MDCLONE

The Ultimate Guide

Synthetic Data for Healthcare Innovation

eBook

A Summary for Providers, Researchers, Life Sciences, the Private Sector, and Governmental Agencies

Table of Contents

ABOUT THIS EBOOK

This eBook is an in-depth, educational guide to synthetic data, its benefits, and its uses for healthcare and research institutions, life science organizations, and governmental agencies.

OVERVIEW

Synthetic data are artificial data created to maximize healthcare data utility while protecting patient privacy.

Could synthetic data impact operations, quality, and research for provider and academic institutions?

Could synthetic data accelerate research for governmental agencies and life science organizations?

Synthetic data are fast gaining traction among healthcare, life science, and government organizations for their ability to let users securely explore information that mimics the characteristics and correlations of an original population.

With patient privacy preserved, synthetic data allow for reliable data exploration, accelerated discovery processes, and open collaboration between entities.

Key benefits of synthetic data include:

- Synthetic data are not connected with any individual patient, so data can be explored without compromising an individual's privacy.
- Data utility is maximized by eliminating constraints typically associated with exploring sensitive, patient-specific data.
- Data collaboration is facilitated both inside and outside the organization.
- Because synthetic data are not specific to real patients, the institutional review board (IRB) or ethics committee approval process is vastly streamlined, greatly reducing the time-to-insight.

01 What are Synthetic Data?

LET'S DEFINE IT

Synthetic data are non-reversible, artificially created data used for analysis while maintaining the attributes of both discrete and non-discrete variables of interest as well as correlations between variables. The statistical properties and intervariable relationships in a synthetic dataset directly reflect the properties of the source data.

Simply put, synthetic data are novel data created to reflect real data without containing any identifiable patient information.

Because the process is non-reversible and there is no one-to-one correspondence between the synthetic data points and those in the original dataset, synthetic data cannot be traced back to individual patients.

Synthetic derivatives of healthcare data are created and collected from actual patient populations. The synthetic datasets share similar statistical properties with the original data, so they can be analyzed and interpreted as if they were the original data.

DID YOU KNOW?

Synthetic data have the same utility and can be analyzed in the same way as the original, real-world dataset — all the while preserving patient privacy.

02 Types of Anonymized Data

Anonymized data are nothing new. What is new is how those data are created, utilized, and accessed — especially when it comes to synthetic data.

De-identification and anonymization are approaches to transforming data while also taking into account indirect identifiers like gender, race, diagnoses, etc. These terms are sometimes used interchangeably. The process removes the association between a data set and the data subject. Other types of anonymized data include:

DE-IDENTIFIED DATA

In de-identified data, patient privacy is achieved by removing, hiding, or replacing personal identifiers of patients, including those referred to as protected health information (PHI) under HIPAA and other regulatory schemes. These identifiers may include names, dates of births, addresses, ZIP codes, emails, and social security numbers. De-identified data are not foolproof or secure and can compromise utility.

MASKED DATA

Data masking is a process that removes or hides information, usually regarding elements that directly identify a person. For instance, in a masked dataset, a patient's address or age that is visible in the original dataset may have been replaced with the value x. Masked data are at risk to be re-identified or reverse engineered to reveal original patient information, posing security and privacy breaches.

Data Type Comparison

	Synthetic	De-identified
Maintains variables of interest	+	–
Maintains intervariable correlations	+	–
Maintains all data elements	+	✖
Re-identification risk	✖	+
Maximum data utility	+	✖
Maintains complete patient privacy	+	✖

Yes +
Partially –
No ✖

03

Healthcare Data Access

Because important patient privacy principles lie at the forefront of healthcare data utilization, and due to the inherent complexity of healthcare data and the data structures that contain them, accessing this immense amount of data is often tedious, process-laden, and resource-heavy.

Oversight, standards, and regulations work to preserve patient privacy but ultimately negatively impact data utility and access.

A consequence of these necessary and valuable privacy regulations is that large amounts of healthcare data are often difficult to access. Further, when patient-level datasets are available, access generally entails time-consuming committee and regulatory board processes.

30%

of the world’s data volume is generated by the healthcare industry today.[1]

DATA ACCESS CHALLENGES

The medical community knows the immense value of real-world data (RWD) from electronic health records (EHRs), medical claims, laboratory results, reports, and additional documentation. But when access to RWD is difficult and time-consuming, it becomes nearly impossible to harness the full breadth of its utility.

Slow research progress, process inefficiencies, siloed systems, convoluted data models, poor resource utilization, stagnation in innovation, and minimal collaboration are just some of the challenges caused by healthcare data inaccessibility.

Top Data Access Challenges
Slow Research Progress
Process Inefficiences
Siloed Systems
Convoluted Data Models

STREAMLINED DATA ACCESS

Accelerating access to RWD can empower research organizations to:

+ Bridge the efficacy-effectiveness gap
+ Population health management
+ Speed up drug research and development
+ Conduct safety and pharmacovigilance audits
+ Expedite translational research
+ Develop treatments that improve patient outcomes

THE GOAL

Scientists, clinicians, and researchers can use synthetic data to efficiently pursue innovation and improve patient care without hindrance.

The goal of generating synthetic data is to give stakeholders fast access to privacy-preserving, fictional but functional datasets that capture the complexities of the original patient-level data.

