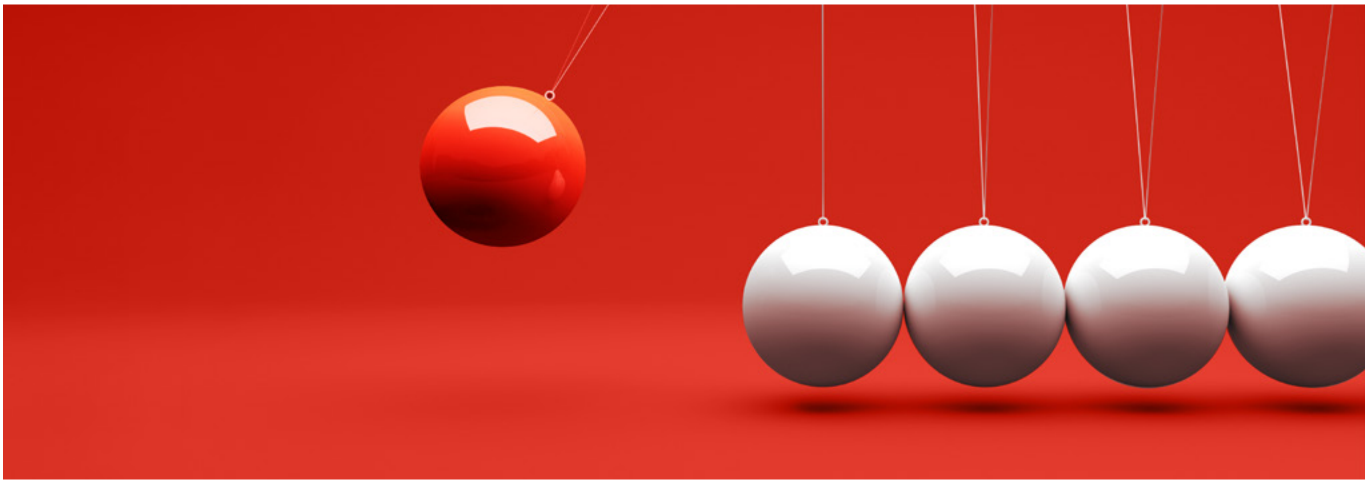
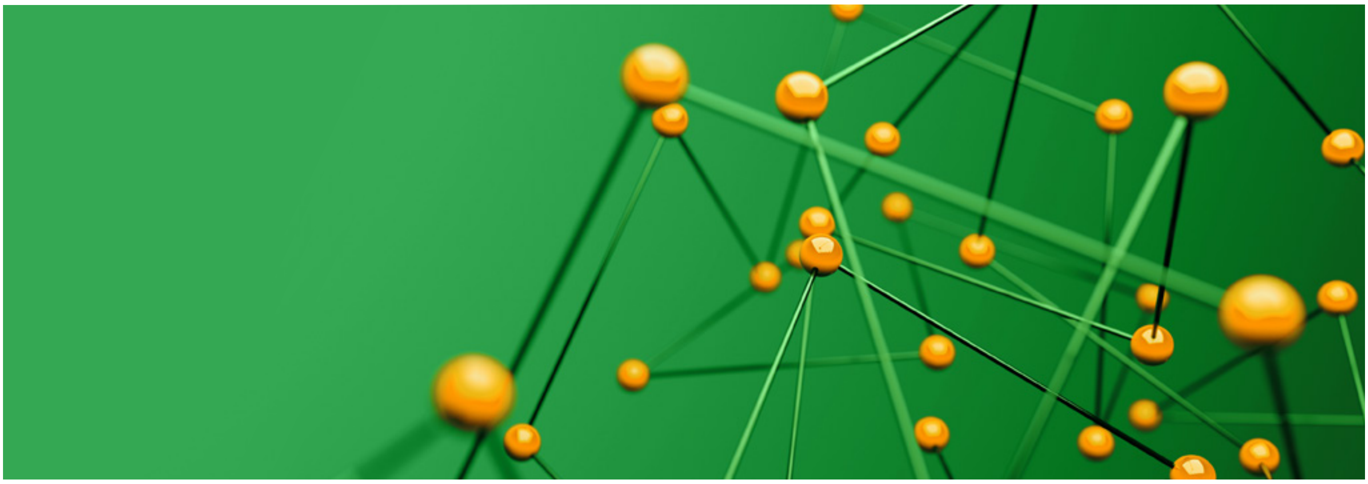




Measuring effectiveness

Three Grand Challenges

The state of the art and opportunities for innovation

Contents

Foreword

At Google, we work with advertisers, agencies and third-party measurement providers to help measure the effectiveness of marketing. We're also fortunate to be building new measurement products ourselves.

This work has provided significant insight on where the state of the art lies, and because we work across countries, industries and platforms, we have a broad perspective. So, while it won't be a complete picture, we think it's time to share our observations about where effectiveness measurement is running up against the boundaries of the possible.

In this paper we'll also go a step further and look just beyond those boundaries. What are the biggest unsolved problems that the industry is grappling with? And where are the opportunities to make real progress in the next three to five years?

We will set out some ambitious goals for the industry, aiming to bring people together, stimulate innovation, and capture the imagination. Three 'Grand Challenges', where progress would make a huge difference in measuring what works across the whole of marketing.

While we can all make contributions, no single party, including Google, will ever 'solve' these problems on its own. But we would like to work with experts in effectiveness measurement to develop new approaches that lie just beyond the cutting edge.

Together, we can make real progress. So we look forward to working with you during 2019 — and beyond.

Author:

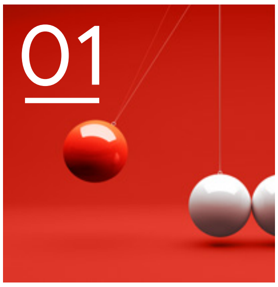
Matthew Taylor
Effectiveness Specialist, UK

Contributors:

Baptiste Tougeron, Effectiveness Specialist, Southern Europe
Ludwig Bruetting, Effectiveness Specialist, DACH/CEE/MENA
Jonas Christensen, Effectiveness Specialist, Northern Europe

Executive summary

Marketers face many obstacles to measuring effectiveness, but there are also exciting opportunities for the future. We therefore propose three 'Grand Challenges' that the industry should tackle, together. They are not the only challenges that exist, but solving them would make a significant impact.

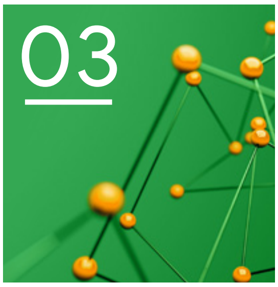


Incrementality: proving cause and effect

Estimating the true effects of marketing interventions when isolated from other factors.

Measuring the long term, today

Using immediately available granular data to estimate how marketing might deliver returns in the future.



Unified methods: a theory of everything

Combining different measurement methods to get a rounded and more accurate view of marketing performance.

01. Incrementality: proving cause and effect

In medicine, proving cause and effect is literally a question of life and death. Survival rates improve with the introduction of a new drug. But is this causation or merely correlation?

In the world of marketing the stakes may not be as high, but the problem remains the same. Sales increase after a campaign, but was this caused by the marketing — or by something else?

The process of estimating the true effect of an intervention (for example marketing) on an outcome (for example sales) is known as 'causal inference'. And common effectiveness measurement methods don't always get it right.

Rather than abandoning these trusted methods altogether, marketers can adopt medicine's approach: a 'hierarchy of evidence' that favours methods higher up the hierarchy. Where it's not possible to employ the 'gold standard', marketers should be aware of the limitations of their research and the uncertainty attached to their results. But experts should take care over how and when they communicate uncertainty, so that it doesn't hinder decision-making.

Randomised controlled experiments: the gold standard

The challenge: Randomised controlled experiments are the most accurate method of measuring causal effects. But they can typically only test one or two things at a time, they can be difficult to administer, and they require user or regional data.

The opportunity: What if marketers could use experiments more often following agreed best practice, use them to measure more factors simultaneously, and apply them across the media mix so that all channels are held to the same standard of evidence?

Better causal inference from observational data

The challenge: When marketers can't run experiments, they must rely on methods that use observational data such as marketing mix modelling[1] or digital attribution. These techniques that are not always good at estimating causal effects.

The opportunity: Algorithms now exist to determine when a particular observational method is a good estimate of causal effects and when it is not. Such algorithms can analyse 'causal diagrams' which encode our understanding of how an intervention causes an outcome alongside other factors.[2] These have seldom been applied to marketing effectiveness.[3] What if we could develop tools allowing marketers to build causal diagrams, to analyse them automatically and to recommend optimal methods for estimating causal effects?

Communicating uncertainty

The challenge: Regardless of the measurement method (even the 'gold standard' is not flawless) we need to acknowledge and understand the error margins present in estimates of marketing effectiveness.

The opportunity: What if effectiveness experts could understand when communicating uncertainty is useful in making marketing decisions, and build tools that surface it in the right way, at the right time and to the right people?

1 Also 'MMM', 'econometrics'
2 (Pearl. 2018). The Seven Tools of Causal Inference with Reflections on Machine Learning.
3 Note that the algorithms are different from (and are a potential enhancement to) Structural Equation Models and Bayesian Networks, which are sometimes used to measure marketing effectiveness.

02. Measuring the long term, today

CEOs must continually balance business decisions that drive long-term growth against the short-term returns demanded by shareholders.

Marketers walk the same tightrope: investing in a brand is essential to growing a sustainable business, but may not drive quarterly sales at a good return on investment. According to some sources,[4] marketers have become too focused on the short term, and this damages effectiveness. A successful long-term strategy requires better measurement of long-term effects, without having to wait years for the results.

Projecting long-term MMM results back into the short term

The challenge: Advanced MMM approaches can estimate the long-term sales delivered by marketing and create a 'multiplier' comparing this to short-term sales. But this may require many years of data and is more expensive than regular MMM, meaning the approach isn't often used. Marketers may also employ multipliers out of context or without understanding their error margins.

The opportunity: What if effectiveness experts could agree best-practice standards for long-term MMM in particular, so that it can measure the effects of smaller campaigns or channels? What if they could come up with guidelines about when long-term multipliers should and should not be used to project long-term effects?

Projecting customer lifetime value (LTV) back into the short term

The challenge: With good customer data, marketers can model the expected LTV of potential customers, so they can be targeted with customised messages or increased spend. But few advertisers have the quality of data and systems required to modify assumptions based on real behaviour and to connect this with automated marketing.

The opportunity: What if developments in machine learning could enable marketers to predict the lifetime value of potential customers using less data, allowing them to adjust marketing spend and continually update predictions based on actual behaviour?

Using online behaviours as a leading indicator of long-term outcomes

The challenge: Marketers often have to wait months or years to measure long-term effects, by which time the world has moved on. Meanwhile, some of the measures which are available in the short-term, such as brand health, are taken infrequently and from small samples. As a result, it's difficult to identify which changes were driven by marketing.

