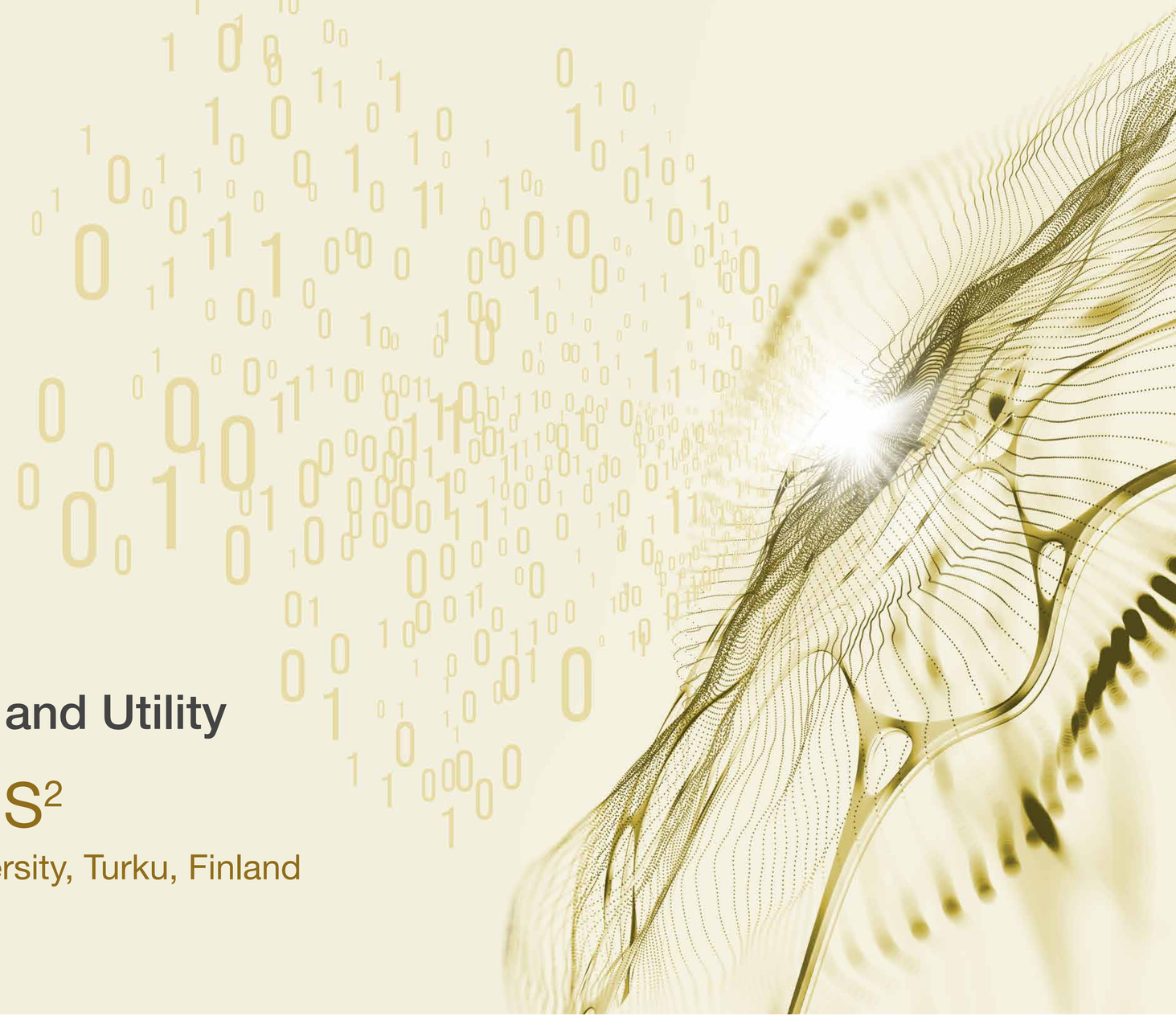
RWD143

SYNTHETIC HEALTH DATA

Synthetic Health Data at a Glance: Striking a Balance between Privacy and Utility

Hermansson LL[1], Parvanian S[2]

[1]BCB Medical Ltd, Djursholm, Sweden, [2]Abo University, Turku, Finland



OBJECTIVES:

Hacking and unauthorized disclosures are two major causes of data breaches in the healthcare industry. It is estimated that just in 2021 more than 40 million healthcare records have been exposed illegally in the United States. Synthetic data which defines as artificially manufactured data has gained significant attention lately due to its capability in protecting patient privacy and augmenting clinical research. According to Gartner, while until 2024, 60% of the data used for AI development will be synthetically generated, synthetic data will entirely exceed real data in AI models by 2030.

METHODS:

We utilized survey data obtained from the eight interviews with Finnish academia, pharma industries, and healthcare authorities on the business potential of synthetic data. The data is presented as a risk assessment evaluation and with required actions on the topic.

RESULTS:

While scientists from academia focus more on improving the quality of generated synthetic data as an alternative to real-world data, industries including authorities emphasize the lack of sufficient evidence on the utility of synthetic data. To address the barriers to unlocking the full potential of synthetic data, we categorized the challenges into three parts: 1- technological knowledge including insufficiency in size and quality of the training dataset, biased datasets, and lack of clinical-quality measures, evaluation metrics, data authentication and security measures 2- political and regulatory challenges due to a lack of refined regulatory frameworks 3- commercialization challenges such as failure to create an efficient business model, time-consuming B2B preparations due to formalities and involved authorities, reduced level of professional acceptance and scientific support material.

CONCLUSIONS:

To unlock the true potential of synthetic data there is a need to enhance algorithms to maximize data utilization with the minimum de-identification, deploy experimentation and human-in-the-loop evaluation metrics, strengthen collaboration networks, and provide high-quality research, and liaison with authorities.

Abstract ID#: **120274**

Submitter's E-mail Address: **lisse-lotte.hermansson@bcbmedical.com**