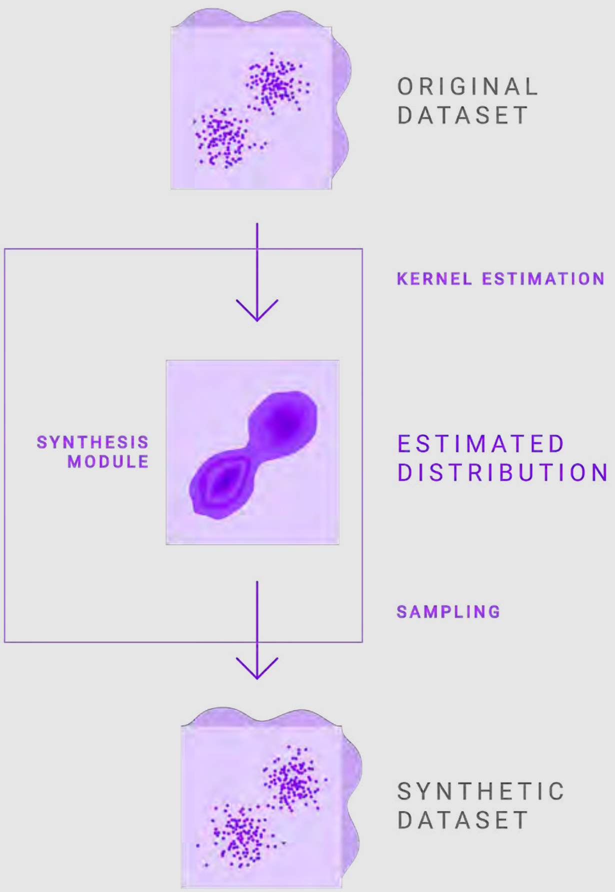
04

Synthetic Data Creation

Synthetic healthcare data are derived from original patient data collected from actual patient populations.

This patient-level data refers to data about patients' health statuses and healthcare delivery routinely collected from sources such as EHRs, claims and billing activities, and product and disease registries.

Data can be extracted from disparate systems and warehouses into one data lake, pooling information together to create a complete patient portrait. From there, synthetic data are created for exploration and discovery.



Synthetic data can be generated in a variety of ways, from computational derivation to deep learning models such as variational autoencoders (VAEs) and generative adversarial networks (GANs).

Computational Derivation and Bayesian Networks

COMPUTATIONAL DERIVATION

Computational derivation is one way in which synthetic data are created.

A cohort of interest is selected from a larger population set, along with the properties of interest in that cohort. Computational algorithms then assess the statistical characteristics of the defined sub-population — distributions, correlations, and intervariable relationships. Finally, the synthetic data engine populates the computed model with novel data points to create a dataset with the same statistical properties and relationships as the original cohort but containing none of the original data points.

BAYESIAN NETWORKS

Bayesian networks — probabilistic graphical models that represent knowledge about variables and their dependencies — can be used to simulate new synthetic data. These networks can be used to make inferences about particular variables of interest in the network, depending on the other variables in the network. However, to simulate synthetic data with Bayesian networks, a user must have existing information about the dataset and must also have a great deal of computational power to compute massive and sparse datasets.

Deep Learning Models: VAEs and GANs

Deep learning models can also be used to generate synthetic data, with the main two techniques being in the neural network class of deep learning systems. Neural networks mimic the way the human brain functions and can be trained to generate patterns through prediction and correction.

VAES

The first type of neural network that is commonly used to create synthetic data is called a variational autoencoder (VAE). Useful for working with data that are imbalanced, VAEs doubly transform a data distribution to minimize errors when reconstructing a dataset. In essence, they are trained to reduce the number of dimensions of the data while losing only a minimal quantity of information. However, they have a high degree of complexity and require a strong theoretical understanding of their functioning.

GANS

A more common deep learning method used to generate synthetic data, especially for image and video generation, is the use of generative adversarial networks (GANs). GANs train two neural networks, respectively known as a generator and a discriminator. The generator creates synthetic data, and the discriminator tries to determine whether the result is real or not. This operates in a continuous loop, and the quality of data being produced by the generator keeps improving. Eventually, the discriminator becomes unable to tell the difference between the real and the synthetic data. In healthcare, GANs can be used to generate synthetic images of skin lesions, for example, as well as X-rays and pathology slides.

05 Value of Synthetic Data

When clinical leaders, health systems, drug developers, and researchers can access, explore, and analyze healthcare data without obstacles, delays, or worry, the potential benefits are immeasurable.

Users working with synthetic data can glean information and draw conclusions they would otherwise not have been able to reach due to privacy boundaries. Synthetic data, which can be shared with other users without violating patient privacy, can facilitate research, aid in the discovery of health patterns, and assist clinicians in treating patients.

Synthetic Data in Healthcare Enables:

- More Efficient Processes and Resource Utilization
- Increased Collaboration Between Stakeholders
- Faster Research
- Improved Patient Outcomes
- Patient Privacy Preservation
- Aggregation of Data from Disparate Sources
- Increased Innovation

More Efficient Processes and Resource Utilization

Synthetic data enable healthcare systems and institutions to develop more efficient processes and care workflows, as well as improve financial and labor resource allocation and utilization.

With synthetic data in hand, non-data analysts can access the data they need to see and understand performance trends for service lines, departments, facilities, and assets. Achieving resource efficiency, identifying care gaps, developing improved care pathways, and avoiding high-cost complications by identifying at-risk populations early are just some of the benefits of having streamlined access to this data.

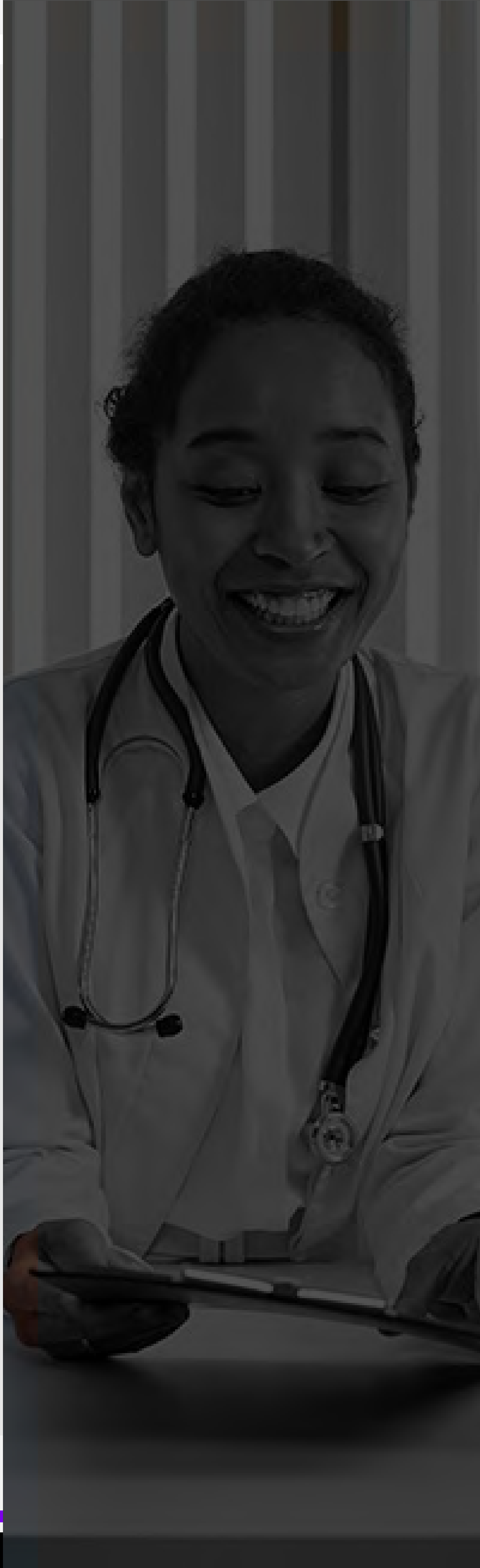
Increased Collaboration Between Stakeholders