The opportunity: Online behaviours (such as search queries) can provide frequent, granular data with very large sample sizes. Research shows that this data is related to brand health[5] and may be a leading indicator of long-term outcomes. What if effectiveness experts could agree best practice on how to measure this? What if they could attach an estimated financial value to the uplifts in online behaviour caused by marketing?

4 Media in Focus, IPA 2017, Mounting Risks to Marketing Effectiveness, Enders Analysis 2017
5 Brand Attitudes and Search Engine Queries, Dotson et al 2017

03. Unified methods: a theory of everything

For decades, scientists have sought a way to marry the theory of the very small (quantum mechanics) with the theory of the very large (classical physics), to develop a so-called 'theory of everything'.

A similar challenge is emerging in effectiveness measurement. Consumer-level models like digital attribution measure at the level of the very small. Aggregate-level models like marketing mix modelling measure the very large. And they can lead to very different results.

Randomised controlled experiments are the gold standard for measuring very specific marketing interventions, but their results are often different again.

The logical next step, therefore, is to find a way to bring together measurement methods to get a rounded view of their effectiveness: a 'theory of everything' for marketing.

Blending one method with another

The challenge: While some advertisers and measurement providers claim they are blending MMM with digital attribution, MMM with experiments, or digital attribution with experiments, they form a minority, and best practice remains unclear.

The opportunity: What if effectiveness experts could come together to explore opportunities for blending methods, examining the pros and cons and setting standards for the industry?

Blending multiple methods at once

The challenge: Beyond traditional methods, effectiveness experts seek to blend data and insight from different sources. This can involve collating and presenting it all in a digestible way, or blending it on an analytical level and presenting a unified final result. Some promising approaches exist, but there are no industry standards.

The opportunity: What if effectiveness experts could agree on the ideal process for bringing together and presenting multiple sources of information? What if researchers could build transparent models to blend data of different granularities (user, segment, geo, aggregated) to get consistent and holistic measures of effectiveness?

The role of expert judgement

The challenge: In the absence of perfect measurement, marketers must navigate uncertainty to make decisions that they think will drive the best future returns, based on their experience and expertise. But such individual wisdom is often held in organisational silos and not combined in a structured way.

The opportunity: What if researchers could demonstrate that sourcing and combining multiple expert opinions delivers better results than individual opinions, and may even be as accurate as some measurement methods? What if effectiveness experts could develop tools that collect and combine these opinions in order to help marketers make decisions?

Conclusions

Across each of the three challenges, some common themes emerge which point towards the future of marketing effectiveness measurement.

Striving for the best, but embracing the possible

Driven by ever greater scrutiny on the bottom line, marketers will increasingly have to aim for the highest standard of evidence. They won't always get there, so they should be pragmatic, aware of any potential uncertainty in estimates, and make decisions nevertheless.

Progress will come with transparency and clear communication

There is a gap between the academic research into effectiveness measurement and the marketing materials of measurement providers. Effectiveness experts need to fill this gap with transparent, open-source research that is robust but also communicated in a way that makes it accessible and digestible for marketers.

Human judgement: more critical than ever

Perhaps the most surprising theme to emerge is that solving these challenges won't put measurement experts out of a job. Rather, expert oversight is likely to become increasingly important: a study by the Boston Consulting Group in 2018 found that "Companies using advanced technology with active human supervision can improve their campaign performance by up to 35%."[6]

Coming together and finding common ground

Our intention in proposing these challenges isn't to try and set the terms of industry engagement, or to force the direction of conversation. Between different advertisers, media platforms and technology providers, there will be a wide range of opinions about the importance of each issue, and the perceived benefit of investing in solutions.

Instead, we hope the challenges will be seen for what they are — an invitation — one that we're extending to passionate marketers, effectiveness experts and data scientists in every corner of the industry. If you have a view on any of the issues raised here, please get in touch — we'd love to hear what you think.

6 BCG: Dividends of Digital Marketing Maturity, 2019

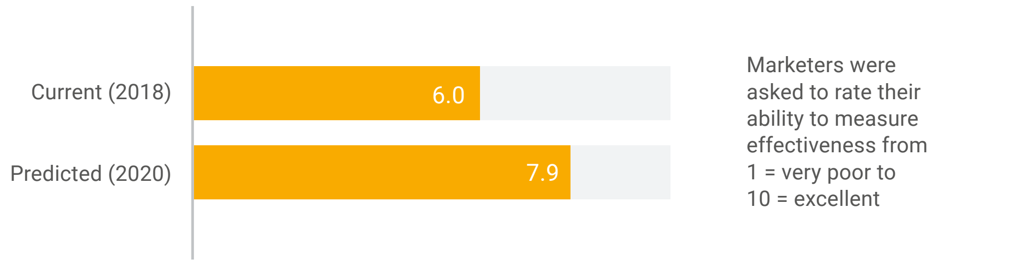
Introduction

In marketing, the term 'effectiveness' is used to describe how campaigns or channels drive both brand effects (for example, awareness, consideration and intent) and business outcomes such as sales. Accurately assessing the marketing that was delivered (for example, reach, frequency) is crucial too. But that presents different challenges, beyond the scope of this paper.

Effectiveness goes hand in hand with 'efficiency': if effectiveness is what marketing delivers in terms of brand and sales, efficiency is how much it cost, and whether it was worth it.

Measuring effectiveness is hard, but marketers are optimistic. When surveyed in mid-2018, a sample of 213 UK client and agency marketers rated themselves 6/10 on their ability to measure effectiveness. But the same group predicted that they would reach 8/10 by 2020.[7]

MARKETERS' OPTIMISM



Source: Greengrass Consulting for IPA & ISBA, 2018

So, how do they plan to get there? In the rest of this paper, we'll outline three of the biggest challenges shaping the future of effectiveness measurement, take a look at the state of the art, and explore opportunities to move beyond it.

7 Marketing Effectiveness: How is it working in practice? Greengrass Consulting for IPA and ISBA, 2018 https://www.isba.org.uk/media/1861/marketing-effectiveness.pdf

Three Grand Challenges

Marketers face many obstacles to measuring effectiveness, but there are also exciting opportunities for the future. Solving the three 'Grand Challenges' we've identified is the first step to making a difference:

01

Incrementality: proving cause and effect

Estimating the true effects of marketing interventions by accurately isolating them from the impact of other factors.

02

Measuring the long term, today

Using immediately available granular data to estimate how today's marketing might deliver returns weeks, months or years from now.

03

Unified methods: a theory of everything

Combining the major effectiveness measurement methods to get a rounded and more accurate view of marketing performance.

01

Incrementality: proving cause and effect

This is a problem as old as marketing itself. Sales rise just after a campaign, but was this caused by the marketing or was it just correlated, and actually caused by something else?

Most measurement methods try to adjust for other factors, isolating the effect of the campaigns or channels they want to measure. But they don't always do this well.

The process of estimating the true causal effect of an intervention (for instance, a marketing campaign) on an outcome (for instance, sales revenue) is known as causal inference. The rules of causal inference can be used to grade measurement methods according to their ability to accurately measure causal effects.[8]

Rank	Measurement method	Description
1	Meta-analysis or systematic review of randomised controlled experiments	Analysis of the overall findings across many randomised controlled experiments.
2	Randomised controlled experiment	Test units (for instance, users, people, geos) are matched into groups that are as similar as possible. The 'treatment' is deliberately applied at random to one group and withheld from the other. The difference in outcomes (for example, sales) is the estimated effect of the treatment (for example, ad exposure).
3	Cohort study	Cohorts (groups) of people are found who just happened to be exposed to the 'treatment' and other groups who were not. Again the difference in outcomes is the estimated effect.
4	Case control study	The histories of a group of people who have the desired outcome (for example, purchase) and a group who do not are compared to see whether they previously had the treatment or not.
5	Case studies or series	Individual cases or a collection of them, not systematically analysed.
6	Expert opinion	Judgement of an expert in the field.

In the fields of science and medicine, causal inference can be literally a matter of life and death. So it's no surprise that, in many countries, for a new drug to be approved it must be tested right at the top of the hierarchy. Randomised controlled experiments will be conducted against a placebo or the best existing treatment, followed by systematic reviews across many experiments in order to validate the findings.

However, experiments can be expensive and difficult to execute, and typically they can only test one or two interventions at a time. So, when and how can we apply them to marketing effectiveness? And if it's not possible to test in this way, what is the best available alternative?

8 There are many different versions of this table in the literature. This is a simplified overview

We've identified three key areas of opportunity:

01 Randomised controlled experiments: the gold standard

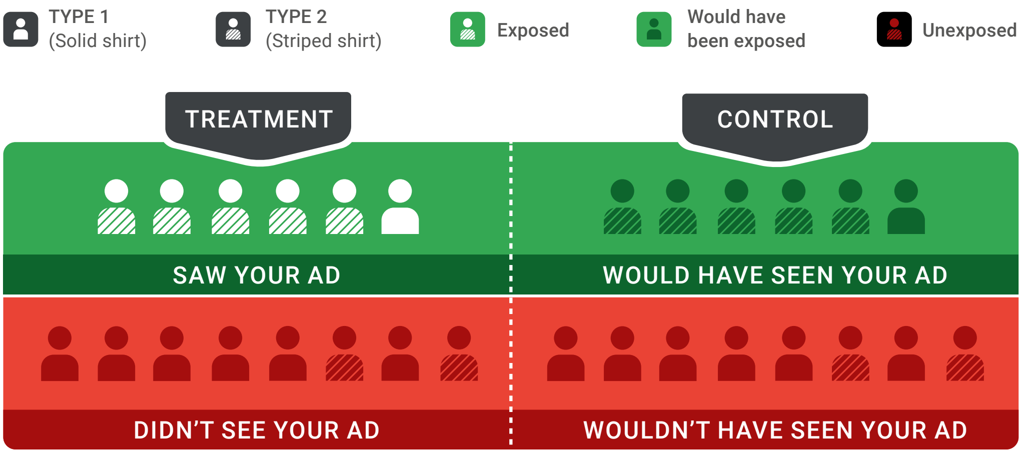
02 Better causal inference from observational data

03 Communicating uncertainty

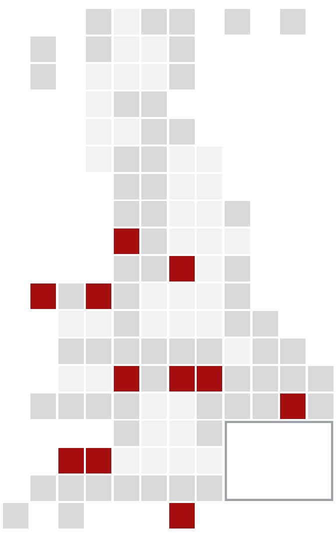
01 Randomised controlled experiments: the gold standard

Randomised controlled experiments have been used in marketing for some time, in the form of user-level and geo-level analyses.

