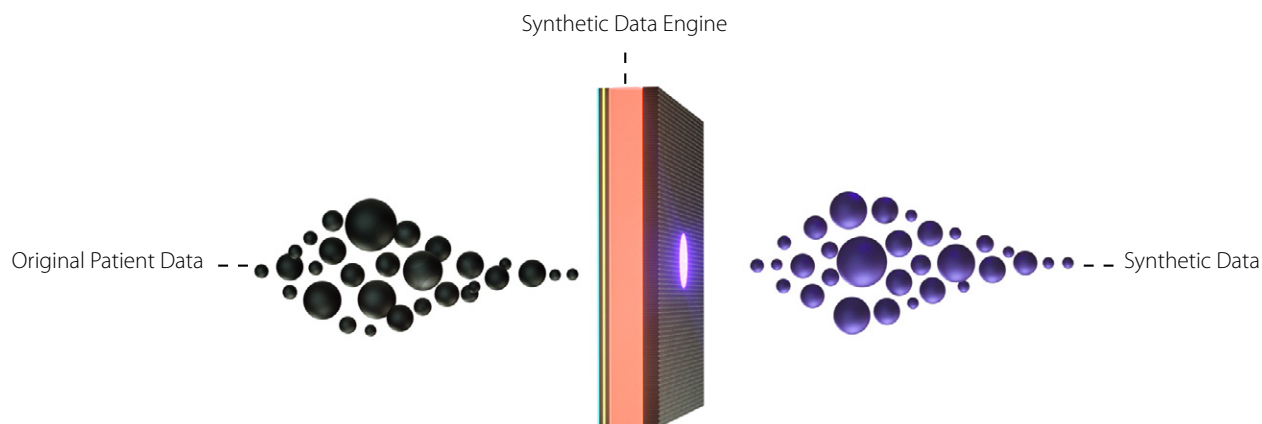# Introducing Synthetic Data

MDCLONE

MDClone Synthetic Data Engine

MDClone's synthetic data overcomes one of the biggest obstacles facing healthcare data analysis: balancing patient privacy and data utilization. Using original data sets, MDClone Synthetic Data Engine creates anonymous data statistically identical to the original but containing no actual patient's information. That way, there's no risk of exposing patient identity and the data can be accessed and analyzed freely.
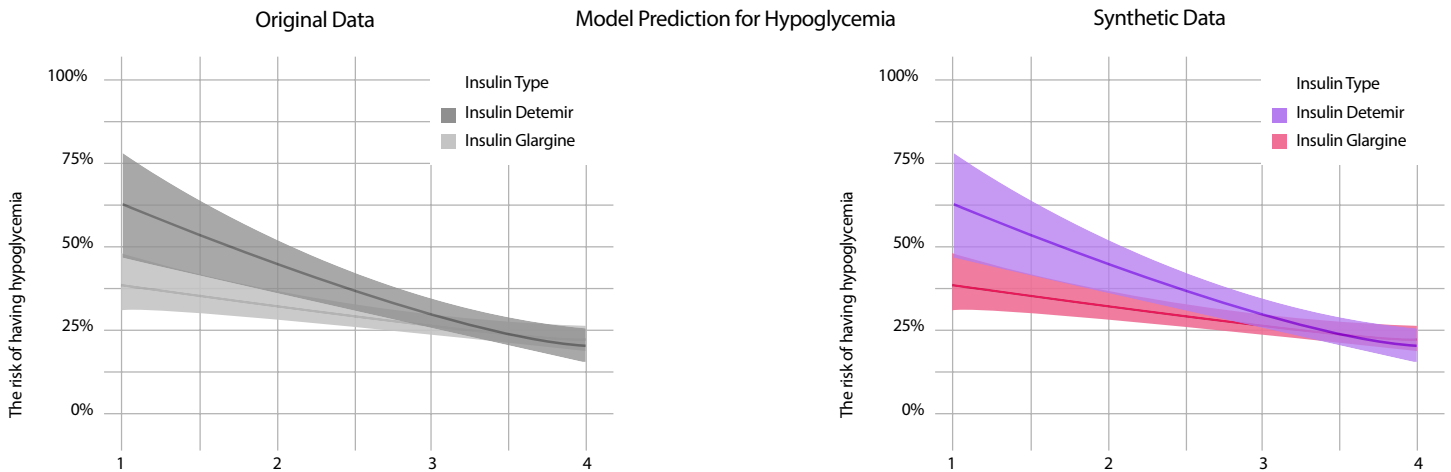
Synthetic Data Engine