Sharing healthcare data can be laborious, expensive, and fraught with obstacles. Between balancing patient privacy and adhering to compliance requirements, individuals, teams, and organizations often find it difficult to share data and work together internally and externally.

Synthetic data opens the door to secure, privacy-guaranteed data-sharing opportunities. With it, organizations can let members of their internal teams explore information and share research with third-party external teams worldwide without undue constraints.

Benefits of increased collaboration between stakeholders in healthcare include[3]:

- Probability of less bias
- Reduced knowledge gaps
- Improved research validity
- Improved efficiency
- Shorter timelines
- Speed-to-insight



BUILT-IN PRIVACY PRESERVATION

Synthetic data offer built-in patient privacy preservation because there is no one-to-one correspondence between patients in the synthetic population and those in the original population.

Faster Research

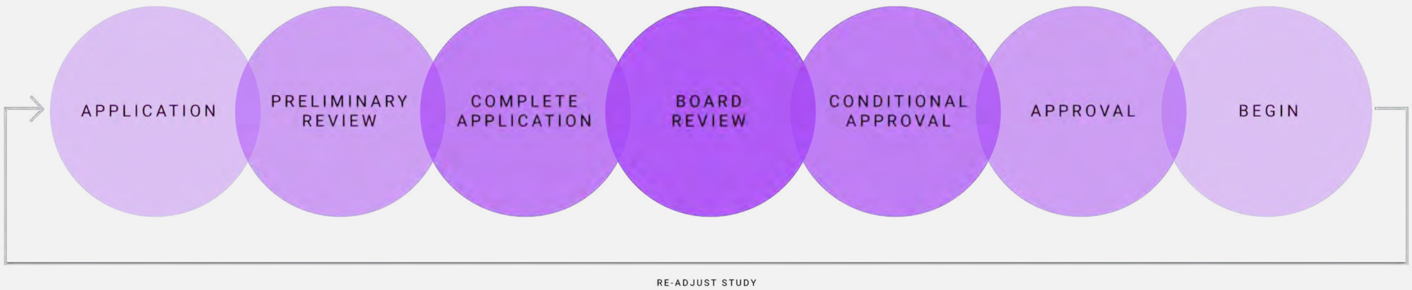
The use of synthetic data allows researchers to rapidly discover and prototype heavily protected health information. It empowers researchers to quickly organize and access data, explore ideas, and find insights that power research, leading to better patient outcomes and impactful healthcare innovation.

By offering maximum data utility while preserving patient privacy, access to synthetic data makes more information available to more people with fewer constraints, which shortens the timeline of discovery from years to days or even hours.

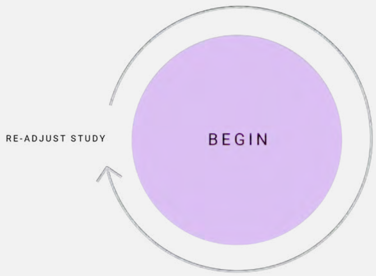
10x

more efficient research processes with synthetic data[4]

RESEARCH PROCESS
WITHOUT SYNTHETIC DATA



A major academic medical center conducted faster research with synthetic data:



RESEARCH PROCESS
WITH SYNTHETIC DATA

During the peak of the coronavirus pandemic, rapid research was crucial to saving lives and understanding the virus and its impact. Using synthetic data, an academic medical center was able to identify trends across its 15 hospital locations.

These trends revealed that in north St. Louis communities, Black patients with COVID-19 were 2.5 times more likely to be admitted to the hospital and, once admitted, four times more likely to end up in intensive care than any other patient group.[5] With this information in hand, the healthcare organization was able to work with public health groups to help support its higher-risk populations.

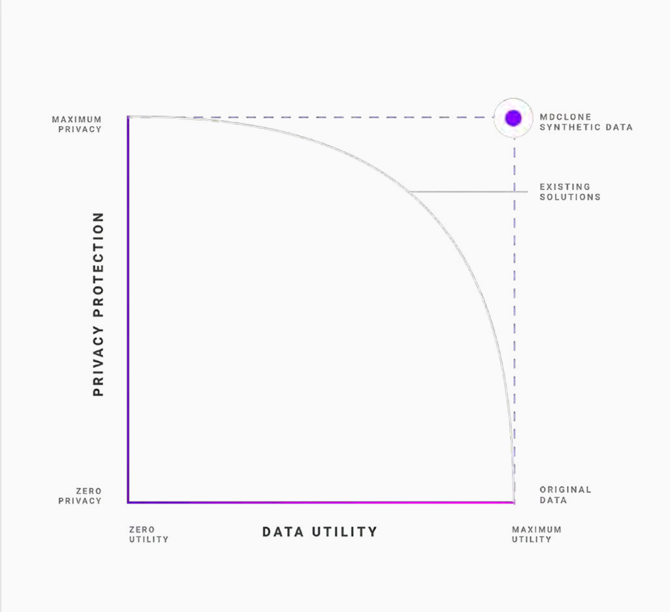
By choosing to work with synthetic data, they were able to identify the most at-risk patient populations during the beginning of the COVID-19 pandemic, when speed-to-insight was critical to responding to the virus.

