



Pragmatism vs. Idealism: Synthetic Data in Market Research

Augmenting Primary Research with
Synthetic Approaches





The insights industry feels as though it's on the cusp of a data revolution. With the emergence of new methods to generate synthetic data, the potential to expand, enhance, and even pioneer new forms of market research is palpable.

That said, the concept of synthetic data sparks debate. Some are eager to "revolutionize consumer insights," while others see it as merely incremental, or even, resist the idea altogether.



Four Types of Synthetic Data

There are four primary types of synthetic data that could be used to augment, or in some cases to replace, primary market research.



Persona-Based Type



Description

Fictional profiles representing typical individuals within a target audience, based on real data patterns but entirely artificial

Common Uses

Hypothesis development, storytelling, internal training

Training Needs

Limited to aggregate level reports

Example

A synthetic patient profile illustrating medication adherence patterns in diabetic patients, useful for hypothesis development and internal training for pharmaceutical sales teams

Boosted Samples Type



Description

Synthetic respondents added to real datasets to fill gaps or balance representation, based on desired distributions

Common Uses

Balancing samples, subgroup analysis

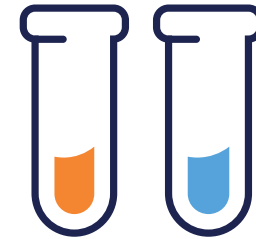
Training Needs

Individual-level data from a quantitative dataset

Example

Adding synthetic respondents to balance representation for hard-to-reach patient sub-groups to help pharmaceutical companies understand unmet needs across all types of patients

Pure Silicon Type



Description

Fully synthetic datasets created without any real respondent data but designed to mirror statistical characteristics of a population

Common Uses

Model training, concept testing, internal research without privacy risks

Training Needs

Limited

Example

A synthetic dataset replicating survey data from patients with rare diseases, designed for internal research on potential treatment plans without compromising patient privacy

Digital Twins Type



Description

Synthetic models of individual respondents, designed to simulate behaviors and preferences over time

Common Uses

Behavioral forecasting, concept testing, market simulation

Training Needs

Individual-level data

Example

Developing a digital twin model of patients with chronic heart disease to simulate reactions to new treatment regimens, aiding in forecasting treatment outcomes and market demand for new medications

Where Synthetic Data Can Add Value



Hypothesis Generation

Synthetic datasets can be invaluable for generating hypotheses about evaluating receptivity to new drug formulations or new devices.



Early-Stage Testing

Synthetic data can efficiently test initial concepts without the need for extensive primary data collection.



Augmenting Limited Datasets

When patient cohort data are limited due to privacy regulations or ethical considerations, synthetic data can fill these gaps, broadening the context for epidemiological studies.



Reducing Respondent Burden

In studies where gathering real patient data is intrusive or ethically complicated, such as in mental health, synthetic data helps reduce the demand on actual patients while still providing comprehensive insights.

Addressing the Challenges Head-on

Capturing Real-World Complexity

Synthetic data may not naturally reflect the unpredictability of real-world behaviors. However, well-constructed models can integrate clinical, demographic, and lifestyle variables, better simulating patient behaviors and treatment outcomes.

Historical Confirmation Bias

Models need periodic updates to reflect current medical advancements and data from the latest medical research and treatment protocols to mitigate confirmation bias.

Real-Time Adaptability

Modern synthetic data models can incorporate timely inputs, such as electronic health records (EHRs) and wearable health devices, making them responsive to immediate patient needs and emerging health trends.

Emerging Trends and Novelty

This aligns with the broader debate on AI's capacity for creativity. While synthetic data may never fully capture the "pulse of the market," or predict novel disease outbreaks, it can be used to simulate various health scenarios, helping pharmaceutical companies and healthcare providers prepare for unexpected developments it can complement human insight rather than replace it.

A Balanced Viewpoint

Synthetic data isn't a silver bullet, and it shouldn't be treated as one. Its greatest value lies in its ability to complement primary research, not replace it. In the right circumstances, synthetic data can help researchers move faster, fill gaps, and test hypotheses efficiently.

As the industry continues to evolve, the judicious use of synthetic data offers a pragmatic path forward, allowing market researchers to better navigate complexity and make informed, confident decisions.



KJT is an evidence-based healthcare consulting firm. We transform and empower businesses through evidence, insight, and collaboration. Founded in 2007, KJT employs more than 50 full-time staff across the United States and is 100% employee-owned.

[Contact Us](#)