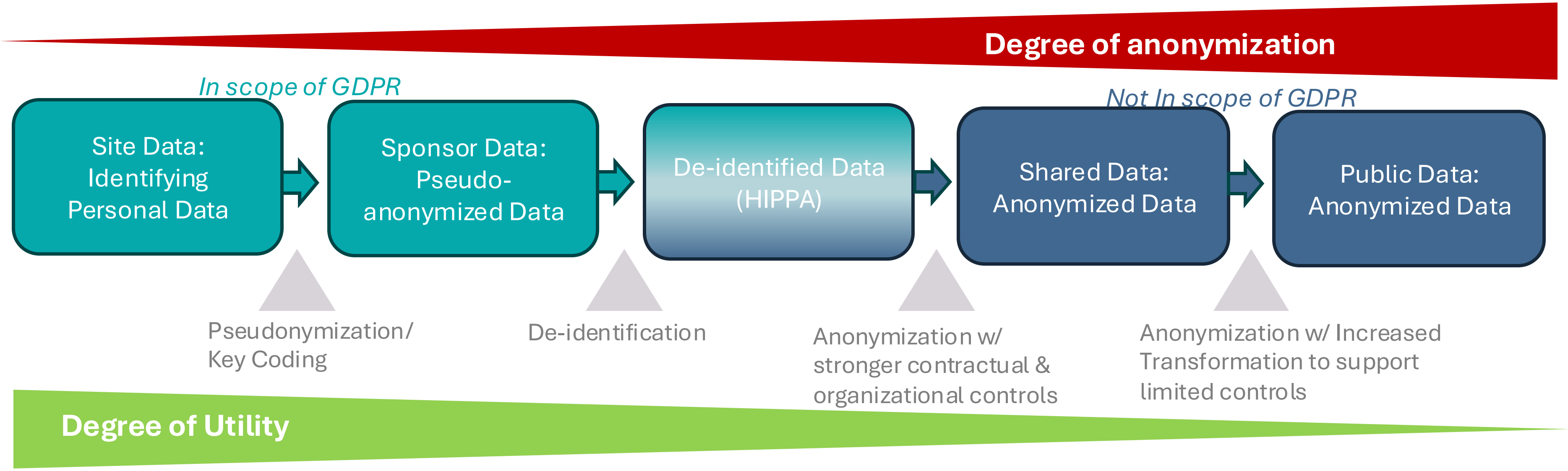
- **More and more disclosures of the same patient data in different formats globally**
  - CSRs published at multiple data cut offs, scientific publications, clinical trial registries, safety data reporting, etc.
- Patients publishing their own data on **social media**, not understanding it can be used to match locations and treatments in clinical trial datasets
- Public nonclinical sources, such as **news reports and police reports** make accessing anything unique or noteworthy easy
- **Data marts** are growing
  - A data mart is a subject-oriented database containing transactional data (rows and columns), which makes it easy to access, organize, and understand. It contains historical data brought together used to understand trends.
- **Large Language Models** make combination of data much easier

Table of Contents

1. Deriving Value from Clinical Trial Transparency Policies, Processes and Standards
2. What is the EFPIA Anonymization Gradient?
3. Applying the Anonymization Gradient to support various data sharing scenarios
4. Conclusions

There is a tradeoff between the level of privacy and data utility in all data sharing scenarios

- Different solutions work best in different scenarios when you consider the various controls that are in place.
- The Anonymization Gradient can help your organization consider various data sharing scenarios



Degree of anonymization
In scope of GDPR
Not In scope of GDPR
Site Data: Identifying Personal Data
Sponsor Data: Pseudo-anonymized Data
De-identified Data (HIPPA)
Shared Data: Anonymized Data
Public Data: Anonymized Data
Pseudonymization/ Key Coding
De-identification
Anonymization w/ stronger contractual & organizational controls
Anonymization w/ Increased Transformation to support limited controls
Degree of Utility

HC Holtzople Consulting

efpia European Federation of Pharmaceutical Industries and Associations

Anonymization requires careful consideration of all controls and the research goals

Anonymization is a protective measure.

Use a combination of the different controls to ensure patients are protected in all data sharing scenarios. This is critical to meet the commitments in the ICF and secure the future of clinical research

Anonymization is not a fully automated process.

Using the same set of predefined anonymization measures in all data sharing scenarios is not recommended and will not drive utility.

Anonymization is not an absolute concept. No "one size fits all."

Anonymization depends on multiple factors, including the context of the data sharing and the recipients.

Anonymization reduces data utility.

The more data is transformed to make it anonymous, the more its research options are reduced.

Consider researcher needs within your anonymization approach. Incentivize research in controlled environments to provide greater utility.

HC
Holtzople
Consulting

Anonymization is not one size fits all

Thank you