Improved Patient Outcomes

When clinician researchers leverage their health systems' historical data, they can realize huge payoffs in terms of timely diagnosis, increased diagnostic accuracy, disease progression expertise, and personalized care pathways. They are empowered to explore patterns they might have noticed, validate their hypotheses with synthetic data, and optimize care delivery as quickly as possible. Patients can reap benefits such as faster intervention, highly personalized treatment, reduced hospitalizations, improved predictive analysis, and reduced morbidity and mortality.

Patient Privacy Preservation

Synthetic data offer built-in patient privacy preservation because there is no one-to-one correspondence between patients in the synthetic population and those in the original population from which it was derived. Synthetic data therefore retain the utility of the original data they are derived from while maintaining patient privacy. When this privacy is guaranteed, researchers, clinicians, and leaders in the industry can explore and make data-driven decisions with ease and speed.



Aggregation of Data from Disparate Sources

The healthcare industry is faced with data aggregation challenges. It is often difficult for health information from diverse information systems to be gathered and shared across various sites at which clinical care, research, and innovation might be taking place. This incompatibility makes it difficult for researchers to aggregate large volumes of data from different sources to test analytical models, develop health IT tools, and other purposes.[6]

When information can be pulled from disparate sources and external systems, pooled into one data lake, and from there, extracted for research and analysis, users can engage with all data elements, including clinical and non-clinical patient-oriented information as well as structured and non-structured information. From there, synthetic data enables maximum data utility by any user in a health system.

Rapid extraction of data from EHRs allows for a deep understanding of patient medical histories:

Israel's largest hospital created synthetic data from its COVID-19 patients and pooled information from its EHRs with those of a large ambulatory health network to better understand its patients' medical histories. Synthetic data enabled clinicians to find links between pre-existing conditions and COVID-19 outcomes.[7]

> Synthetic data enabled clinicians to find links between pre-existing conditions and COVID-19 outcomes.[7]

Increased Innovation

Synthetic data accelerates all forms of discovery — whether that be drug discovery or uncovering gaps in care. Because synthetic data derivatives are not considered to be that of human subjects, studies that use synthetic data allow researchers and clinicians to access data without the typical IRB process. They can test theories, analyze models, share data, and collaborate effectively to instantly accelerate innovation.

Valuable insights into patients' histories with synthetic data:

The use of synthetic data has given clinicians in Israel's largest hospital the ability to work with that data and gain insights after a few hours of training. The speed to research and discoveries through synthetic data leads to better operations and efficiency, improved patient safety, and better patient outcomes.[8]

06 Validity of Synthetic Data

Synthetic data provides statistically similar results to real-world data.[9]

A groundbreaking study analyzing medical research results based on synthetic data and their relation to real data results concluded that the "use of synthetic structured data provides a close estimate to real data results and is thus a powerful tool in shaping research hypotheses and accessing estimated analyses."[10]

In another study comprising three diverse use cases, the validity of synthetic data was tested by comparing results with the original data. In each case, results from synthetic data were found to be statistically similar to results from the original data, further validating the utility of synthetic data for research, leading to faster insights and improved data sharing.[9]

These studies demonstrate that synthetic data is a valid tool for research and discovery. The use of synthetic data ultimately gives clinicians and researchers new opportunities to study patient populations that are statistically similar to an original patient population without compromising patient privacy.

VALIDATING SYNTHETIC DATA FOR RESEARCH

Results were sufficiently statistically similar **($P > 0.05$)** between the synthetic derivative and the real data to draw the same conclusions.[9]

READ THE RESEARCH PAPER ↗

Proven with Research

Publications supporting the accuracy of synthetic data when compared to real data results:

EXPLORE RESEARCH PUBLICATIONS ↗

RESEARCH PAPERS

Foraker, Randi E., Sean C. Yu., Aditi Gupta, et al. **Spot the Difference: Comparing Results of Analyses from Real Patient Data and Synthetic Derivatives.** *JAMIA Open* 3, no. 4 (2020): 557–566.

doi:10.1093/jamiaopen/ooaa060. ↗

Guo A, Foraker RE, MacGregor RM, Masood FM, Cupps BP, Pasque MK. **The use of synthetic electronic health record data and deep learning to improve timing of high-risk heart failure surgical intervention by predicting proximity to catastrophic decompensation**. Frontiers Digital Health 2020.

doi:10.3389/fdgth.2020.576945 ↗

Zamstein N, Neumann AJ, Foraker RE et al. **Prediction of COVID-19 case severity using synthetic data derived from the National COVID Cohort Collaborative.** Poster presented at: AMIA Annual Symposium; October 30–November 3, 2021; San Diego, CA. Poster Session 2, Board Number 117.

Summary of presentation ↗

07

Synthetic Data Applications

The various applications of synthetic data are extensive and can include:

- Academic Research
- Healthcare Organizations
- Federal Agencies
- Life Sciences

Academic research institutions, healthcare organizations, federal agencies, and life sciences organizations can use synthetic data to speed hypothesis testing, increase the pace of research, improve operational efficiency and care quality, and facilitate third-party access to data, while enabling collaboration within and between organizations.

Obstacles to data sharing — primarily privacy concerns — are vastly reduced with synthetic data, allowing organizations to make data-driven decisions for the populations they care for.