Original Patient Data

Synthetic Data

Preserving Complex Patterns

MDClone's Synthetic Data Engine works with any cohort of interest, examining the original data and based on the statistics alone, generating a new data set. Unlike existing methods of de-identification in which real patients are manipulated yet still presented, synthetic data contains no actual patient's information, only the relevant statistical stories.

MDClone's Synthetic Data Engine has been in use at healthcare organisations for almost two years, and it works over and over again, now on hundreds of questions. Here are some examples from one of our partners.
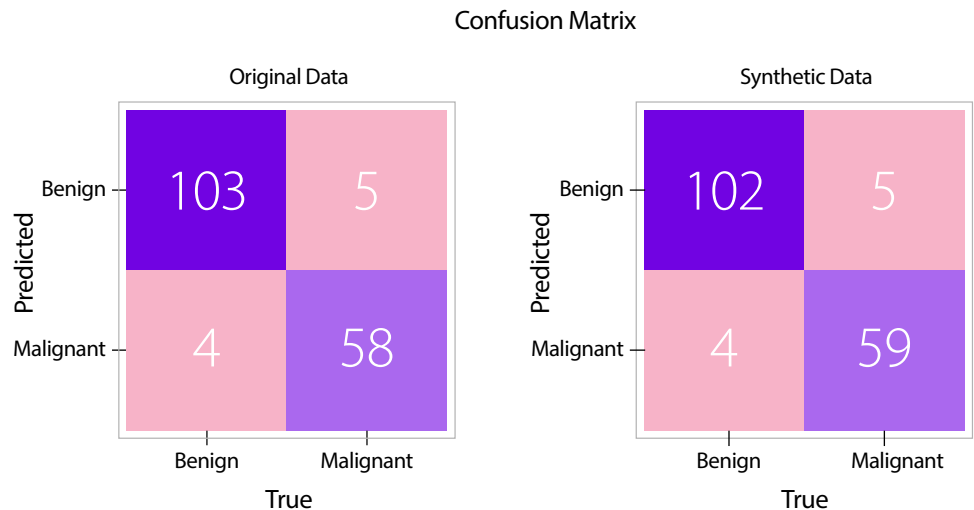
Example 1 **Hyperglycemia**

Comparison prediction based on a stepwise logistic regression model with 11 variables and all their interactions. Models built with original and synthetic data.

Model Prediction for Hypoglycemia

Original Data — Insulin Type: Insulin Detemir, Insulin Glargine

Synthetic Data — Insulin Type: Insulin Detemir, Insulin Glargine

Example 2 **Breast Cancer Prediction**

Random forest machine learning algorithm to predict breast tumor diagnosis from cytopathology data. Synthetic data was proven to be as efficient as the original data, as can be seen by comparing the Confusion Matrix.

Confusion Matrix

Original Data

Predicted \ True	Benign	Malignant
Benign	103	5
Malignant	4	58

Synthetic Data

Predicted \ True	Benign	Malignant
Benign	102	5
Malignant	4	59

About MDClone

MDClone introduces the world's first Healthcare Data Sandbox, unlocking healthcare data to enable limitless exploration, discovery and collaboration.
The Sandbox is a big data platform that eliminates the barriers between data and those who can use it to transform care. Powered by its breakthrough Synthetic Data Engine, the Sandbox overcomes patient privacy restrictions, and together with its robust analytic capabilities data and insights are now at anyone's fingertips.

Get in touch
info@MDClone.com
www.MDClone.com
© 2019 MDCLONE LTD.

MDCLONE