



A captive

Techniques to measure audiences and advertising effectiveness are increasingly sophisticated, but the challenges of cross-platform viewing – and a lack of common standards – mean research is needed more than ever to get to the truth of what people think. By **Tim Phillips**

In a 1982 study designed to show how advertising affects sales behaviour, research company Information Resources Inc (IRI) identified locations in the US with a single cable provider and a grocery store that accounted for 90% of all food purchases. It then performed an experiment in which TV advertising was manipulated so that half the families saw one set of ads, and the other half saw a completely different set. The 3,000 households used ID cards to log their purchases.

The result would not have been a surprise to John Wanamaker, who famously claimed that an unknown 50% of his advertising budget was wasted: “In 360 tests in which the

only variable was advertising weight – the amount of television advertising to which consumers are exposed – increased advertising led to more sales only about half the time,” concluded Magid Abraham, at that time president of IRI and, later, a founder of comScore, in the *Harvard Business Review* in 1990. He warned advertisers not to assume that all advertising worked, or that the effects of their campaigns increased over time – both were rules of thumb that, in the absence of sophisticated measures of effectiveness, were widely believed at the time.

Today, few marketers would assume that advertising offered



audience

guaranteed positive return, or deny the existence of diminishing – or negative – returns to campaign saturation. Our ability to measure audiences and effectiveness has been transformed by online panels, passive measurement and emotional response tracking.

We might have more tools in 2017, but the complexity of the problem is exponentially greater. There's a cross-platform measurement challenge: how do we know how many times each person has been exposed to an ad and, in the digital environment – where, according to comScore, 74% of people use the internet on more than one platform – how do we understand what a

person watches, rather than just counting what a device displays?

There's also the question of who does the counting and auditing when publishers and platforms measure themselves according to different standards. For example, Facebook considers an ad viewable if it is on screen. Twitter charges marketers if the ad is 100% in view for three seconds. But GroupM bases its measurements on 100% of the ad being visible for 50% of its duration, with audio on – and by the Media Rating Council (MRC) standard, at least half a display ad must be visible for one second to count as being 'viewed'. It is two seconds for digital video.

Then there's the challenge of measuring the effect of the advertising on the business. Without some model of how viewers behave, or knowledge of what they think, we can't estimate effectiveness in moving key performance indicators (KPIs), or in creating a sales-based return on investment (ROI).

If some effects are easier to measure than others, do we end up creating advertisements that pander to those statistics? Should we try to attribute ROI? Can emotional measurement be more effective than rational measurement as a predictor of KPI or sales uplift – and for what type of advertising? ▶



RAISING EYEBROWS AND RAISING SUBSCRIPTIONS

In 2015, *The Economist* found that its circulation had begun to plateau. Its iconic 'white on red' billboard execution had, for many years, positioned the magazine as the go-to title for businesspeople. "The downside to this success was that many people outside of this group felt alienated from the brand," says Michael Brown, insight director at UM Insight.

The Economist engaged creative agency Proximity London, and media agency UM London, to design and execute a new programme, to reposition it with some eyebrow-raising copy: 'Bad back? Do it doggy style' was one example, about research on how to copulate in comfort. In its desire to position *The Economist* as a publisher for those beyond its

stereotypical readership, the team identified younger readers as a golden target. Its digital advertising would show online banner ads in relevant news contexts, with punchy copy, in an environment where young people habitually graze. If the campaign worked, prospective subscribers would click through a display ad, read a free article on the topic in *The Economist*, and have what research identified as an 'Economist Epiphany' – a key driver in converting new subscriptions.

The Economist used Mediabrands Marketing Sciences to deliver its digital advertising evaluation. "As media researchers, advertising performance evaluation is our bread and butter," says Brown,

who decided on a cookie/tag approach. Each creative execution within the initiative – from online display banners to ads in Facebook – had a simple piece of code (a 'tag') written into it. Research Now's online consumer panel uses cookies on its devices that watch for the tags.

The core KPIs were: relevance; intention to subscribe; recommendation and intention to share content; plus a social buzz analysis, with sentiment and context interpretation through Brandwatch. On one hand, *The Economist* wanted to drive subscriptions; on the other "it was especially critical that we measure globally and robustly, in order to make sure that some of the cheekier or more playful creative that we ran did not alienate either

our prospects or the existing reader base", says Mark Cripps, EVP of brand and digital marketing at *The Economist*.

He adds that the campaign, through regular measurement, was able to optimise the activity in real-time by changing creative and placement.

"To provide a sense of the programme's success, in its first year alone it was exposed to millions of prospects, drove more than 64,000 new subscriptions and delivered £57m in lifetime subscriber value," Cripps explains. This was 25 times the investment. "The approach has allowed millions of new readers to discover our content, and the way in which the programme was evaluated was a key factor in maximising this outcome."