OVERCOMING OBSTACLES

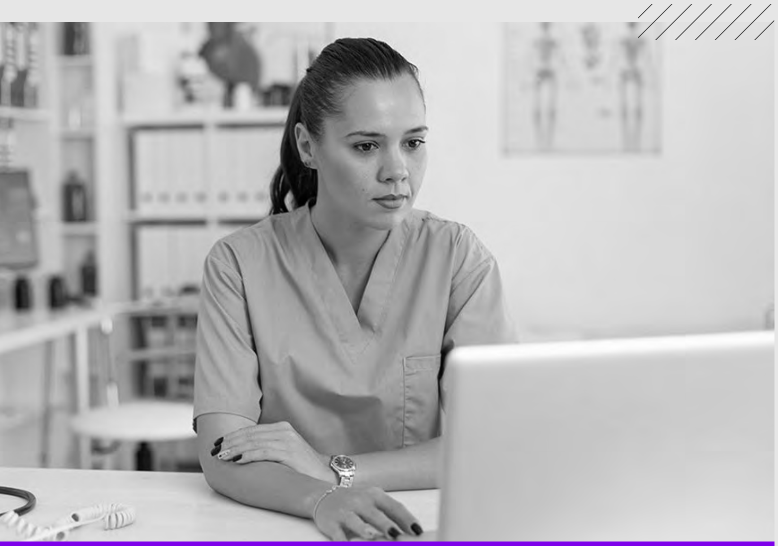
Obstacles to data sharing — primarily privacy concerns — are vastly reduced with synthetic data, allowing organizations to make data-driven decisions for the populations they care for.

For Academic Research

To accelerate the pace of data-driven research, academic medical centers can give investigators, students, and research professionals at their institutions the ability to interact with data in real time.

> Synthetic data allows users to deploy research projects faster and independently.

Synthetic data may provide a solution for researchers who aim to generate and share data in support of precision healthcare. Data synthesis enables the creation and analysis of synthetic derivatives as if they were the original data, providing — for this purpose — a superior alternative to data de-identification.



Here's how a major research university can use synthetic data for research:

A major research university aims to accelerate the pace of data-driven research via the use of synthetic data. Specifically, it ensures that every investigator at its university can interact with data instantaneously to advance research and discovery.

POSSIBLE OUTCOMES:

+ Interact with several diagnoses and procedures and view the size of a cohort population in seconds
+ Adjust queries dynamically and check project feasibility before submitting for original data access
+ Improve timing of high-risk surgical intervention using supervised deep learning and machine learning algorithms
+ Identify of patients to predict mortality within specific time frames from the first diagnosis by machine learning and deep learning models

For Healthcare Organizations

Major provider and health system organizations can use synthetic data to improve management protocols and outcomes for their patients.

With faster access to data, synthetic data allows organizations to access broader sets of data, refine hypotheses, streamline knowledge extraction and discovery, and deliver better care.

Here's how a major health can use synthetic data to improve management protocols and outcomes for its patients:

A health system aims to build a program for managing and improving outcomes in a specific population. More specifically, the program needs to identify patients in early stage diagnosis and deliver fully integrated and coordinated care.

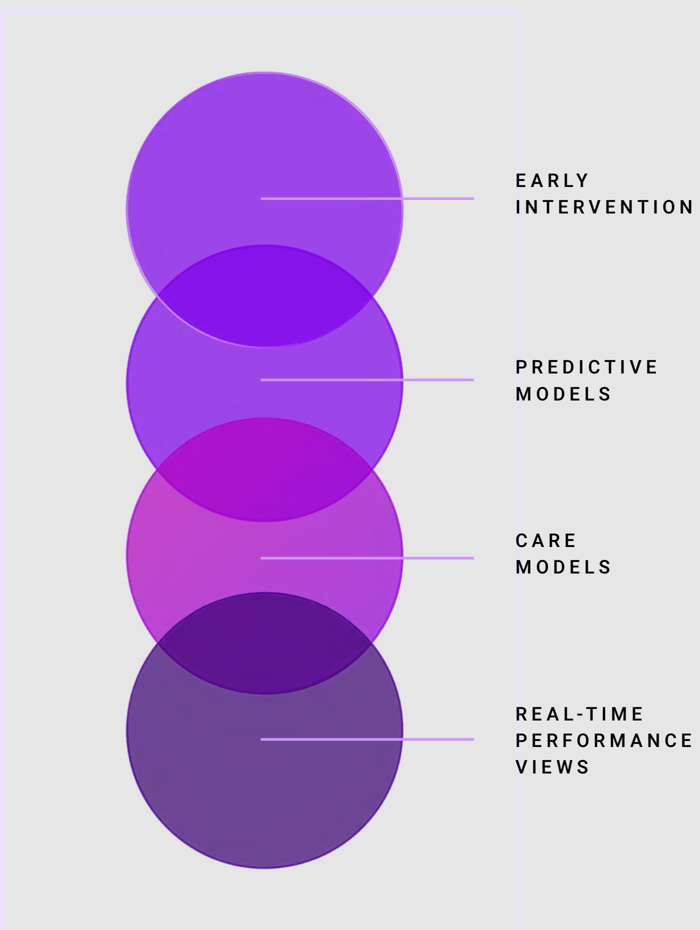
Because late diagnosis and the presence of two or more comorbidities exacerbate the cost and complexities of certain conditions, the health system needs customized care interventions to manage the conditions effectively.

POSSIBLE OUTCOMES:

+ Deploy algorithms that accurately identify patients by stage and presence of specific conditions
+ Stratify patients by gaps in care and time from the last encounter
+ Identify cases of comorbidity clusters, lab trajectories, CPT codes, ICD-10 codes, medical claims data, and other operational constructs
+ Increase predictive analytics capabilities

For Governmental Agencies

Governmental agencies aim to understand health risk factors and improve care delivery and outcomes for their populations. They also want to drive operational improvements and efficiency.



Here's how a major governmental agency can use synthetic data to better understand and improve the health of its population:

A major governmental healthcare system aims to use synthetic data to empower a wider network of clinical and operational staff to explore data and discover insights that can be used to impact lives.

Potential projects could include chronic disease management, precision medicine, health equity, and COVID-19.

POSSIBLE OUTCOMES:

+ Create predictive models
+ Test and refine care models and initiatives for its population
+ Obtain real-time views of the performance of its facilities and service lines
+ Identify leading indicators of a disease to enable early intervention for patients most at risk

For Life Sciences

Given the enormous financial investment and exceedingly long time frame involved in launching new, life-saving therapies, pharmaceutical companies lose revenue and competitive advantage when timelines stretch out.

