

Synthetic Data in Biostatistics: Definitions, Regulatory Implications, and Practical Applications

Chair: Khadija Rantell (MHRA, UK), Pierre Mancini (Sanofi, FR)

Acknowledgment:

Jenny Devenport (Merck Healthcare KGaA) and
Florian Klinglmüller (Austrian Agency for Health and Food Safety)

Outline of the session



Fireside chat:

Professors Theresa Stadler (Swiss Data Science Centre) and Matthias Templ (University of Applied Sciences and Arts Northwestern Switzerland)



Talk: Use of synthetic data in clinical trials: a stakeholder's journey to overcome limitations of standard clinical trials Panel

Speakers: Andrea Manfrin (MHRA, UK) and Federico Mattiello, Francesco Brizzi (Roche, CH)



Panel discussion:

Speakers, Prof Kit Roes (EMA, NL) and Lukas Aguirre Dávila (Paul-Ehrlich-Institut [PEI], DE)

Synthetic Data Fireside Chat Questions

Conversation with Professors Theresa Stadler and Matthias Templ

1. To set a common baseline for today, how do you define synthetic data, and what makes data truly synthetic in a regulatory context?
2. What makes modern synthetic data generation fundamentally different from classical simulations such as Monte Carlo simulations?
3. Synthetic data inherit characteristics of the data used to generate them.
What utility metrics should be used for assessing quality, and do we need a different framework for analysing such data?

Synthetic Data Fireside Chat Questions

Conversation with Professors Theresa Stadler and Matthias Templ

4. Where are you seeing the most promising traction for synthetic data today?
5. Does synthetic data solve privacy data-sharing issues better than traditional anonymization methods and, if yes, how do we evaluate its benefit?
6. Where can we safely use synthetic data to augment datasets and sources of evidence?
What are the typical use cases for data augmentation?

Synthetic Data Fireside Chat Questions

Conversation with Professors Theresa Stadler and Matthias Templ

7. How do we generate synthetic data that preserve complex multi-dimensional relationships across clinical trial tables without losing the joint distribution?
8. **To close our discussion: What critical question have we not touched on today that everyone needs to ask before they trust synthetic data for a trial?**

Thank you