For example, products such as Brand Lift allocate specific users either to a test group where they are shown the ad being evaluated; or to a control group where the ads are withheld, with those users seeing the next ad in the auction instead. The difference in brand metrics between the test and control groups is known as the 'lift' due to the ad.[9]



When there is no user-level data, many advertisers and measurement providers use geo tests instead. Geographic regions are matched into a test group and a control group, where each group has similar sales patterns and other relevant factors. The ads are then delivered in test geos and withheld from control geos, and the results compared.[10]



The key point about both these methods is that the treatment (in this case advertising) is randomised across similar users or geos, creating a test group that should differ from the control group only because they have seen the advertising. Therefore, other factors that could cause an uplift in search or sales are 'designed out' before the experiment. Such randomised tests ought to be the gold standard in measuring effectiveness, and we should aim for them where possible.

However, the process presents some challenges. The tests require careful design, and it's easy to omit important factors. For example, imagine conducting a geo test on cat food ads, and missing the fact that a regional promotion was run in test areas just beforehand. The offer makes consumers stock up on cat food, so that sales fall just when the ads might have pushed them up. The ads may still drive some sales, but the overall decline makes effectiveness impossible to measure.

9 A revolution in measuring ad effectiveness: knowing who would have been exposed https://www.thinkwithgoogle.com/intl/en-gb/marketing-resources/data-measurement/a-revolution-in-measuring-ad-effectiveness/
10 Measuring ad effectiveness using geo experiments (Vaver et al 2011) https://ai.google/research/pubs/pub38355

In this kind of test, it can also be hard to measure several factors at once. Let's say you want to run a geo test across 10 regions, to compare the effectiveness of 6-second and 20-second online video formats. To do this you would need four 'cells' of regions matched as closely as possible, which might be quite difficult. If you want to compare creative type at the same time, you'll need six cells; and if you add targeting types too, you'll need 12. The test now requires more cells than there are regions, and is impossible to conduct.

Video	6-sec	20-sec
Test	1	2
Control	3	4

Video	6-sec	20-sec
Awareness Creative	1	2
Purchase Creative	3	4
Control	5	6

	Demographic targeting		Interest targeting	
Video	**6-sec**	**20-sec**	**6-sec**	**20-sec**
Awareness Creative	1	2	3	4
Purchase Creative	5	6	7	8
Control	9	10	11	12

The execution of experiments can also present challenges that are specific to marketing. In a drug trial it's fairly easy to control who receives one drug and who does not – and it's even possible to estimate how many people in the treatment group didn't actually take the drug, or did so at the wrong dose.

In marketing the lines may be more blurred. In geo experiments, for example, someone living in a control region might travel into a test region and be exposed by mistake – but it's hard to know if and when that's happened. Likewise, for panel experiments there can be blind spots where exposure to certain channels or devices can't be tracked.

Moreover, if the tests are in themselves quite 'narrow' (for example, you're targeting people interested in a product in a competitive auction with a budget cap) it can be difficult to deliver enough of the treatment in the test cells (for example, number of impressions per region) to get a statistically significant read on the effects.

There can also be a clash of objectives. Marketers are paid to meet brand and sales growth targets across their whole group of target consumers. How many of these consumers are they prepared to keep sacrificing as control groups purely to measure effectiveness?

Despite these challenges, the appropriate use of randomised controlled experiments should continue to grow in marketing — the size of the prize is worth the effort and best practice approaches are freely available.[11]

What if... we could use controlled experiments more often in marketing to agreed standards, giving a true picture of how marketing causes brand and sales uplifts?

What if... we could find ways to solve the design and execution challenges so that it's easier to test multiple factors and measure cross-media effects?

What if... we could apply controlled experiments to all marketing channels so that, as the use of experiments grows for online advertising, offline advertising is held to the same standards of evidence?

11 For example: 1) Estimating Ad Effectiveness using geo experiments in a time-based regression framework (Vaver et al 2017) https://ai.google/research/pubs/pub45950. 2) Prepare for the Unexpected: A Guide to Testing and Learning with Incrementality Measurement (Facebook, 2019)

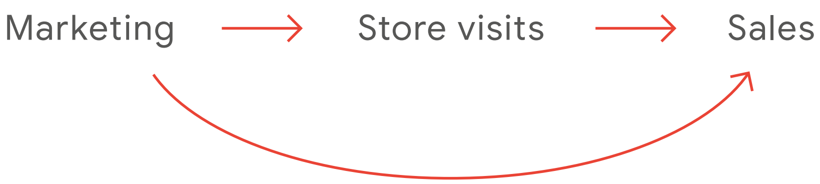
02 Better causal inference from observational data

It is clear that marketers can't measure everything with experiments. So, how can they get better at causal inference from observational data?

A 'causal revolution'

Researchers can start by defining a hypothesis of how marketing might work for the challenge at hand. This hypothesis can then be visualised as a 'causal diagram' illustrating how one factor affects another.

For example, imagine a retail brand trying to drive sales in-store. The marketer's causal diagram shows how marketing causes store visits, which in turn causes sales. It also shows how marketing causes sales directly, by influencing people who would have visited the store anyway.



Effectiveness experts currently approach this problem using 'Structural Equation Modelling' (SEM)[12] which uses a system of equations to connect the variables.

1. Sales = (a x Marketing) + (b x Store visits) + Other factors + Error
2. Store visits = (c x Marketing) + Other factors + Error

In this example, the first equation estimates the effect of marketing and store visits on sales. The second equation estimates the proportion of store visits which were actually driven by marketing.[13] The results can be used to reattribute credit for sales and estimate the direct and indirect effects of marketing via the 'mediating' variable of store visits.

12 SEM is also known as 'nested' or 'multi-stage' modelling. Alternative approaches exist, for example Bayesian Networks (see Bayesian Statistics and Marketing, Rossi et al)
13 Both adjust for other factors not shown in the diagram

However, the reality is more complicated than this. For example, season can have a big effect: marketing spend goes up in busy seasons, more people visit stores and those who do visit may buy more. So season in this case is a 'confounding' variable, which affects both the intervention (marketing) and the outcome (sales).

Typically season would be one of the variables adjusted for among the 'other factors' shown in the simple SEM described previously. But including confounding variables in a regression can lead to biased estimates.[14] So in this case the SEM, while intuitive, has some shortcomings.

Thanks to 30 years of research, an approach called the 'Structural Causal Model' (SCM) now offers one potential solution to this complexity.[15] The SCM formalises causal diagrams so they can be analysed by algorithms and provide answers to two key questions in the context of marketing effectiveness:

1. Can the causal effect of marketing on sales be accurately estimated from observed data only?
2. If so, what type of estimation method should be used and how should it adjust for various 'mediating' or 'confounding' or other types of variables on the path between marketing and sales?

Advocates of the SCM argue that it has led to major breakthroughs in social science and epidemiology: a 'causal revolution' that has enabled researchers to answer questions they had been wrestling with for decades.[16]

However, to date there has been little research on the application of causal diagrams and SCM algorithms to the SEM approaches used in marketing effectiveness.[17]

What if... researchers could work with marketers and effectiveness experts to explore the potential of the Structural Causal Model (SCM) to check the validity of Structural Equation Modelling (SEM) approaches?

What if... we could develop tools allowing marketers to build causal diagrams reflecting their understanding of how their brand works, to analyse these diagrams automatically, and to recommend optimal methods for estimating causal effects?

What if... this could spark a 'causal revolution' in the measurement of marketing effectiveness, improving accuracy across all methods?

14 (Skelly et al 2012) Assessing bias: the importance of considering confounding
15 (Pearl, 2018) The Seven Tools of Causal Inference with Reflections on Machine Learning.
16 (Pearl, MacKenzie, 2018) The Book of Why
17 (Google, 2018) Bias Correction for Paid Search in Media Mix Modeling is a rare example of such research https://ai.google/research/pubs/pub46861

03

Communicating uncertainty

So far, we have seen that common methods of measuring marketing effectiveness can't always estimate true causal effects. We've also suggested a need to embed understanding of causality into effectiveness — bringing about a 'causal revolution'.

Until then, we need to better understand and communicate the 'health warnings' that come with existing methods. And we should continually question how helpful it is to surface the level of uncertainty in our estimates to the different stakeholders who make marketing decisions.

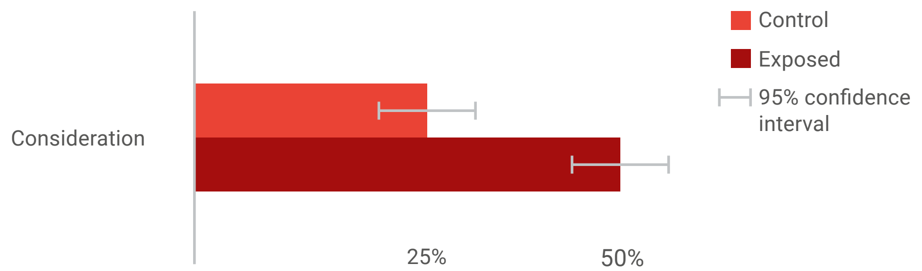
A 'known' unknown: confidence intervals and statistical significance

It is always possible that the element of chance can contribute to a small difference between test and control groups during a particular experiment. Because of this, it's important to find ways to measure uncertainty, even in randomised controlled experiments where estimates of causal effects are accurate.

One common measure of uncertainty is the 'confidence interval'[18] - a range of values around the central estimate.

Consider an example. Here we can see the 'lift' in consideration between control users and users exposed to YouTube ads in a fictional but realistic study. The consideration estimate for control users is 25% and the 95% confidence interval is between 20% and 30%. For exposed users the estimate is 50% and the confidence interval between 45% and 55%.

CONFIDENCE INTERVALS - BRAND LIFT EXAMPLE



Source: Illustrative example of Brand Lift results, YouTube

Marketers looking for a single, straightforward estimate can clearly see the data they need (50% minus 25% is a lift of 25 percentage points). Those who wish to use the estimates for planning purposes can use the uncertainty to simulate alternative scenarios.

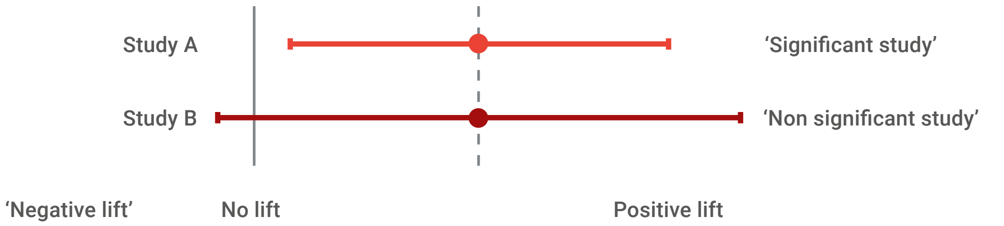
While not all marketers will want to trawl through the detailed statistics, research shows that the right graphical representation can make it easier to interpret the data.[19] By visualising these concepts in marketing effectiveness tools, we can help the more technical stakeholders understand the quality of evidence in front of them.

