When Synthetic Data Works:
A Validation of Augmented Sample Usage

Dig insights' validation of Fairgen's sample boosting method

Why This Matters

Synthetic data is one of the most talked-about innovations in market research, but trust has lagged behind the hype. Insights pros and business leaders need more than bold claims, they need proof of when, how, and who it works for.

That's why Dig Insights conducted an independent, multi-metric validation of **Fairgen's synthetic boosting method**, a technique that uses machine learning to strengthen the statistical reliability of small, underrepresented subgroups in survey data.]

What We Found Out

- **Accuracy can be gained**
  stronger distributions and more reliable relationships between variables, especially in niche subgroups.
- **Sample size effectively doubled**
  In most cases, the boosted subgroups performed as if they had twice as many respondents.
- **Cost efficiency potential**
  More reliable insights from niche audiences without **additional field spend.**
- **Dependency on inputs**
  Boosting works best with robust starting data; weak inputs risk amplifying noise.

For researchers, this means the confidence to apply synthetic data when the right conditions are in place. For business leaders, it means faster, lower-cost paths to reliable insights without compromising quality.

Introduction: The Promise And Proof Gap

Synthetic data - data that's generated by algorithms or models rather than collected through primary data collection, has captured the attention of the insights industry for good reason. Whether it's simulating survey responses, creating synthetic interview participants, or boosting small subgroups within a dataset, the promise is faster timelines, lower costs, and access to insights from hard-to-reach audiences.

But there's a gap.

While the market is flooded with bold claims, very few are backed by **independent, transparent validation.** Public "proof" is often limited to vendor-led studies (which call into question biases), limited metrics, or one-off case examples, leaving you with a credibility problem.

Dig Insights set out to close this "proof gap." Partnering with Fairgen, we ran a rigorous, multi-metric evaluation of one method: **boosting small subsamples.** Our goal was to see not only if it worked, but also to raise the bar for how synthetic data is tested, validated, and ultimately trusted in the insights industry

Dig Insights

Dig's Role as an Independent Validator

1

Utility & curiosity

Like other research providers, we want to know when and how it works so we can confidently apply it.

2

Setting a higher standard

Most public "validations" are narrow, relying on a single dataset or metric. Our approach tested multiple audiences across two datasets and applied several complementary metrics. This reduces the risk of bias and creates a validation that both researchers and decision-makers can trust.

3

Transparency

Fairgen allowed us to publish results regardless of outcome and shared details of their process. That gave us the confidence to design a test to the same standards we apply in client-facing research.

Importantly, Dig has no commercial ties to Fairgen. Our only stake was in ensuring the validation met the standards we would trust in our own client work.

Understanding Fairgen's Approach

Synthetic data can be:

- **Fully synthetic –** generated from scratch (e.g., by a large language model).
- **Augmented –** boosting small real-world subsamples using machine learning.

Fairgen focuses on augmentation, boosting subgroups with 15–350 respondents (0.1–20% of a study sample). We began our validation here because:

- **Theoretical grounding:** augmenting within a larger dataset leverages context from the rest of the sample.
- **Clear claims:** Fairgen's promises were specific and measurable, making them testable.

Dig Insights

Methodology

Our goal was simple: test whether Fairgen's boosting method could enhance the accuracy and effective sample size of small subgroups in survey data, without compromising research quality.

Datasets

We ran the validation on two very different datasets:

Telecom company: n = 26,372

Quick Service Restaurant (QSR): n = 2,258

This contrast (large national vs. smaller dataset) let us see how the method performs under different conditions.

Subgroups

We focused on the kinds of audiences researchers often struggle to analyze due to small sample sizes:

Telecom: Saskatchewan (n=21), Nova Scotia (n=40), 18–24-year-olds, newcomers, and "No Tweets."

QSR: Young adults (≤21, n=10), older adults (60+, n=35), rare QSR visitors, Asian/Pacific Islanders, and high-income respondents.

Train/Test Split

Each dataset was split into:

Training set – fed to Fairgen's algorithm to learn **patterns in the data (e.g., how demographics link to behaviors). Using these patterns, the algorithm then applied the boosting process to generate synthetic respondents for underrepresented subgroups.

Test set – kept separate (never exposed to Fairgen's algorithm) as the ground truth to compare both raw and boosted subgroups against.

Evaluation metrics:

We applied three complementary metrics:

1. Distribution accuracy (univariate/marginal)
   Do the boosted responses line up more closely with the true results from the larger dataset?
   - Measured by Mean Absolute Error (MAE) on response distributions: the average difference between boosted subgroup response distributions and those in the ground-truth dataset.
     Lower MAE = less error, more accurate results.
2. Relationship accuracy (pairwise)
   Do patterns hold up more reliably after boosting?
   - Measured by Mean Absolute Error (MAE) on correlations: i.e., ****whether boosted subgroup patterns move closer to those in the ground-truth dataset.
     Lower MAE = stronger, more reliable patterns.
   - Example: Age may predict fast-food visits in the larger dataset. In a small subgroup, that pattern can get distorted, does boosting pull it back toward the true trend?
3. Effective Sample Size (ESS)
   After boosting, how big is the new subgroup sample?
   - When we boost data, it improves reliability. ESS translates that improvement into sample size terms. For example, boosting can make results from 20 people behave as reliably as results from 50. In simple terms, a small improvement in accuracy (from ±7% to ±5%) is like doubling the sample size — an "ESS boost factor" of 2.