Using synthetic data, life sciences companies might[12]:

+ Understand how a particular drug product impacts a disease not included in the original drug indications
+ Secure previously inaccessible datasets, such as high-quality, real-world data siloed in health systems' EHRs, claims databases, and lab information systems
+ Significantly expedite site selection and subject recruitment
+ Share data and collaborate efficiently with health systems, payers, and related institutions to solve bigger problems faster

Because synthetic data can provide faster, easier access to a broad range of real-world datasets, companies can also streamline post-marketing and outcomes research projects as well as utilize resources more efficiently to better support the commercialization of their drug portfolios.

Here's how a life sciences company can use synthetic data to study the impacts of drug products on particular diseases:

A life sciences company aims to understand how a particular drug product impacts a disease not included in the original drug indications.

POSSIBLE OUTCOMES:

+ Perform studies to understand the drug product's impact (efficacy) on this new disease
+ Determine the value and performance of the drug product on the disease
+ Complete a full analysis in less than a month, with less effort expended than if they were to have conducted a study without synthetic data
+ Life sciences companies can use synthetic data to propel their businesses forward and better understand drug products and diseases, share data, and achieve a greater return on investment

08

Synthetic Data in AI and Machine Learning Systems

Artificial intelligence is increasing in sophistication and popularity, making the value of its applications across multiple industries immeasurable. AI is now used in almost every sector — from healthcare to finance to e-commerce.

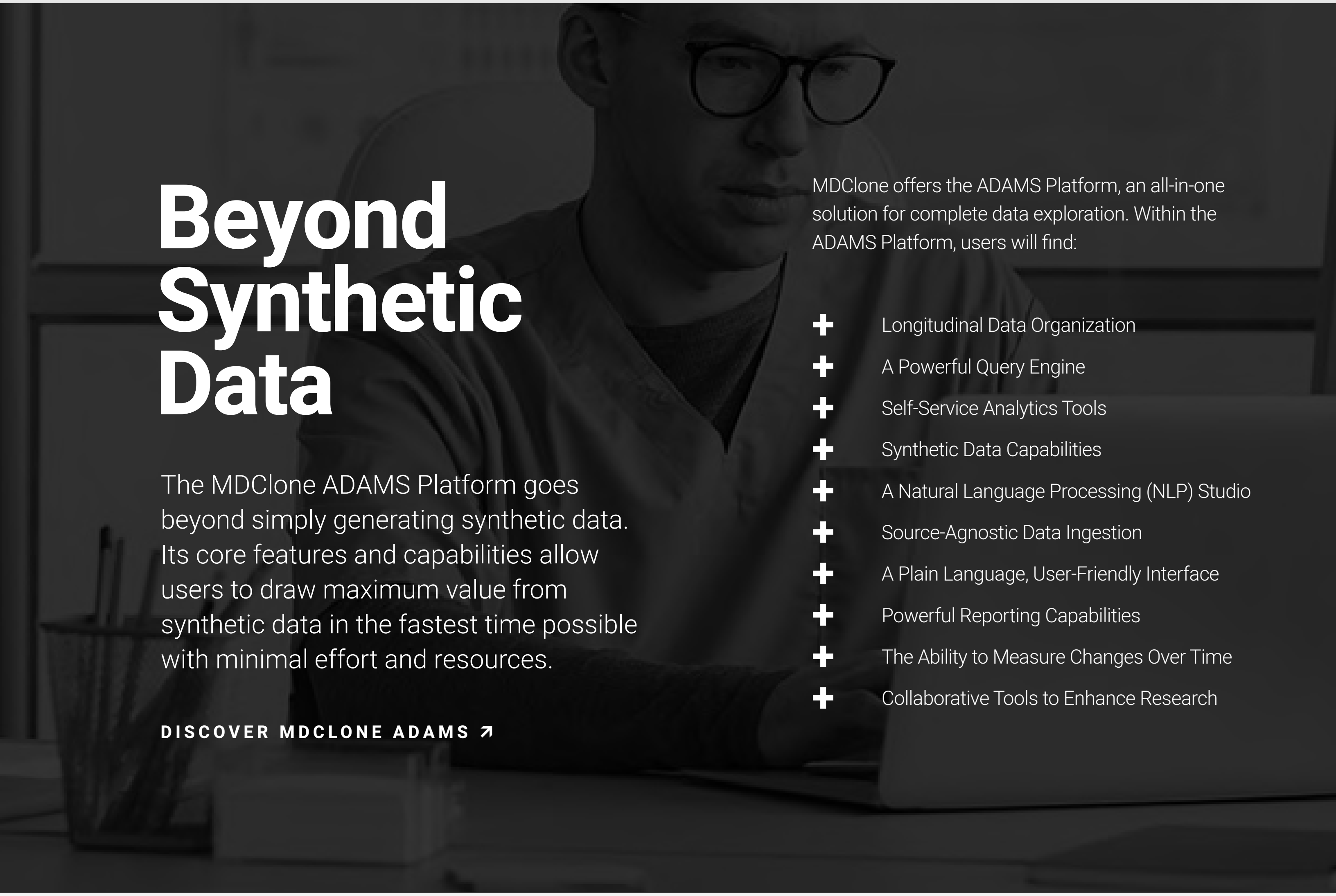
Machine learning algorithms — computational instructions that power AI systems — are trained with large amounts of data. Generally, the more data available, the better the performance of a machine learning model.

However, gaining access to large amounts of data for training these AI systems is challenging. With synthetic data, researchers can build the repositories of data they need to train their algorithms.

Machine learning models trained on synthetic data retain their validity when applied to the original population from which the synthetic training set was derived. In other words, AI experts can use easily accessible synthetic data to develop their machine learning engines, which can then be used to make valid predictions and conclusions about real-world patients.[9,11]

In essence, synthetic data can help AI technology maintain compliance and data protection methods to avoid the use of personal data within AI models.

09 The MDClone Difference



Beyond Synthetic Data

The MDClone ADAMS Platform goes beyond simply generating synthetic data. Its core features and capabilities allow users to draw maximum value from synthetic data in the fastest time possible with minimal effort and resources.