18 Or 'credible interval' in Bayesian statistics
19 https://ig.ft.com/science-of-charts/ (subscription required)

Related to confidence intervals is the idea of whether a difference, or 'lift', is 'significant' or not. This is often treated as a binary concept, simple for anyone to understand.

In the example below comparing two studies, the 95% confidence interval of the second study crosses zero, so the result is considered not to be significant based on that threshold.

ESTIMATES OF BRAND LIFT FROM TWO DIFFERENT STUDIES



Source: Illustrative example of Brand Lift results, YouTube

However, the reality is more nuanced and there is a movement from some scientists to take this into account.[20] In the example above, notice how the point estimate — the average observed effect — is the same in both studies, so they are not in conflict. Also the confidence interval of the second study only overlaps zero by a small margin, so perhaps we should take a more flexible view.

Last but not least, there are situations (e.g. multi-stage econometric models) where confidence intervals and significance are challenging to calculate at all. So the results may not be valid or make sense, even if the model can be shown to predict outcomes well.

Overall, it is clear that we should take care as we communicate the confidence intervals around marketing effectiveness estimates, to different stakeholders.

An 'unknown' unknown: bias

Let's zoom out for a moment to look at uncertainty in the wider context of the causal revolution.

Here, the extent to which causal inference methods can be used to measure causal effects is captured in the idea of 'bias'. Bias can be understood as how much a method overestimates or underestimates the true effect. Unfortunately, it is often impossible to measure: it's an 'unknown' unknown.

The Structural Causal Model[21] can use causal diagrams to help to identify when and where biases are likely to occur — and even whether they are likely to lead to overestimates or underestimates. If we can find a way to present these findings to marketers without going into the technical details, we have a real opportunity to improve understanding.

But how might the idea of measuring and communicating uncertainty land in the real world of marketing?

20 Scientists Rise Up Against Statistical Significance: https://www.nature.com/articles/d41586-019-00857-9
21 See previous section

How marketers might react to uncertain estimates

Marketers who relish statistics and causal inference might well embrace increased communication of uncertainty. If we find new ways to highlight and visualise the accuracy of our effectiveness measurements, this could inspire a fresh approach to the decision-making process.

However, many marketers will find that as more uncertainty is brought to the surface, they are overwhelmed with information and begin to doubt their decisions. Useful measurement methods which drive significant parts of the marketing economy risk being written off if marketers adopt a more 'purist' view of measurement.

For example, consider marketing mix modelling.[22] This method takes two to three years of weekly aggregated sales data and unpicks marketing effects from other factors, using the resulting model to predict sales under different investment scenarios. Statisticians can point to many elements of uncertainty in such models,[23] with some measurement experts suggesting they should not be used at all for certain purposes.[24] Yet modellers can ensure good predictive power through holdout testing, and many advertisers still find MMMs helpful in making decisions — surely their use would end quickly if their predictions were always wrong?

It is likely that some audiences (for example, measurement experts) may benefit from exposure to numerical error margins in order to understand the quality of evidence in front of them, while others (for example, planners, buyers) may not, as they are already faced with too many sources of information. Instead, they might rely on advice from the experts or even simple colour-coding (for example, red, amber, green in order of increasing quality).

For this reason, it seems that surfacing all the error margins present in measurement methods to all audiences is not the answer. Any initiative dedicated to analysing uncertainty needs careful application in order to both help marketers understand results clearly and allow them to make decisions confidently.

There are several opportunities here for research and development:

What if... we could build measurement tools that automatically identify uncertainty in effectiveness estimates — even when it can't be quantified?

What if... we could better understand how marketers process uncertainty in effectiveness estimates and how this affects their decision-making?

What if... we could build measurement tools that take this into account and highlight uncertainty in a way that is useful to decision-making, for the right people and at the right time?

22 MMM, econometrics
23 Challenges and Opportunities in Media Mix Modelling (Chan and Perry 2017)
24 Forecasting Advertising and Media Effects on Sales: Econometrics and Alternatives (Ehrenberg Bass 2018)

02

Measuring the long term, today

Using immediately available granular data to estimate how marketing might deliver returns in the future

In the business world, companies are valued, invested in, and bought or sold on the basis of their potential future value — and the current drivers of that value.

Any investor must weigh up the history and current position of the company, predicting how actions taken now might increase its future growth. These decisions can come with a high cost in terms of short-term revenue or upheaval to the business.

Marketers must walk the same tightrope: investing in the future of their brands will result in a sustainable business, but driving profit now is also important. Businesses that work out how to weigh short-term vs. long-term gain will thrive.

Today's advertisers can struggle to balance brand-building with sales activation. Some observers believe that the shortening tenures of marketers have resulted in more focus on quarterly or annual targets.[25] Others argue that less brand investment results in long-term harm.[26]

In markets like the UK,[27] online advertising – with its real-time performance metrics – now accounts for more than half of overall advertising budgets. Understandably, some have blamed the rise of online advertising for a focus on short-term profit. However, simply observing two things growing in tandem does not prove that one has caused the other. Correlation does not always mean causation.

In fact, it is just as possible to build brand equity online as it is to drive the next sale. In addition, many 'traditional' media (for example, outdoor, TV) are now delivered digitally, so the previous distinctions are no longer accurate.

Further debate about the respective merits of 'digital' and 'traditional' media is unlikely to be helpful to marketers. Instead, they need better measurement of the long term, today. They need metrics and methods that can estimate long-term effects accurately and quickly.

There are two key areas to explore:

01 Measuring long-term effects

02 Projecting the long-term value of marketing

25 Mounting Risks to Marketing Effectiveness: Enders Analysis 2017
26 Les Binet and Peter Field for the IPA Media in Focus, 2017
27 https://www.iabuk.com/news-article/aawarc-uk-advertising-spend-reaches-ps236bn

01 Measuring long-term effects

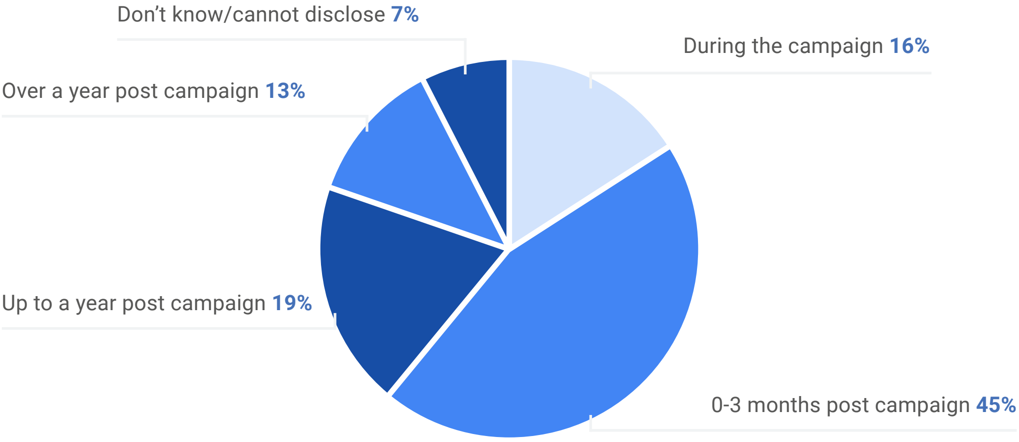
To understand how marketing drives long-term effects, we must understand both what long-term means, and how outcomes are measured.

What does 'long-term' mean?

The definition of long-term in marketing can vary from one sector to the next. Milk has a purchase cycle of a couple of days, toothpaste several weeks, and cars or mortgages several years. Marketers also follow business cycles which, again, can differ from company to company. To one that makes plans annually, the long term could be five years from now. But if a company makes plans quarterly, the long term might be just six months away.

Recent research in the UK showed that 61% of marketers measure success solely during the campaign or in the first three months afterwards and a mere 13% measure success 'over a year post campaign'.[28]

TIMEFRAME TO MEASURE MARKETING SUCCESS



Source: ISBA, Demystifying the Role of Attribution member survey, October 2018

It could be that, broadly, the long term is considered to be anything longer than three months. But given that many adults remember marketing they saw or heard when they were children, it's clear that marketing effects can last far longer than this.

Whether effects last three months or three decades, we need to be able to measure them, or at least project what they might be with a greater degree of certainty.

28 ISBA: Demystifying the Role of Attribution (2018)

Measuring long-term effects in aggregate

In many cases, marketers don't have access to customer-level data and must look at aggregated long-term changes in outcomes, from brand awareness through to sales and profit. They must then estimate the effects of marketing in driving those changes, adjusting for other relevant factors.

This is typically done via long-term marketing mix modelling, a process that encompasses a number of methods developed by MMM providers.

Method	How it works	Challenges
Time-varying parameters for marketing effects	Allowing the 'weights' that describe marketing effects in the model to vary over time to reflect changes in performance	Doesn't actually measure long-term effects, just measures changes in short-term performance over time
Long-term adstock	Using adstock rates which carry over marketing levels (TVR, impressions etc) into future periods and estimating their effect on sales over months and years (rather than weeks in regular models)	Effects are only detectable for the biggest channels and campaigns, where marketing drives a large proportion of overall sales
'Cleaning the base' (decomposing sales into short and long-term components)	Using techniques such as UCM, VAR or BSTS[29] to split outcomes into short-term spikes and long-term trends. Separately, modelling the effect of marketing on each time series.	Typically requires more data than regular MMM (four years' weekly data or more to properly identify seasonal effects rather than two to three years)
'Multi-stage models' (integrating brand and intermediate outcomes before sales)	Building a system of equations which looks at direct and indirect effects of marketing via other variables (for example, Marketing > Brand Interest > Web Visits > Sales)	Requires good data on each outcome. Still hard to detect effect of smaller channels or campaigns across all models and over longer 'lag' periods

The final two methods are well-established, and could be considered the best choices currently available to measure the long-term, aggregate effects of marketing. But the additional data, analysis and additional cost required mean that they are not yet in mainstream use. More importantly, they only deliver results after several years — too late for marketers who need to make decisions in days or weeks.

29 UCM = Unobserved Component Modelling, VAR = Vector Autoregression, BSTS = Bayesian Structural Time Series

Measuring long-term effects with customer data

With access to customer data, measurement of long-term outcomes is arguably more accurate. A marketer might be able to see what a particular customer has bought and when, and how much revenue or profit resulted over different time periods. This is often called the 'lifetime lalue' or LTV of that customer.

This data also allows a brand to understand which sales come from existing customers (loyalty) or from new customers (increased market penetration), a distinction that can be important in understanding brand performance.