By looking at numbers (MAE on responses), patterns (MAE on correlations), and effective size (ESS), we ensured the improvements weren't just cosmetic, they made the data more reliable for real-world decisions.

Dig Insights

Results

Across both studies, Fairgen's boosting turned groups normally "too small to study" into reliable, analyzable audiences. It had the biggest impact on the smallest subgroups in large datasets and delivered consistent gains across all groups in smaller datasets.

Study 1 – Telecom (Large dataset)

- **Overall accuracy improved:** MAE dropped from 3.7% → 2.8%.
- **Biggest lift for smallest groups:** Saskatchewan (n=21) cut error by 2.5 points and more than doubled ESS (+153%). Nova Scotia (n=40) more than doubled ESS (+165%).
- **Mid-sized groups saw little change:** 18–24-year-olds and newcomers showed no real improvement, suggesting boosting works best on the smallest subsamples.
- **Stronger patterns:** Correlations between variables (e.g., age and behavior) held up more clearly after boosting. Average correlation error dropped by 0.03.

		Distributions		Correlations		ESS	
Segment	Original n	MAE Original [%]	MAE with Boost	MAE	MAE with Boost	Boosted n equivalent	ESS Boost Factor
Sask	21	6.9%	4.3%	0.18	0.13	53	+153%
Nova Scotia	40	4.9%	3.0%	0.14	0.09	106	+165%
NoTweets	237	2.0%	1.5%	0.06	0.04	426	+80%
18-24 yrs	157	2.4%	2.6%	0.07	0.05	130	-17%
Newcomers	198	2.4%	2.5%	0.07	0.05	174	-12%
Average		3.7%	2.8%	0.10	0.07		+74%

Table 1.
Boosting made the hard-to-analyze groups (small provinces and niche segments) behave like larger samples, cutting error rates and strengthening patterns.

Results

Across both studies, Fairgen's boosting turned groups normally "too small to study" into reliable, analyzable audiences. It had the biggest impact on the smallest subgroups in large datasets and delivered consistent gains across all groups in smaller datasets.

Study 2 – QSR (Smaller dataset)

- **Overall accuracy improved:** MAE dropped from 11.1% → 7.6%
- **Every subgroup benefited:** All five improved, with average error down 3.5 points.
- **Patterns clearer across the board:** In every subgroup, the usual relationships (correlations: like how age or income predict behavior) held up more reliably after boosting. Correlation error fell by 0.08 on average.
- **Sample power multiplied:** Effective Sample Size increased everywhere, averaging +124% — meaning results behaved as if samples were more than twice as large.

		Distributions		Correlations		ESS	
Segment	Original n	MAE Original [%]	MAE with Boost	MAE	MAE with Boost	Boosted n equivalent	ESS Boost Factor
<=21 years old	10	12.1%	7.0%	0.33	0.23	29	+190%
60+ years old	35	5.5%	3.1%	0.18	0.11	104	+198%
Rarely Eats At QSR	15	22.1%	15.8%	0.20	0.12	29	+94%
Asian / Pacific Islander	12	7.5%	5.6%	0.28	0.23	21	+71%
High Income ($175k+)	11	8.6%	6.5%	0.34	0.24	19	+69%
Average		11.1%	7.6%	0.26	0.19		+124%

Table 2.
Gains were universal: lower error, stronger patterns, and bigger effective samples across every subgroup.

Limitations

While our validation shows meaningful gains, it's important to recognize where we did **not** test, and where challenges remain:

- **Beyond pairwise correlations**
  Our analysis focused on distributions accuracy and pairwise correlations. We did not evaluate the effect of boosting on more complex multivariable models (e.g., drivers analyses).

- **Operational complexity**
  Applying multiple boosts to the same dataset is not straightforward, which creates workflow challenges for real-world adoption.

- **Uneven gains**
  The strongest improvements were seen in the smallest subgroups; mid-sized groups showed little to no change. This points to conditions where boosting is most effective, and where it may add limited value.

Conclusion

Across two very different test settings: one broad dataset and one smaller dataset, Fairgen's boosting method consistently improved results. Both distribution accuracy (univariate/marginals) and relationships between variables were more reliable in the boosted subgroups. The smallest subgroups gained the most, often more than doubling their effective sample size, while broader groups held steady. **This means researchers can draw faster, more confident insights from niche audiences without extra fieldwork.**

At the same time, results show clear boundaries: Fairgen's method works best when starting data is robust; weak inputs risk amplifying noise rather than signal. Operationally, applying multiple boosts presents challenges, and this validation did not assess performance on multivariable models.

For Dig Insights, the bigger lesson is that synthetic augmentation can reshape research, but only if adoption is grounded in transparent testing and clear proof of when it works. That's why we see Fairgen's approach as a credible step forward, and why we're committed to rigorous validation that builds trust in new methods.