DISCOVER MDCLONE ADAMS ↗

MDClone offers the ADAMS Platform, an all-in-one solution for complete data exploration. Within the ADAMS Platform, users will find:

- Longitudinal Data Organization
- A Powerful Query Engine
- Self-Service Analytics Tools
- Synthetic Data Capabilities
- A Natural Language Processing (NLP) Studio
- Source-Agnostic Data Ingestion
- A Plain Language, User-Friendly Interface
- Powerful Reporting Capabilities
- The Ability to Measure Changes Over Time
- Collaborative Tools to Enhance Research

SELF-SERVICE INTERACTION

Generating and using synthetic data within the MDClone ADAMS Platform is fast and empowers users in an organization to access data on their own terms. By definition, self-service means that anyone can interact in real time with data without technical mediators, external teams, or data experts.

With the MDClone ADAMS Platform and its synthetic data capabilities, users can quickly organize and access information, explore ideas, and generate insights instantly and in their own time. There is no need to wait for external teams to extract data or technical experts to navigate the data. Instant, self-service access makes possible an interactive, iterative discovery process to unlock the knowledge hidden in raw data.

By definition, self-service means that anyone can interact in real time with data without technical mediators, external teams, or data experts.

DYNAMIC DATA EXPLORATION

The MDClone ADAMS Platform offers a dynamic and fluid process for data exploration, analysis, and action by any user. With MDClone's synthetic data engine, synthetic data can be created or accessed by querying original data. The user asks the platform a "question" by defining the sub-population they want to explore and the particular population characteristics they are interested in. The platform answers the "question" by creating a new dataset that encapsulates the properties and characteristics the user cares about.

Research moves faster, and innovation excels when there are no constraints on how quality data can be explored.

PICTURED:
A user-friendly interface featuring the ADAMS Platform Query Tool

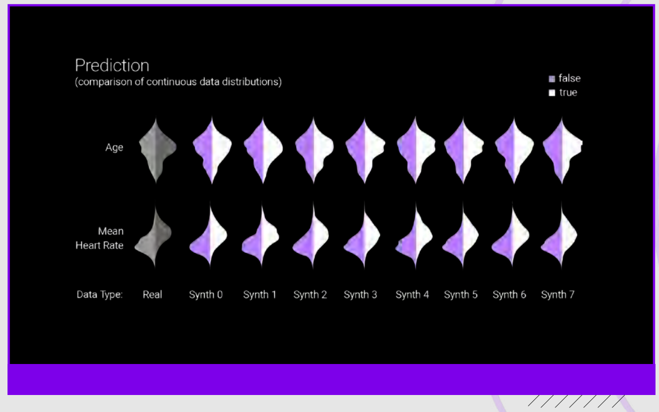
With MDClone, users can:

- Access and test ideas, instantly creating models
- Build entire studies and get insights in their own time

INSTANT CONVERSION OF DATA

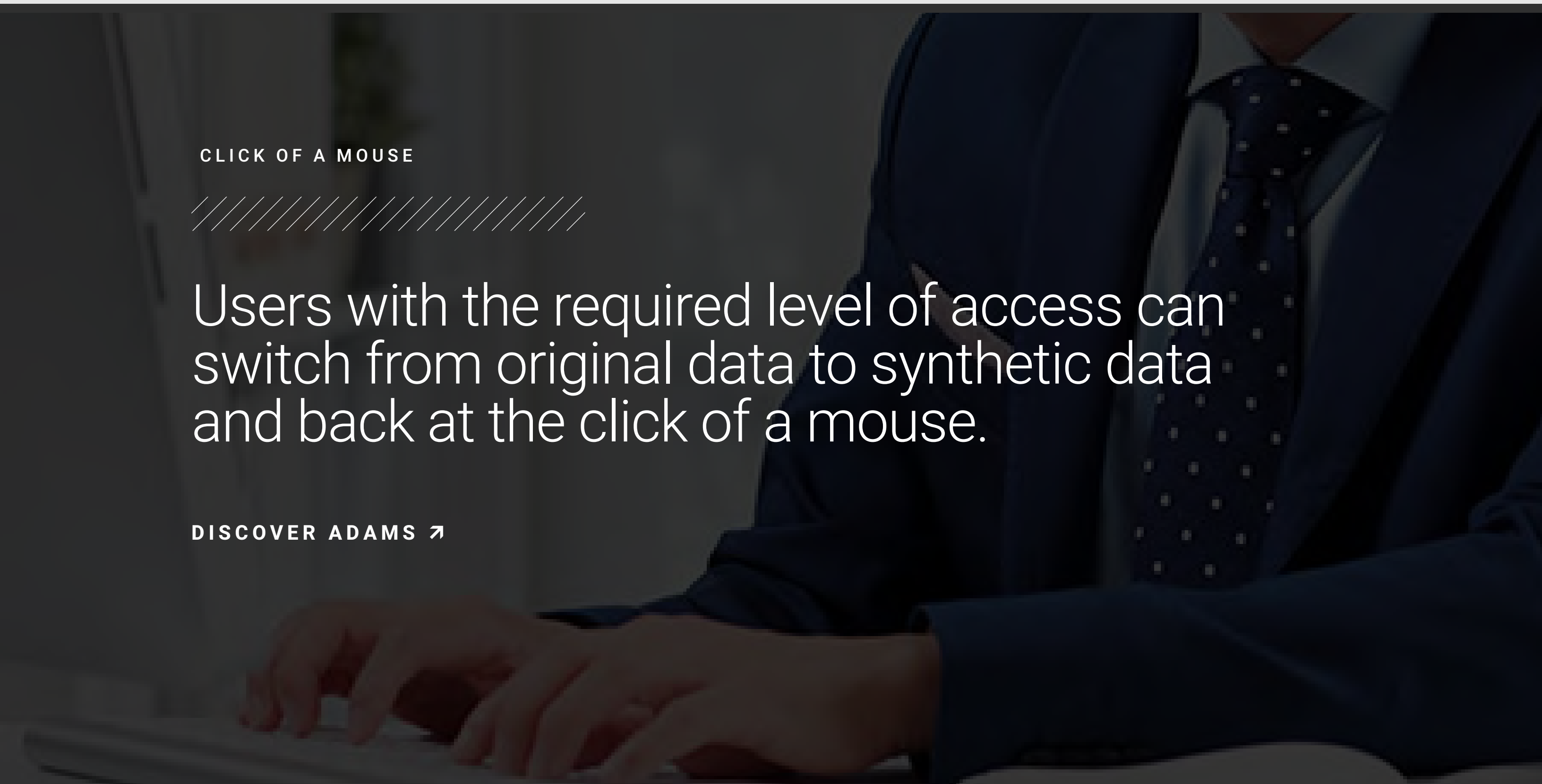
MDClone lets users with the required level of access switch from original data to synthetic data and back at the click of a mouse.