But the challenge of measuring the marketing effects remains. Should credit be divided between any channels or campaigns the customer was exposed to before their first purchase? Should we include all interactions since then? And should the customer's existing perception of the product or service also take a share?

Effectiveness experts will have their own approaches which can flex to accommodate the needs of the business and the available data. But it is clear that, while having customer data is useful, it does not fully solve the problem of long-term measurement.

02 Projecting the long-term effects of marketing, now

If MMM and LTV approaches can estimate the long-term effects of marketing, but marketers need quick results, what can they do? First, they need rules to help project these results back into the short term. Second, they need to know which outcome metrics best provide leading indicators for long-term success.

Projecting long-term results back into the short term

The results of long-term MMM break down sales due to marketing into short-term and long-term sales, usually calculating a long-term 'multiplier':

$$\text{Long-term multiplier} = \frac{\text{Long-term sales}}{\text{Short-term sales}}$$

It's possible to take the long-term multiplier from one study and apply it in another context, but is it the right approach? Consider the same brand, measuring short-term results in 2019, and scaling them up using a long-term multiplier from a study covering 2013-2018. Performance may vary in 2019 and the true multiplier may be different (when it is finally measured in 2020 or later) meaning the new study is subject to error.

Applying such multipliers out of context, across different brands, channels or sectors may give rise to even more inaccuracy. And for multipliers to be checked against reality, marketers will need to wait.

With access to customer data, projecting back into the short-term is a more sophisticated process. It is possible to build models based on immediately available signals that project the expected LTV of different potential customers and make automatic adjustments: targeting them with customised marketing, bidding more to reach them, or both.

These predictions will also be subject to error. But the actual behaviour of customers can be studied in real time, so that the accuracy of models can be constantly tested and updated.

For example, imagine that in March a segment of customers is predicted to be lower value, but in April a proportion of them actually go on to purchase several times. From that point, when offered the chance to reach similar but new customers, the marketer's buying platform can automatically bid more to secure the opportunity to reach them.

Overall, three opportunities emerge:

What if... effectiveness experts could agree best-practice for long-term MMM on aggregated data, in particular enabling it to measure the effects of smaller campaigns or channels?

What if... they could come up with rules about when long-term multipliers should and should not be used to project long-term effects?

What if... developments in machine learning could enable more marketers to predict the lifetime value of potential customers, allowing them to pay more to reach them and to continually update predictions based on actual behaviour?

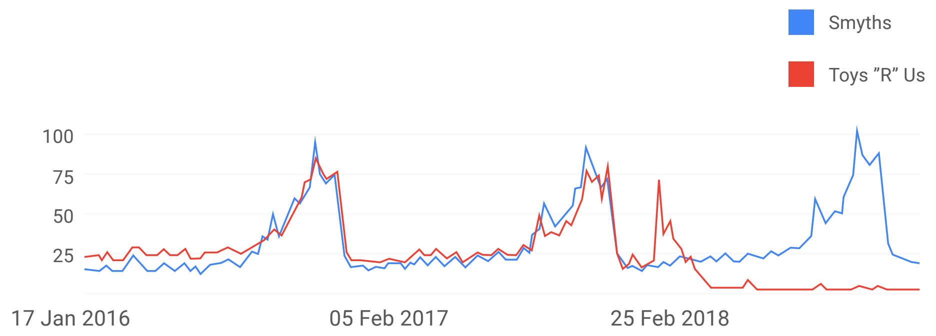
Using online behaviours as a leading indicator of long-term outcomes

Brand tracking surveys report at weekly or monthly intervals based on small samples. In contrast, online behavioural signals are continuous and based on much larger samples or the whole market. If researchers can ascertain which behavioural signals give an accurate representation of brand metrics, these could be decomposed to analyse how they are driven by marketing channels and campaigns.

For instance, in some circumstances search query volume has been shown to be a good indicator of brand 'salience', sometimes known as top-of-mind awareness — the ability of a consumer to spontaneously recall the brand when asked to think about the category.[30]

Consider retail brand Smyths Toys, which grew in the UK from 2016 to 2019, while competitor Toys 'R' Us faltered in 2017 and went bust in early 2018. The search query volume over the period reflects this, allowing for the obvious seasonal patterns of interest in the toy category. In 2016, Toys "R" Us' volume exceeded Smyths'; in 2017 it lagged behind; and in 2018 it fell to a low base after a brief spike resulting from headlines covering the failure of the business.[31]

INDEXED SEARCH QUERY VOLUME: SMYTHS TOYS AND TOYS 'R' US, UK 2016-19



This isn't to say that we should discard brand tracking altogether and replace it with search query volume. Rather, it raises the question: which brand metrics (for example, salience, consideration, purchase intent) are reflected in which search terms or topics?

In 2016, researchers ran a study among 1,511 users who gave permission to link their search behaviour with their responses to brand tracking questions about the smartphone and automotive categories.[32] The researchers found that users who are actively shopping in a category were more likely to search for brands in that category. As users moved from being aware of a brand to intending to purchase it, they were increasingly more likely to search for that brand, with the greatest gains as customers went from recognition to familiarity and from familiarity to consideration.

30 Digital Behaviour Analytics http://www.millwardbrown.com/mb-global/what-we-do/media-digital/digital-behavior-analytics
31 Source: Google Trends, February 2019
32 Brand Attitudes and Search Engine Queries https://ai.google/research/pubs/pub45740

However, they also found that users who owned a particular automotive or smartphone brand were much more likely to search for that brand, even when they were not 'in market' to buy a new one. This suggests that a substantial volume of brand search in these categories was unrelated to shopping or product search. Clearly, more research is required to understand how different sets of search query volume (keywords, terms, categories) might be a proxy for (or leading indicator of) important brand metrics.

More broadly, some effectiveness experts have explored the possibility that 'share of search' could be a leading indicator for 'share of market'. The idea is that the proportion of brand search that is for your brand may predict your share of sales in a category.

And there are various other online behaviours in addition to search query volume that could be used at scale to understand brand health. Some measurement vendors, for example, build a composite metric of brand interest from various search and social signals.

Overall, measurement of online behaviour presents some exciting opportunities:

What if... researchers could establish clear relationships between specific online behaviours such as search queries and brand metrics usually captured in surveys (for example, recall or purchase intent)?

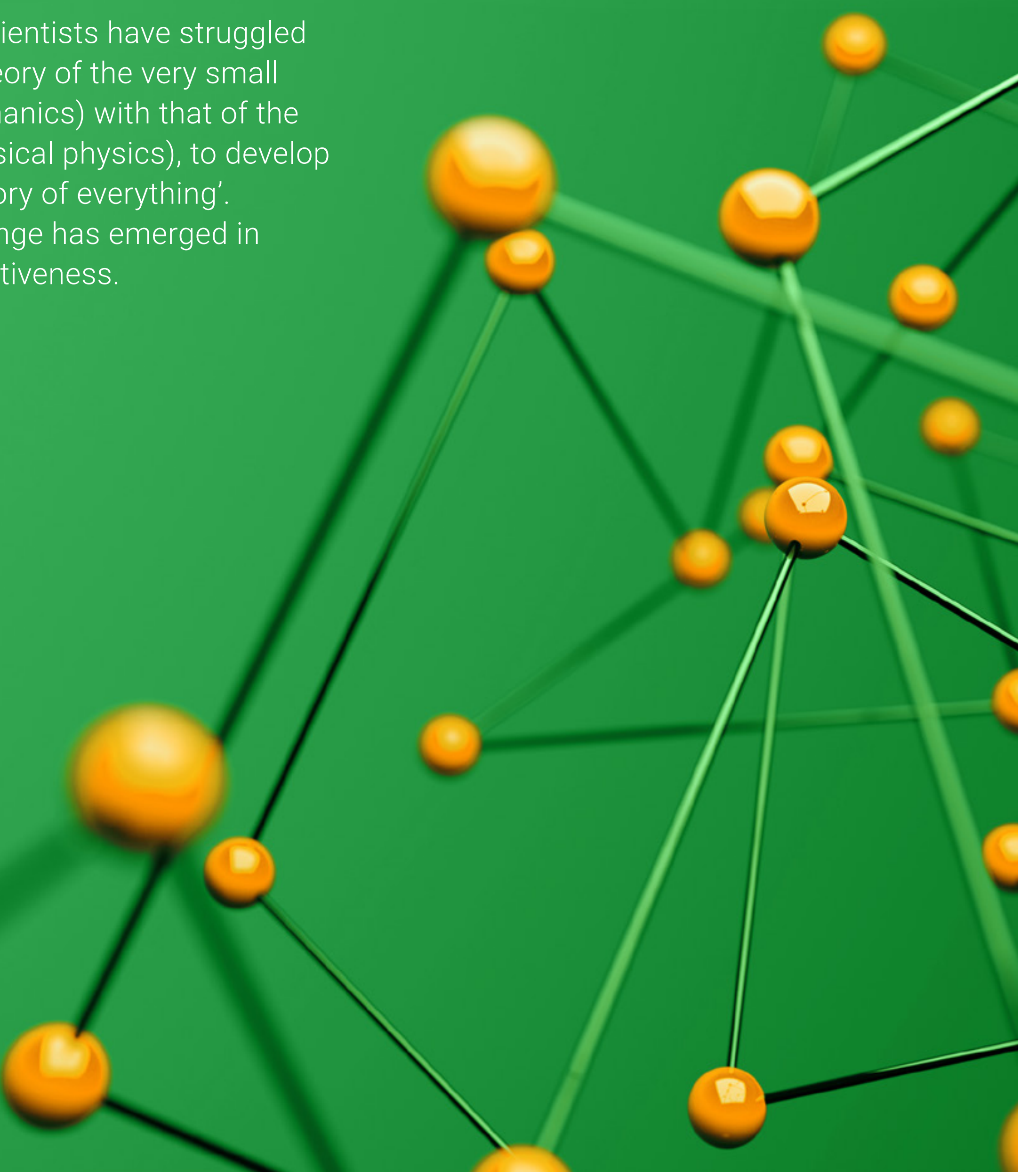
What if... effectiveness experts could reliably use this data as a proxy or leading indicator for brand health/market share, either alongside or instead of traditional survey-based research?

What if... effectiveness experts could accurately estimate the impact of increases in brand metrics on long-term sales, and project this back into the short term, so that they could assign a potential future value to current marketing activity?

03

Unified methods: a theory of everything

For decades, scientists have struggled to marry the theory of the very small (quantum mechanics) with that of the very large (classical physics), to develop a so-called 'theory of everything'. A similar challenge has emerged in marketing effectiveness.



Consumer-level models like digital attribution measure at the level of the very small. Granular units such as users, cookies or devices, advertising impressions, clicks or views are recorded, together with the target outcome (whether they resulted in a web visit or a conversion). By analysing thousands of these 'paths to purchase', attribution models can estimate the effect of single ads and formats in particular channels.

