BlueGen ai

Scenario Exploration and Simulation for Policy Planning

What-If Simulations with Synthetic Populations for Official Statistics

09-10-2025 | Edwin Kooistra

Introduction

Public decisions increasingly hinge on questions about futures we have not observed. What if a new pandemic emerges? What if electric vehicle adoption accelerates? What if our population ages faster than forecast? Traditional datasets are anchored in historical reality and constrained by privacy rules, and they struggle to answer these counterfactuals at the speed that policy requires.

Why synthetic data enables scenario analysis: fast access, privacy by design, high value.

Simulation ready synthetic data closes this gap. By generating privacy safe and statistically faithful populations that mirror real world dynamics, statistical offices can run microsimulations and what if scenarios without exposing any individual's data. Analysts can vary assumptions such as policies, behaviors, and shocks; quantify impacts and uncertainty; and compare options side by side before public resources are committed.

This approach complements official statistics rather than replacing them. It extends today's toolkit, moving from "what happened" to "what could happen," while remaining within governance guardrails. With clear intended uses, validated utility, and documented privacy protections, NSOs can provide ministries and researchers with a repeatable and auditable pathway to evidence based foresight across domains such as health crises, the energy transition, and demographic change.

Scenario Exploration and "What-If" Simulation Analysis

Perhaps the most exciting application for synthetic data in official statistics is the ability to conduct rich "what-if" analyses through simulations. Policy makers constantly ask questions like: *"What if a new pandemic strikes?" "What if our population lives a decade longer on average?" "What if we adopt electric vehicles and green energy faster than expected?"* Traditionally, answering these scenario questions is difficult because we only have data from **historical reality**, not from hypothetical futures. This is where synthetic data – especially **simulation-ready synthetic populations** and data sets – can be a game changer. By creating synthetic data that represents the *unobserved* scenario (e.g. a future population or a rare disaster event), analysts can run **microsimulations** or model-based analyses to explore outcomes in a risk-free way.

Solid workflow for repeatable and auditable what-if analysis.

Health crises (pandemics):

The COVID-19 pandemic showed that policymakers need to rapidly evaluate interventions (lockdowns, vaccination strategies, etc.) before data from those interventions even exist. Synthetic populations and health data can be used to simulate disease spread and healthcare demand under various strategies. For instance, one could generate a synthetic dataset of patients with certain comorbidities, then simulate how a novel virus would propagate and what the hospitalization rates might be, under different containment policies. A study published in *NPJ Digital Medicine* demonstrated this approach in the context of demographic changes: researchers *"created a synthesized dataset and tested various healthcare policy scenarios"* under an aging population, effectively integrating multiple surveys into a unified synthetic population for simulation.

By re-weighting a baseline population to reflect future age structures, they could explore outcomes like health service demand under optimistic vs. pessimistic scenarios. The results allowed policymakers to observe potential ranges of impacts (e.g. on doctor visits, prescription rates) in a simulated future. *"The use of synthetic data in this example [enabled] policymakers to make decisions on resource allocation after analyzing a range of possible scenarios"*, giving insight into future healthcare demand and resource utilization.

	Baseline	Intervention A	Intervention B
Peak recordings			
Total recordings			
IC occupancy			
Lead time wave			
Expected healthcare costs			

Projected impact per policy option with uncertainty ranges.

Energy transition:

Achieving climate goals and shifting to sustainable energy systems is another domain where "what-if" analysis is critical. Energy infrastructure investments and policies have long-term impacts that need careful evaluation. Synthetic data is being used to build so-called **digital twins** of energy systems – virtual environments populated with synthetic agents (households, power grid components, etc.) that react to changes.

For example, BlueGen.ai collaborated with Dutch energy companies to generate synthetic smart meter data representing households' electricity usage. By using synthetic smart meter data, grid operators can simulate how adding thousands of electric vehicles or solar panels would impact the network, **without** having to collect years of real data (and without breaching consumer privacy on actual energy usage). Similarly, specialized firms like RUNWITHIT Synthetics have created **hyper-localized synthetic populations** of cities to model energy futures – including peoples' behaviors, new technologies, and policy changes – so that city planners can virtually test interventions (e.g. a new energy pricing scheme or a blackout scenario) and see the outcomes on grid stability and emissions.

