MMR POV – In a Nutshell

Synthetic Data

Mat Lintern, Simon Harris, Dr. Steve Ferris, Dr. Jonathan Sands

Why synthetic data is a hot topic – and where MMR stands

The role of synthetic data in product testing is certainly a hot topic! With huge product heritage, expertise and in-house Statistics and Data Science teams, MMR are perfectly placed to advise on situations in which synthetic data could play a role, as well as highlighting important considerations when making the decision to use synthetic data.

We've invested in significant experimentation in this space and are keen to leverage these findings through client collaboration, helping to navigate and overcome some of the challenges and risks we've observed.

In a nutshell

There are lots of exciting ways that synthetic data can be used in research, from Personas to Digital Twins. But here we focus on one particular application: boosting small quant base sizes with synthetic data.

To be clear: The approaches used to synthetically boost data **work technically** (the good ones anyway –there are several watchouts!)

But we're **sceptical about using this synthetic data to boost small quant base sizes purely with the intention of adding statistical confidence to results – in fact we think this is wrong!**

Important limitations you should be aware of

While other applications such as Digital Twins and personas can be different, the synthetic data solutions we've explored in this boosting context generate their data solely from the consumer information the model is fed.

They basically fill the gaps between the people you've interviewed, creating additional rows of simulated respondents.

Broader data sources and macro trends are not leveraged, therefore **confidence in the results is entirely dependent on how representative the original dataset is**.

And you can't bend the laws of sampling!

The smaller the base size, the greater the level of uncertainty around the results. That's just how it works, and simulating data in this way doesn't change that!

Request our full point of view article for more on this.

Hence, synthetically boosting from a small base size shouldn't (in our opinion) increase confidence in results – even though some suppliers claim otherwise.

Product testing brings additional complexity

People just like different things – we find this in every test, in every category, in every country. Some people like the sweet one, some the sour. Some like the crunchy one, some the chewier one. It's another fundamental law (this time of product testing) that can't be ignored!

This phenomenon always leads to preference segments (which exist even if they aren't analyzed as part of the project). While you can balance for demographics or product usage at recruitment, it's just not possible (in normal situations) to balance for the size and influence of taste preference segments – and this can skew results appreciably in small-scale product testing.

We're actively exploring solutions for this at MMR… but rest assured it's not an easy one to solve, at least not in an 'agile' way!
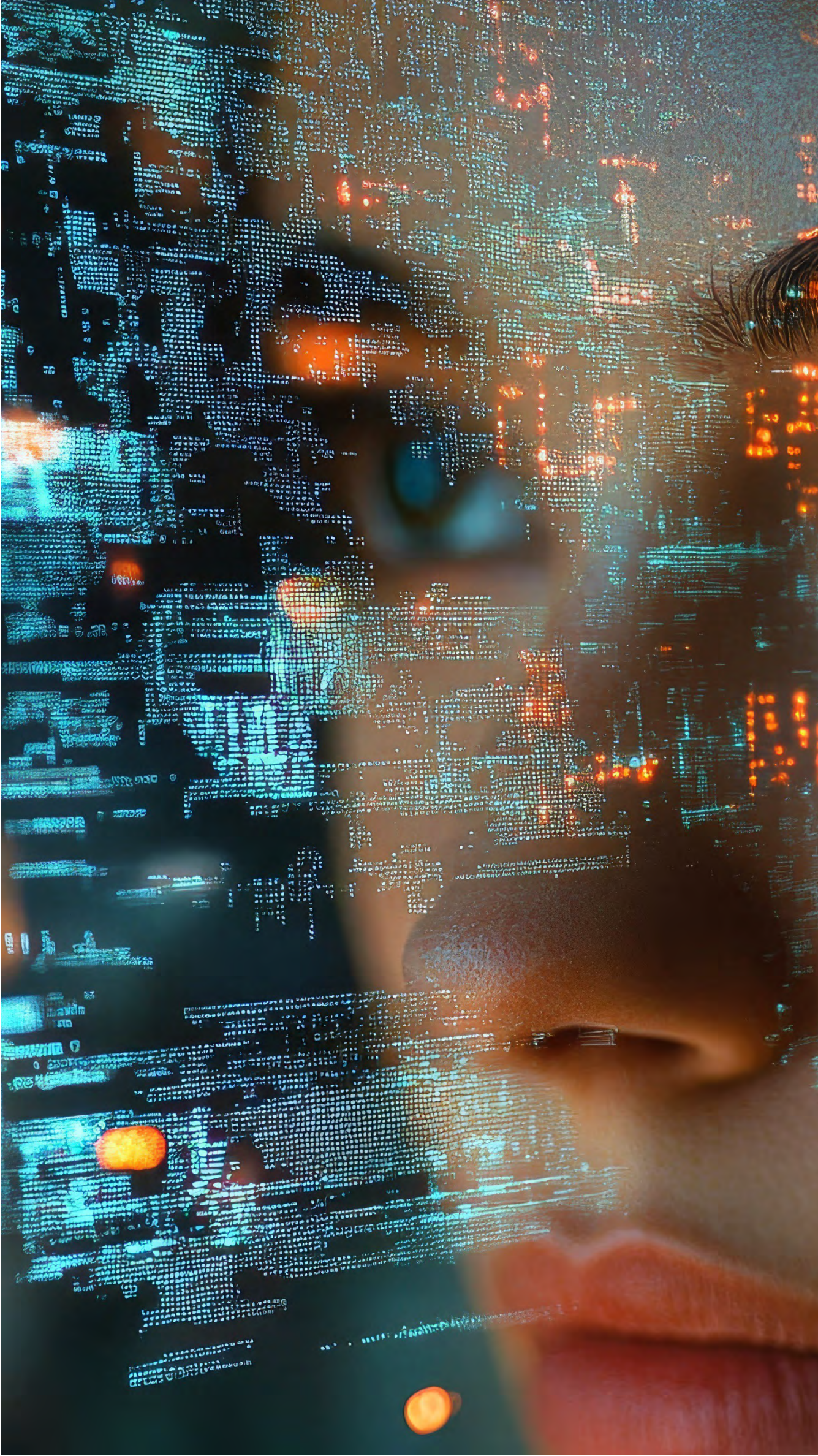
False confidence vs. real risk