Aggregate-level models like marketing mix modelling/econometrics measure the very large. They take weekly sales data and compare it with weekly data on all the measurable factors that might affect sales, including both digital and traditional marketing. Using two to three years of data, MMM can estimate the effect of whole campaigns or channels over a longer period of time.

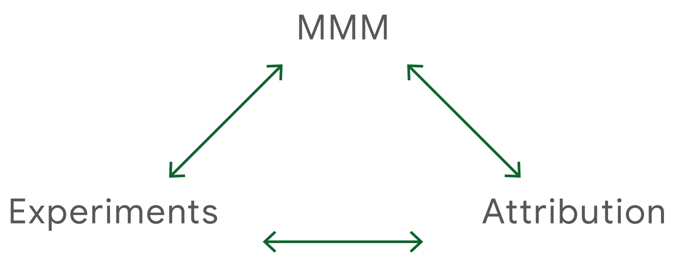
As some countries are already past the point where 50% of ad spend is delivered online[33] and as e-commerce increases as a proportion of all sales, the use of consumer-level and aggregate-level models is also starting to reach parity.[34] So it's no surprise that marketers are asking for their own theory of everything: a way of bringing together the very small with the very large to get a complete, rounded view of marketing effectiveness.

> Today's increasingly complex media environment means once-reliable marketing performance measurement techniques, such as marketing mix and attribution models, fail to properly credit marketing tactics with a customer action or influence. B2C marketers must embrace a new measurement standard — unified measurement — that will measure marketing's entire value and identify the best ways to optimize customer interactions.
>
> Source: Forrester, 2018[35]

33 https://www.iabuk.com/news-article/aawarc-uk-advertising-spend-reaches-ps236bn
34 ISBA: Demystifying the role of attribution (2018) — 48% use MMM, 43% use digital attribution
35 Customer-Obsessed Marketing Demands Unified Measurement (Forrester, 2018)

In our work with advertisers, agencies and measurement providers we are frequently asked about combining the major effectiveness measurement methods. Three methods, and the potential links between them, come up most often.

1. Marketing mix modelling (MMM)
2. Digital attribution
3. Experiments



Typically, MMM is used for big-picture budget setting between large channels or campaigns, alongside other marketing tactics like price and promotion. Digital attribution is used for day-to-day management of online advertising, and experiments are used to test specific hypotheses that are important to the advertiser.

Because they are used for different purposes, each of these methods may well be commissioned by different teams within an advertiser's organisation, and delivered by different third-party providers. In this paper we will not discuss the strategies needed to break down such silos. Where blending methods becomes relevant, we will assume that marketing teams can work together across providers, or with one provider.

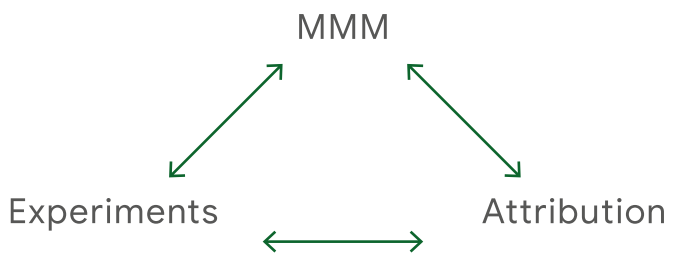
In this section we'll explore potential combinations of methods through the lens of three opportunity areas:

01 Blending one method with another

02 Blending multiple methods at once

03 The role of structured expert judgement

01 Blending one method with another

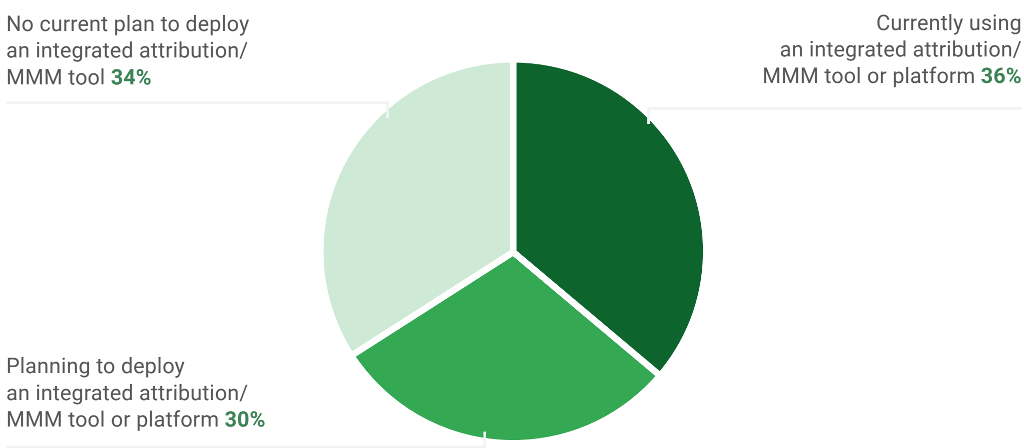


There are three sides to our triangle, so we'll explore each side in turn to see what promise it holds.

Blending MMM and digital attribution

In some of the leading markets for online advertising and e-commerce, a significant proportion of advertisers claim that they are already blending MMM and attribution. As early as 2015, a survey of marketers in the US and the UK found that one third integrated these methods, with another third planning to do so.[36]

USE OF MMM AND DIGITAL ATTRIBUTION, 2015

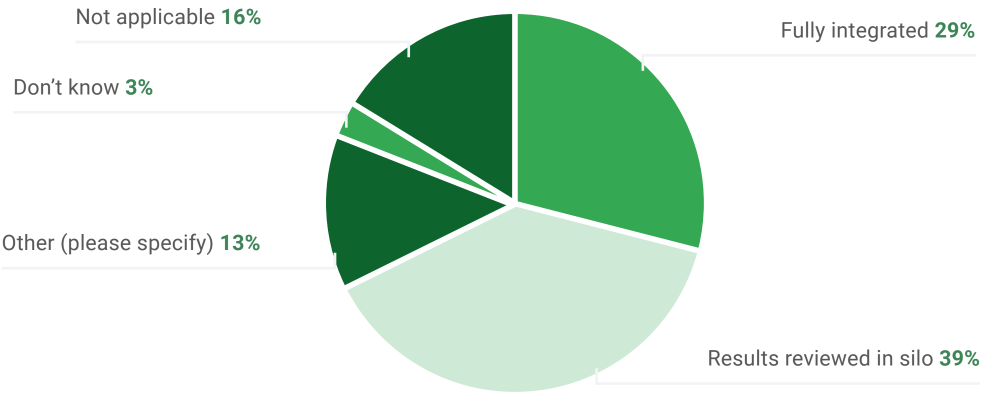


Base: 151 US and UK decision-makers or influencers responsible for advertising budgets
Source: A commissioned study conducted by Forrester on behalf of Google, August 2015

36 Forrester/Google Study 2015

But fast-forward three years and little has changed, at least in the UK. A recent survey by ISBA found almost the same result.[37]

USE OF MMM AND DIGITAL ATTRIBUTION, 2018



Source: ISBA, Demystifying the Role of Attribution Member Survey, October 2018

During 2018 we interviewed around 40 marketers at large brands in the UK about their effectiveness practices, to understand how we can better support their measurement work. While this wasn't a survey, examples of 'fully integrated' MMM and digital attribution were cited in fewer than a third of these conversations.

When integration was mentioned, it was more the case that the advanced marketers now had effectiveness leaders whose role it was to understand marketing effectiveness as a whole, and consider results from different methods.[38] These leaders were typically very aware of the pros and cons of MMM and digital attribution and were often blending insight from them, rather than integrating them at a technical level.

It is likely that this slow progress has been caused in part by some of the new challenges faced by digital attribution. Rising user expectations around privacy are changing the ways marketers can reach and measure audiences online. As the digital ecosystem has evolved, measurement providers have had to prioritise developing new attribution solutions over integrating them with other methods.

Despite these challenges, digital attribution remains an important method in the effectiveness toolkit, and blending it with MMM is clearly of interest to two-thirds of advertisers in markets like the US and UK.

37 Incorporated Society of British Advertisers
38 Culture First, IPA 2017, pages 8-9 https://effworks.co.uk/wp-content/uploads/2017/10/Culture-First-Final.pdf

Matching scope and granularity

MMM and digital attribution can estimate the same key metrics of sales contribution and return on ad spend, but it's important to note that their default scope and granularity tend to be very different.

Rank	MMM	Digital Attribution
Sales scope	Online and offline	Online only
Advertising scope	Whole year or total campaign	Individual channel executions
Channel/ Format	High-level online and offline channels, for example 'TV', 'online video', or 'total YouTube'	Detailed online formats, for example, 'YouTube Bumper ads'
Targeting	All targeting types used by the channel or campaign measured	Particular targeting types, for example, 'Sports Interest audience'
Time period reported on	Usually whole year, weekly reporting not recommended	By campaign, weekly breakdown could be made to match MMM
Lagged sales included	'Adstock' is modelled to capture lagged sales during and after the end of each campaign	A 'look-back window' can capture lagged sales during and after the end of the campaign

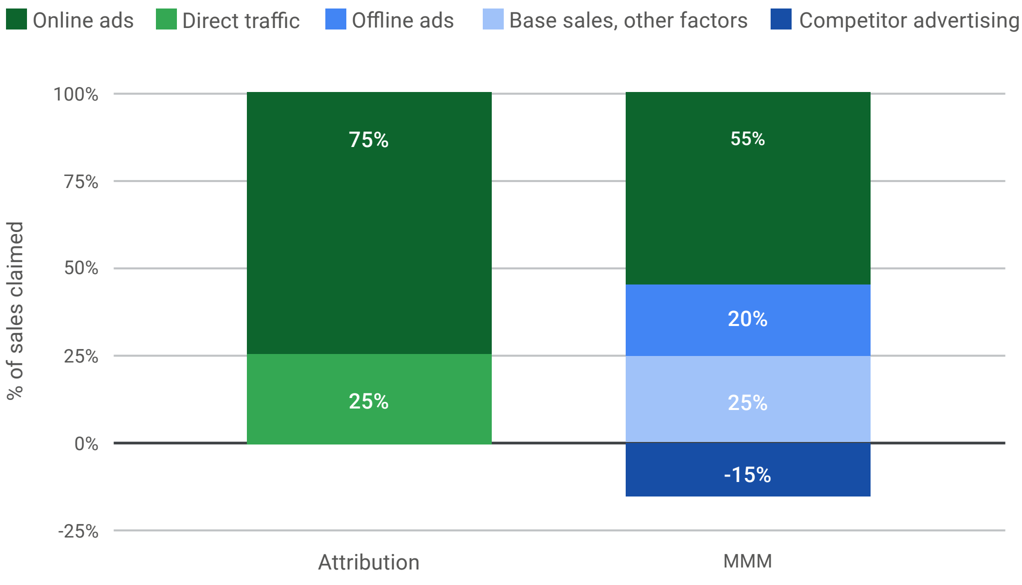
Before comparing or blending results, we need to think about which results to consider, aiming to compare like-for-like where possible. When we can't do this, we should clearly state our assumptions and beware of possible differences.