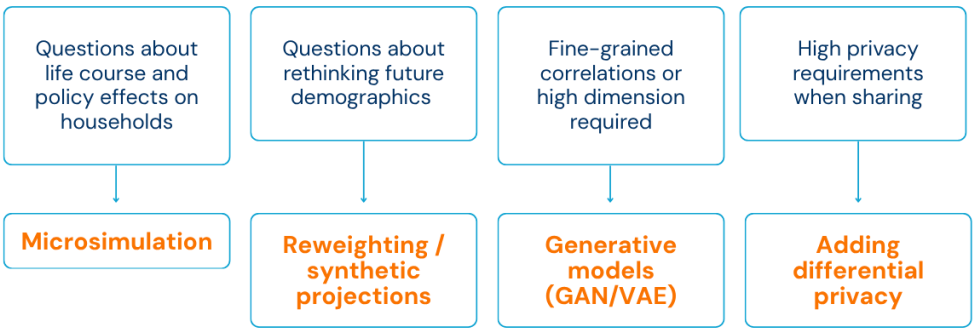
Key indicators for grid management under EV and solar scenarios.

Demographic shifts (aging populations):

Many countries are experiencing aging populations and need to anticipate impacts on pensions, healthcare, and the labor market. Synthetic demographic data can help here as well. One approach is to create a **"synthetic population"** – basically a simulated census of individuals with attributes statistically similar to the real population – and then age that population forward in time. Policymakers can then pose "what-if" questions to the model: *What if we raise the retirement age? What if medical innovations improve senior health?* – and see how the synthetic cohort's outcomes change. The value of these simulations is that they provide data-driven evidence for policy choices in areas where waiting for real-world evidence would be impractical or too late.

Beyond these examples, synthetic scenario analysis is being applied in many other domains. Disaster management agencies use synthetic data to model **natural disasters** and emergency responses – for example, generating synthetic data on households and infrastructures to simulate an earthquake's impact on a city and the effect of different evacuation plans. In all cases, the combination of **purpose-built synthetic data** and **simulation engines** forms a kind of "virtual laboratory" for policy experimentation.

Gartner analysts describe this advanced use: *"Purpose-built data is ideal for scenario planning and running intelligent simulations,"* and when detailed enough, *"synthetic data covering entire scenarios could predict future behavior of assets, processes, and customers,"* which is invaluable for planning. Such forward-looking use of synthetic data, akin to creating a *digital twin* of society or systems, is still emerging but holds great promise for evidence-based policymaking.



Choose the approach that suits your data, demand, and governance requirements.

Conclusion and Outlook

What if simulation on synthetic populations turns official statistics from a retrospective lens into a forward looking decision engine. It enables governments to test policies safely, explore trade offs transparently, and act faster under uncertainty, without compromising confidentiality.

To operationalize this capability, NSOs should:

1. Standardize a streamlined workflow from data synthesis to scenario design and result review.
2. Validate utility through distributional similarity and downstream model performance, and assure privacy through disclosure risk controls and, where needed, differential privacy.
3. Document assumptions, limitations, and intended use for each scenario product.
4. Iterate by monitoring real world outcomes and refining models, assumptions, and synthesis settings.

Executed well, simulation ready synthetic data becomes a core public asset that supports faster, fairer, and more resilient policy choices when it matters most.

Sources:

1. Miriam van der Sangen, *"Synthetic data opens up possibilities in the statistical field,"* CBS (Statistics Netherlands), May 2023. cbs.nlcbs.nl
2. Miriam van der Sangen, reporting on UNECE Synthetic Data Guide release, quoting Stats Canada and CBS researchers on synthetic data benefits and challenges. cbs.nlcbs.nl
3. Christopher Jones (UNECE) & Manel Slokom (CBS) on synthetic data accuracy, use for testing, and definition (CBS News article, 2023). cbs.nlcbs.nl
4. Statistics Canada Data Science Network, *"Unlocking the power of data synthesis with the starter guide on synthetic data,"* 2023 – Use cases and methods for synthetic data in NSOs. statcan.gc.castatcan.gc.ca
5. Mary Branscombe, *CIO.com – "Synthetic data's fine line between reward and disaster,"* May 2025. (Gartner insights on synthetic data growth and use in AI and simulations.) cio.comcio.com
6. BlueGen.ai Case Study – Dubai Statistics Establishment (2023) – CEO Younus Al Nasser on using synthetic data for innovation with privacy. bluegen.ai
7. BlueGen.ai Case Study – Alliander (Dutch energy grid) (2022) – Synthetic smart meter data for grid planning (privacy-compliant). bluegen.ai
8. BlueGen.ai Case Study – IQVIA (Health analytics) – Quality of synthetic data yielding same insights as real data. bluegen.ai
9. Mauro Giuffrè et al., *npj Digital Medicine (2023)* – Perspective on synthetic data in healthcare; includes policy scenario simulation for demographic aging. nature.comnature.com
10. Consumer Data Research Centre (UK) – *"Synthetic Population and Scenario Projection – Future Elderly Model,"* describing microsimulation for aging population policy. cdrc.ac.ukcdrc.ac.uk