It's easy to "prove" synthetic data is a close proxy to the small sample you interviewed (i.e. it mirrors the means and even the data structure). But that's just showing it matches what you already collected. It's self-fulfilling!

The real problem is whether the small sample gave the right read in the first place – and that's very hard to determine.

That's why we use larger base sizes in research: to reduce risk! And that's our main concern. If we're treating this application of synthetic data as a means of 'magically' turning 50 into 200 and upping your statistical confidence as a result – well, we just don't think it does. It's modern-day alchemy!

So, where can synthetic data help?

Synthetic datasets do unlock the potential for more advanced analysis techniques (which we're excited about) – which can be particularly helpful in modelling product optimization scenarios.

But they don't reduce the risk associated with a smaller starting sample – a powerful stats technique very much relies on the quality of data that feeds it!

What MMR is exploring

We aren't against the use of synthetic boosting, especially where it can allow the use of more advanced modelling techniques. As such, **we've developed an approach, designed specifically for product testing, that overcomes some of the technical fragilities that exist in other models**. We're very happy to use this with clients to explore the area and help to understand the risks, benefits and use cases – we just want to urge everyone to go into this with their eyes open to the compromises and the risks.

Further to this we're looking to develop more sophisticated approaches:

Leveraging our historic product testing data

Tackling nuances specific to product research (e.g. taste segment variability)

Investigating better recruitment methods to boost representativeness in smaller samples

But, this is a journey, and unlikely to be a quick one. WATCH THIS SPACE!

Practical alternatives

We understand the importance of having the option to reduce cost and time in product testing. So, as we say above, we're right up to speed with the latest in synthetic boosting techniques.
But we're also keen to suggest an alternative for clients. Instead of essentially magnifying a small base size (think about magnifying a low-res photo!), an **alternative to consider is augmenting the small number of people you've spoken to with richer detail:**

Run a product test with a smaller base size, accepting manageable risk levels

Set a meaningful decision threshold (e.g. 0.x), i.e. decide what you regard as a meaningful difference and look for that rather than relying on significance testing on a small base

Use agile sensory techniques to provide a richer level of product diagnostic detail from a trained panel – which MMR's experts can match onto the consumer data

Leverage the consumers you speak to more – using MMR's advanced conversational AI tools to delve deeper into their product experience

This provides much deeper diagnostic insight, with greater agility and lower cost, while avoiding the risk of false confidence that can be associated with synthetically boosted samples.

Let's collaborate

We'd love to explore scenarios and solutions with our clients, and to advise on when to use synthetic data – and when not to.

GET IN TOUCH

Simon Harris
Innovation Director: Product Excellence
Nova Innovation, MMR

S.Harris@mmr-research.com

Or speak to your MMR Account Director.