Comparing results

There are a number of ways in which we might expect the results of MMM and digital attribution to be different. Imagine a fictional sports retail brand called *Eagle Sports*, which runs online and offline marketing to drive online and offline sales. Their measurement provider *Panda Analytics* is commissioned to develop both MMM and digital attribution models in parallel.

Let's say that *Eagle Sports* is particularly interested in measuring the effectiveness of online ads around the 2018 Soccer World Cup, and that *Panda Analytics* provides these decompositions of online sales over the period:

DECOMPOSITION OF ONLINE SALES FOR 'EAGLE SPORTS' DURING 2018 WORLD CUP CAMPAIGNS



Here, attribution divides the credit for sales between online ads and direct traffic, where MMM also credits base sales and offline channels. For this reason, one criticism of attribution is that it overstates the effect of online channels.

However, note that credit allocated to different factors in MMM isn't always positive: some factors, like competitor spend, act as a drag on sales.

So, it is not necessarily true that attribution always overstates the effect of the online channels. In fact, these channels might be wrongly taking credit for both rises *and* falls in sales that were actually caused by other factors.

Integration approaches

Once we make results comparable, allowing us to see how different factors take credit for sales in attribution and MMM, how might we combine insights from the two methods?

At the moment, advertisers, agencies and measurement companies use various approaches:

Route	Method	Opportunities	Challenges
1. Use digital attribution to improve MMM	Use digital attribution to add detail to MMM	Break down 'buckets' in MMM: for example, if MMM measures 100 sales due to 'online video', break this down into 'YouTube' and 'other online video' according to the proportions observed in digital attribution	• Assumes that digital attribution assigns credit correctly • Assumes that the effect of a channel or campaign is the same offline as it is online • Unclear how uncertainty in each estimate should be reflected in the combined estimate
	Use digital attribution to constrain MMM or to help it learn the effect size	Help MMM better reflect effect sizes seen via digital attribution	
2.Use MMM to improve digital attribution	Use MMM results to redistribute credit in the attribution model	Give credit to factors not normally considered in digital attribution, for example, offline advertising, price and promotion, competitor marketing, even the weather or the economy	• Assumes that MMM assigns credit correctly • Assumes that the effect of a factor is the same online as it is offline • Unclear how uncertainty in each estimate should be reflected in the combined estimate
	Use MMM data to add new touchpoints in the attribution model		
3.Keep MMM and digital attribution separate	Don't combine results; instead use them to form a 'range' where the true answer may lie	Save the time and effort needed to integrate results, trust marketers' experience to use them to make decisions	• Results often don't agree which can be confusing, or can undermine faith in both methods

However, there is little public information on how such integration is achieved, so more open-source research is required.

What if... measurement providers could share how they are integrating MMM and digital attribution without revealing all their intellectual property, allowing for a more informed debate about pros, cons and best practice?

What if... studies could be calibrated against gold-standard results from experiments, to ascertain whether combining information from MMM and digital attribution results in more accurate estimates than when they remain separate?

What if... we could develop a set of rules to help marketers decide whether, and when, the additional insight might be worth the extra effort and cost?

Blending MMM and experiments

Blending MMM with randomised controlled experiments could be useful to advertisers in most sectors. Unlike digital attribution, this process does not require user data, as MMM uses aggregated weekly data instead. Experiments can be randomised using units of interest, like stores, regions or brands.

These two methods should also be a good match in helping to correct their respective flaws:

- MMM measures correlations between predictors and sales, and is not always good at estimating causal effects; but good experiments are the gold standard in causal inference.
- Experiments can only measure the effects of one or two things at the same time, under particular market conditions; but MMM can test all relevant variables over several years of real market conditions.
- Experiments require forward planning and control over exactly how marketing is delivered; but MMM can use observational data viewed retrospectively.
- Experiments are typically applied to measure short-term marketing impact; but MMM can be used to estimate long-term effects.

Comparability is an issue, as it is when blending any two methods. Again, experiments are usually much more granular than MMM. They often measure specific ad executions across one campaign, in particular formats within one channel, whereas MMM measures a whole channel or total campaign and reports over a year or more.

However, these methods do estimate based on the same metrics of 'sales attributed to marketing' and the resulting Return On Ad Spend (ROAS). So comparisons can be made, as long as this is done with appropriate care.

Integration approaches

When and how could these two methods be blended? These are some of the options:

Route	Method	Opportunities	Challenges
1. [Reactive] Use experiments to improve MMM	Use experiments to constrain MMM or to help it learn the effect size	Help MMM better reflect the effect size seen in the experiment, which should be more accurate	Unclear how uncertainty in each estimate should be reflected in the combined estimate[39]
2. [Reactive] Use MMM to explain experiment results	Use MMM as context to understand and explain experiment results	Understand whether an experiment ruled out other factors correctly, for example, was a promotion running that wasn't included in the design?	Few challenges, other than the fact that MMM and experiments often take place in silo
3. [Proactive] Design MMM and experiments together	Run regional-level MMM studies backed up by a programme of randomised, controlled geo experiments using the same regions	Experiments can measure accurate effect sizes for marketing interventions and create useful variation in the data, which makes it easier for MMM to get a read on the same marketing interventions	Significant work required to align all data

The proactive approach presents some challenges. Planning and delivering a regular programme of experiments while aligning them with the MMM would require precision marketing, good data, close collaboration and additional work. But the rewards could be a more accurate picture of marketing effectiveness.

What if… we could develop clear standards around calibrating MMMs by using relevant results from experiments?

What if… we could show how data collected for MMMs could explain experiment results by identifying factors missed by the design?

What if… we could systematically design and deliver a programme of geo-experiments and regular MMM updates together, using the same regions?

39 Clearer with Bayesian MMM using credible intervals, but still needs explanation

Blending digital attribution and experiments

The comparability of digital attribution to experiments depends on whether the experiments are run at user-level or geo-level.

Digital attribution	Method	User experiment	Geo experiment
Scope	Many campaign executions or particular ad formats	Covers one or two specific campaign executions or particular ad formats	
Geography	Usually national level but with enough conversions results may be obtained at sub-regional level	Usually national level	TV region, digital marketing area or store level
Data period	Daily data over the course of a campaign	Varies but often daily data over several weeks	Varies but often weekly or daily data over 6-12 weeks
Reporting period	Usually whole campaign + some look-back window to allow for lagged sales	Usually whole campaign (sometimes + further period to allow for lagged sales)	Results from the period of the experiment + some further period to allow for lagged sales
Metric reported for 'effect size'	Sales attributed to marketing intervention based on comparison of its role in converting and non-converting paths	Uplift in sales between test and control users	Uplift in sales between test and control geos

The scope of digital attribution is more likely to match with that of experiments, with fewer adjustments required than when blending experiments with MMM. But again, the nature of the data (observed vs. controlled) is different for each. And we might reasonably assume that properly designed and executed experiments will result in better estimates of causal effect than those produced by digital attribution.

So, how can digital attribution learn from experiments? Here we can see similar approaches to those used when blending MMM with experiments.

Integration approaches

Route	Method	Opportunities	Challenges
1. [Reactive] Use experiment results to improve digital attribution	Use experiment results to constrain digital attribution or to help it learn the effect size	Help the digital attribution better reflect the effect size seen in the experiment, which should be more accurate	Unclear how uncertainty in each estimate should be reflected in the combined estimate
2. [Proactive] Design digital attribution and experiments together	Design user-level (or even geo-level) experiments to align with digital attribution	Experiments can measure accurate effect sizes for marketing interventions and create useful variation in the data, which makes it easier for digital attribution to get a read on the same marketing interventions	Requires significant realignment of media delivery and measurement

Again, more research is required into these approaches to understand the implications for marketers in terms of accuracy of results.

What if… effectiveness experts could develop clear best practice around improving digital attribution by using experiment results?

What if… effectiveness and marketing leaders could systematically design and deliver a programme of experiments for their brands, to align with digital attribution and make results more accurate?

02

Blending multiple methods at once

When we looked at pairwise blending of MMM, experiments and digital attribution, we saw various ways in which each pair might be designed to work better together, and how their results might be integrated to provide more holistic insight.

But what about unified methods or processes that can take the different types of data collected and present or blend them together simultaneously? And could it be possible to integrate other information available to modern marketers: brand tracking, data from customer panels, and even expert opinion?

There will likely be many routes to achieving this, but promising avenues for exploration include:

1. A unified effectiveness process
2. Fusing data with hierarchical models
3. Agent-based models and simulated data

A unified effectiveness process

We have seen that aligning data from different sources and results from different measurement methods is not always straightforward. For many advertisers, there is much work to be done before they can begin considering integration.

A unified effectiveness process would be an approach rather than a single measurement method, bringing together and organising data and insights from multiple sources in order to simplify the process for marketing decision-makers.

Some marketers are doing this already. They have a clear vision of the data at their disposal and distinct roles for different measurement methods. In addition, they know how to present results at the right time to help stakeholders make decisions.

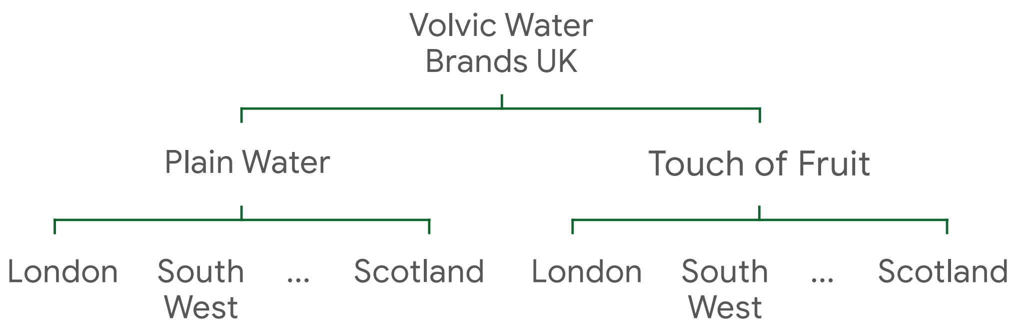
What if... effectiveness experts could identify the ideal process for gathering data and results from many different sources, presenting it to stakeholders in the right way and at the right time?

What if... measurement providers could build tools to ingest and present effectiveness results from many different sources without blending to produce a single answer, but still avoiding confusion or information overload?

Fusing user and aggregated data in a single model

Hierarchical models use data at different levels of granularity when estimating the effects of predictor variables (such as marketing interventions) on an outcome (for example, sales).