After data has been explored and the project scope, hypotheses, and theories have been reviewed using synthetic data, for example, users can obtain IRB approval for the original dataset to move the project forward. The queries and visualizations already developed for exploring the synthetic data can be applied instantly to the original data without repeating any work.



> MDClone really, really makes data to access easy. It saves months of work because it enables a quick search into patients' data that can give a researcher a sense of the sample size, which allows an easy estimation of the study aim.

—Dr. Nir Horesh, Senior General Surgeon, Sheba Medical Center



CLICK OF A MOUSE

Users with the required level of access can switch from original data to synthetic data and back at the click of a mouse.

DISCOVER ADAMS ↗

A GAME-CHANGER

> "The beauty of synthetic data is that it allows us to quickly create datasets that look and feel just like the real data that are generated every time we interact with patients. Synthetic data is really a game-changer."

–Philip R. O. Payne, PhD, FACMI, FAMIA, FAIMBE, FIAHSI, Founding Director of the Institute for Informatics (I2) at Washington University in St. Louis

ABOUT THE MDCLONE ADAMS PLATFORM

The MDClone ADAMS Platform is a self-service data analytics environment that empowers users to quickly organize and access information, accelerate research, drive better patient outcomes, empower healthcare teams to action, and create impactful healthcare innovation.

Put your organization on a continuous learning path with real-time access to insights with our pioneering healthcare data platform that breaks down barriers in data exploration.

ABOUT THE GLOBAL NETWORK

The Global Network is a member-driven collaborative, empowering the largest and most innovative healthcare organizations to work together to perform research and improve operational and patient outcomes faster. Member sites can invite collaborators to explore synthetic data within a secure, controlled, collaborative environment.

The Global Network currently has more than 20 member sites around the world. Powered by MDClone, the network connects healthcare organizations to life sciences to facilitate partnership, research, and healthcare advancement in an unparalleled networking model.

> It's clear to me that synthetic data is the key to moving population health forward. The more I work with MDClone, the more convinced I am that it's a game-changer for all of us.

—Dr. Alan Forster, VP, Innovation and Quality at the Ottawa Hospital; Vice Chair, Quality and Clinical Services, Department of Medicine, Faculty of Medicine, University of Ottawa

Partner With Us

MDClone is a technology firm focused on unlocking healthcare data and empowering exploration, discovery, and collaboration.

MDClone democratizes data and empowers clinicians, researchers, and executives to explore, drive action, and work together to improve patients' health.

LEARN MORE ↗

REFERENCES

1. "The Healthcare Data Explosion." *RBC Capital Markets.* https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion.

2. MDClone. "Intermountain Healthcare Drives Remarkable Results with MDClone." https://www.mdclone.com/intermountain-healthcare.

3. Nass, Sharyl J., Erin Balogh, and Margie Patlak. 2010. *Extending the Spectrum of Precompetitive Collaboration in Oncology Research: Workshop Summary.* Washington, DC: National Academies Press.

4. Davis, Jessica. "How Synthetic Data Accelerates Coronavirus Research." *Big Data Analytics* (blog). *InformationWeek*, August 7, 2020. https://www.informationweek.com/big-data-analytics/how-synthetic-data-accelerates-coronavirus-research.

5. MDClone. "Case Study: Intermountain Healthcare." https://www.mdclone.com/case-study-Intermountain-healthcare.

6. The Office of the National Coordinator for Health Information Technology. 2021. *Synthetic Health Data Challenge Frequently Asked Questions*. https://www.healthit.gov/sites/default/files/page/2021-10/Synthetic%20Health%20Data%20Project%20FAQs_508.pdf

7. Lieber, Dov. "The People in This Medical Research are Fake. The Innovations are Real." *The Wall Street Journal*, April 6, 2021. https://www.wsj.com/articles/the-people-in-this-medical-research-are-fake-the-innovations-are-real-11617717623.

8. Sheba Tel HaShomer City of Health. 2022. *Sheba Medical Center: The Smart Hospital Leading the World.* https://issuu.com/bizclikmedia/docs/bro-sheba-medical-center-march2022.

9. Foraker, Randi E., Sean C. Yu., Aditi Gupta, et al. "Spot the Difference: Comparing Results of Analyses from Real Patient Data and Synthetic Derivatives." *JAMIA Open* 3, no. 4 (2020): 557–566. doi:10.1093/jamiaopen/ooaa060.

10. Benaim, Anat Reiner, Ronit Almog, Yuri Gorelik, et al. "Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison from Five Observational Studies." *JMIR Medical Informatics* 8, no. 2 (2020). doi:10.2196/16492.

11. Guo, Aixia, Randi E. Foraker, Robert M. MacGregor, Faraz M. Masood, Brian P. Cupps, and Michael K. Pasque. "The Use of Synthetic Electronic Health Record Data and Deep Learning to Improve Timing of High-Risk Heart Failure Surgical Intervention by Predicting Proximity to Catastrophic Decompensation." *Frontiers in Digital Health* 2 (2020). doi:10.3389/fdgth.2020.576945.

12. Stone, Andrew. "Synthetic Data: Pharma's Next Big Thing?" *Reuters*, March 22, 2022. https://www.reutersevents.com/pharma/clinical/synthetic-data-pharmas-next-big-thing.

MDCLONE

REQUEST A DEMO ↗

MDCLONE.COM

Unlock healthcare data.
Transform care.