For example, imagine an MMM run across nine regions and five sub-brands, but reporting at national level. Rather than being constructed as 9 x 5 = 45 separate models, a hierarchical model would connect them all in a national model which can report in total, but also learn from the variations observed across the sub-brand and geo levels. The resulting picture can be seen as a 'hierarchy', hence the name.



The estimated effect size for a particular marketing intervention — in this case for YouTube on Plain Water sales in London — can learn from the effect size estimated for YouTube on Touch of Fruit in Scotland and vice versa. Across all these permutations, the overall estimate of YouTube effects should become more accurate.

This is just one example[40] for aggregated data using MMM.[41] But the same principal could be adapted to fuse aggregated (for example, weekly) and disaggregated (user-level or household-level) data into a unified model. Such models could also learn from panel data, brand tracking data, and the results of controlled experiments.

Media effects measured at user level could learn from the effects measured at geo level and total level. And each of these levels could also learn from other results pulled in from outside the model. The overall result could be a single, unified view of effectiveness.

In advanced markets such as the US, some measurement providers have already put these techniques into practice, selling 'fusion' solutions that claim to report very granular insights across online and offline marketing and sales channels. But because these solutions are their intellectual property, there is little published research on how they work. It is therefore unclear how such methods rate in terms of their ability to measure causal effects, or how the uncertainty in such estimates should be calculated or communicated.

40 https://www.thinkwithgoogle.com/intl/en-gb/success-stories/uk-success-stories/how-volvic-ipsos-mma-found-refreshing-digital-insights-putting-better-data-econometrics/
41 Geo-level Bayesian Hierarchical Media Mix Modeling (Google, 2017) https://ai.google/research/pubs/pub46000

What is clear is that unified measurement requires unified data, a challenge that should not be underestimated. In our experience, aligning and labelling user and aggregated datasets can result in significant work across advertiser and agency silos.

When it comes to such 'fusion' models, more research is required to explore the potential methods and the accuracy of results, and how these can be best communicated to non-technical marketers.

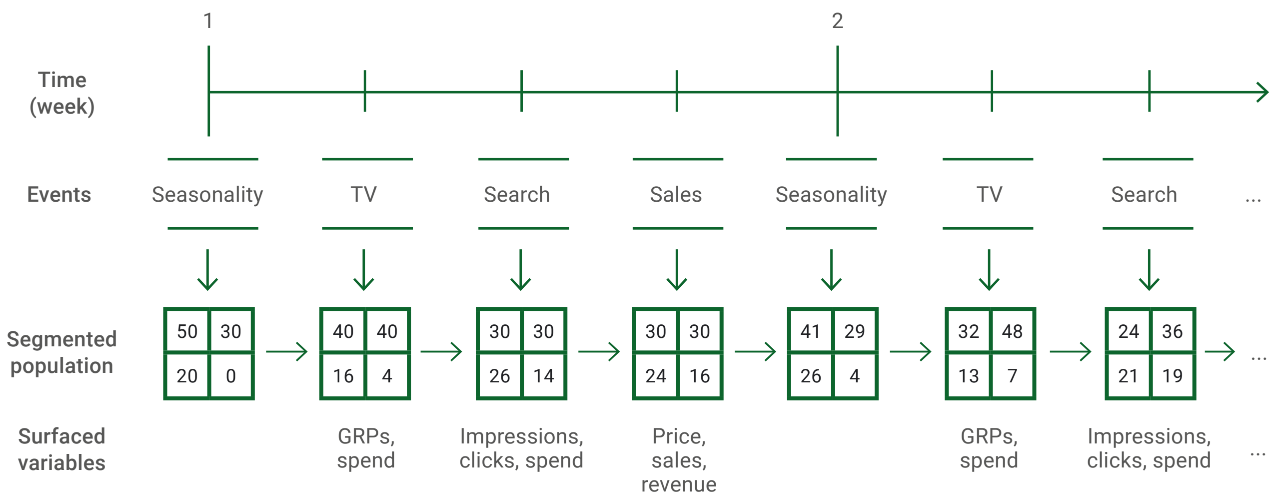
What if… measurement providers could share more details of how they are applying 'fusion' models to combine aggregated (for example, weekly, geo) and disaggregated (for instance, user-level or household-level) data?

What if… these models could be tested against the usual methods (and some 'ground truth' from experiments or simulations) to determine how accurate they are?

What if… effectiveness experts could develop best practice, and a set of rules to help marketers decide when the additional insight might be worth the extra effort and cost?

Agent-based models and simulated data

Agent-based modelling revolves around the idea of creating a population of imaginary customers (agents), allowing forces to act on them (such as marketing interventions) and observing how many of them 'purchase' a product over time. The result is simulated sales data generated purely by the forces you've specified. This can be done at both an individual customer level and a customer segment level.[42]



42 See Introduction to the Aggregate Marketing System Simulator (Google 2017) for an example of segment based ABM https://ai.google/research/pubs/pub45996

Because the agents or segments are not real, the only factors that act on them are those that you specify. It can be both hard and time-consuming to construct models that accurately reflect reality, but by comparing the results with real sales data and continually updating assumptions, each iteration brings you closer.

There are some unique benefits to this approach. Because you are specifying the factors, any effects you observe are true ‘causal effects’. Assuming your model is a good approximation of reality, with all the relevant factors included, you can adjust one intervention and observe how it genuinely affects sales.

For the same reason, you can use the data generated by a good agent-based model as a reference point — a source of ground truth — to evaluate other measurement methods. For example, the authors of a recent Google paper[43] used simulated data to compare the performance of national-level and geo-level hierarchical MMMs in recovering the ‘true’ marketing ROI specified by their simulation.

Again, more research is required here. It will be necessary to apply each of these methods to further marketing effectiveness challenges in order to explore their pros and cons.

What if... we could use tools like agent-based modelling (ABM) to build convincing simulations of sales for particular brands?

What if... we could use these to understand how marketing is actually working, rather than trying to decompose observational data?

What if... we could use the ABM as ground truth against which to judge various effectiveness measurement methods?

43 Geo-Level Bayesian Hierarchical Media Mix Modelling https://ai.google/research/pubs/pub46000

03 The role of structured expert judgement

Marketing effectiveness experts don't always believe that the additional insight from blending existing methods is worth the effort.

Some might even be tempted to write off methods such as MMM or digital attribution, feeling that their flaws can lead to poor decisions, for example when predicting the impact of future investments. These same marketers might accept the results of randomised controlled experiments, but it is not possible to use experiments in every instance.

So what should marketers do? Should they rely on incomplete evidence or rely on their gut feeling — the instinct they have honed over their experience of developing and delivering campaigns? But this instinct can be influenced by the cognitive biases that we all face in processing information and making decisions. In fact, research shows that marketers' intuitions about effectiveness are often wrong.[44]

A recent paper from the Ehrenberg-Bass Institute[45] takes the following view:

> Managers already know a lot about the effects that changes in the marketing mix are likely to have, from accumulated experience. Moreover, much of this knowledge is difficult or impossible to include in econometric models... So rather than using econometrics, we can use managerial knowledge in a structured way to aid forecasting. It is important it is structured, as unaided judgements are typically unreliable.
>
> Source: Ehrenberg-Bass Institute, 2018

44 Marketers' Intuitions about the Sales Effectiveness of Advertisements, Hartnett et al 2016
45 Forecasting advertising and media effects on sales: Econometrics and alternatives

Earlier we discussed how observational methods like MMM don't always measure causal effects well.

Rather than abandoning these methods altogether, perhaps a workable middle way is to: 1) use the latest research in causal inference to improve them, and 2) do a better job of surfacing the uncertainty in their estimates.

However, the idea of 'structured expert judgement' holds promise, potentially offering a way to combine the experience of multiple experts when evaluating or predicting the effectiveness of marketing. The Ehrenberg-Bass paper suggests the following approach:

- Prepare a set of media mix proposals together with the expected external conditions that will also affect sales
- Ask five or more experts with diverse but relevant knowledge to forecast the change in brand sales for each of the proposals
- Use the forecasts to build a model of how the experts believe changes in marketing spend will affect sales, adjusted for other factors
- Use this model to forecast the effect of new media mix proposals
- Compare predictions with reality after a period of time, to refine the model

So far, this approach does not appear to have been trialled on a real marketing challenge, so it is too early to judge its value. But it does present a logical conclusion to the thread running through this paper: that analytical evidence should be combined with the understanding of experienced marketers.

What if... we could develop methods of structured expert judgement that can be shown to produce accurate predictions of media performance?

What if... we could make it easy for marketers to source these opinions frequently and at scale, to make regular predictions which can be calibrated against real results?

What if... we could build platforms allowing marketers to view the results of structured expert judgement alongside the results of other methods (ranked for how they compare against the gold standard) to allow more intelligent, unified decision-making?

.

Conclusions

Some common themes emerge across each of the three Grand Challenges, which point towards the future of effectiveness measurement in marketing.

Striving for the best, but embracing the possible

Driven by ever-increasing scrutiny on the bottom line, marketers will increasingly have to aim towards the highest standard of evidence — randomised controlled experiments. But these will not always be possible, and in some use cases, such as prediction, the results may not be helpful.

In these instances, marketers will have to rely on other methods, making them as strong as possible and being honest with themselves about the accuracy of the results. Measuring effectiveness as well as we can, and being pragmatic about the results, is surely better than not measuring at all for fear of not reaching perfection.

Progress through transparency and clear communication

In all three challenges there are interesting new methods in the academic literature that could be applied to marketing effectiveness measurement – and in some cases, have been developed already. Some measurement providers have made progress in these areas but have not shared the details, for fear of compromising their intellectual property.

The industry needs to fill this gap with transparent, open-source research that is robust and clearly communicated so as to be accessible to non-technical but measurement-savvy marketers.

Human judgement: more critical than ever

Perhaps the most surprising theme to emerge is that solving these major challenges in effectiveness will still require expert human input. In fact, the way that we process and absorb analytical results from information streams will be critical — as will the way that we learn to live with uncertainty, rather than single numbers.

Rather than leaving measurement experts redundant, technology will help us with both of these challenges: and this has already been proven with randomised controlled experiments. A 2018 study by the Boston Consulting Group found that: "Companies using advanced technology with active human supervision can improve their campaign performance by up to 35%."[46]

Coming together and finding common ground

Our intention in proposing these challenges is not to set the terms of industry engagement, or to force the direction of conversation. Between different advertisers, media platforms and technology providers, there will be a wide range of opinions about the importance of each issue, and the perceived benefit of investing in solutions.

Instead, we hope the challenges will be seen for what they are — an invitation — one we're extending to passionate marketers, effectiveness experts and data scientists in every corner of the industry. If you have a view on any of these issues, please get in touch — we'd love to hear what you think.

46 BCG: Dividends of Digital Marketing Maturity, 2019

R E D W O O D_
Designed and produced by
Redwood London

Learn more with Google