▶ **A single source of truth**

One of the advantages of the simplicity of IRI's experiment was that all the data was first-party, collected at the respondent level. This is also one of the benefits of the ADimension suite of products that Research Now offers its clients – but at a far more nuanced level.

“We have two products: one for audience validation, the other to measure reach and frequency,” explains Liam Corcoran, Research Now vice-president of advertising and audience measurement. “But we also help our clients to understand which is the best-performing channel for their campaign. The real benefit when we measure is that we have a single-source panel.”

Research Now's knowledge of its panel is deeper than most. By using partners such as Avios, Nectar and Trainline to recruit panel members it knows a lot about the preferences and habits of its panel, not just their

“ A multiplatform, multichannel world creates questions about how difficult it is to measure ”

demographic data. On average, Research Now's panel has 350 attributes linked to a single person – for some, it has up to 600. Added to that, it can track the digital behaviour of panellists using first-party cookies; in the UK, it uses 300,000 device IDs on 110,000 panellists – so it doesn't just provide granular results on how a campaign is changing KPIs in a particular segment, but can also discover which devices are influential in doing it.

This has proved to be unexpectedly informative. In 2015, for example, Research Now did an experiment that might be considered a more-sophisticated descendent of cable

splitting. It tracked device use and change in KPIs for a controlled experiment, featuring 18m impressions of campaigns for retailer John Lewis and car manufacturer Seat, across 18 publisher sites. It manipulated which advertisements the respondents were exposed to – and on which platforms – and found that exposing consumers to advertising across two platforms was powerful.

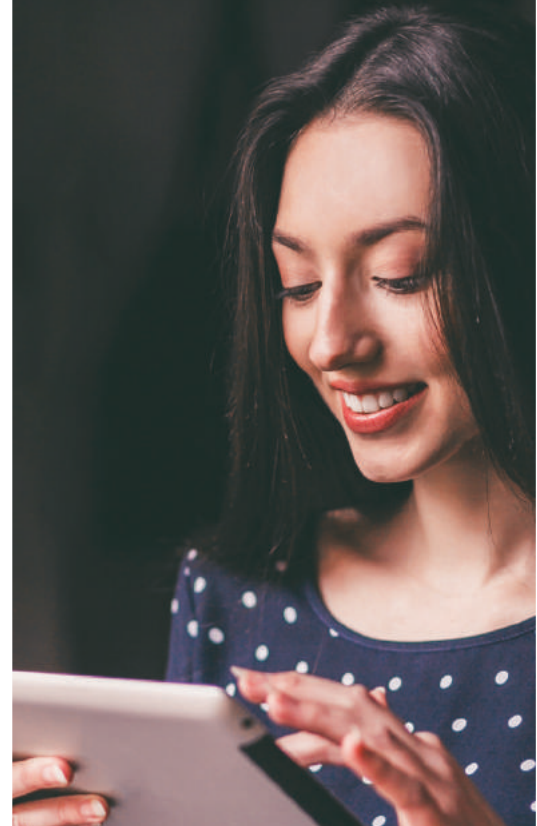
When the ad was viewed on two devices out of a desktop computer, tablet and mobile, awareness rose from 20% to 60%. Meanwhile, consideration went up from 6% to 30% and recommendation 0% to 15%, while 52% said they would think differently about the brand, compared with 5% who were exposed on just one device.

This is just one example of how a multiplatform, multichannel world offers huge advantages to advertisers who use it well – but it also starts to create questions about how difficult it is to measure. Research Now can track in-app use using its panel, but this is not the case for social media. “The biggest challenge we see at the moment is on platforms such as Facebook and Instagram,” says Corcoran.

Facebook's data policy tells users that: ‘We may provide these partners with information about the reach and effectiveness of their advertising without providing information that personally identifies you.’ It won't pass on what it defines as personally identifiable information. “Therefore there's a whole channel we can't measure,” Corcoran explains. “We feel frustration about the consistency of approach. But we supply first-party data, from a single source, so we can be sure the data we have is of good quality.”

Transparency and harmonisation

Phil Shaw, a director at Ipsos Connect, understands the consequences of this for brands: ▶



Ad and Audience Insight with Research Now

www.researchnow.co.uk
infoUK@researchnow.com
+44 (0) 207 084 3000

▶ “Large advertisers, such as P&G and Unilever, and ad agencies are concerned that there are publishers who control their own platforms and report their own results. They are wanting to see more independent verification.”

It’s impossible to measure the effectiveness of advertising unless we know, with some consistency, who has seen it. But the various methods by which advertisers, broadcasters and publishers work

“If you argue brand advertising isn’t working on digital platforms, you need to poke a stick at the measurement”

together to measure audiences in the joint-industry currencies have long been controversial. Measuring the impact of advertising across purely digital channels is hard enough, but bring in radio, TV, press or out-of-home – all of which, inevitably, acquire data in different ways, and some of which mix passive measurement, surveys and diaries – and it is impossible to

ever be confident that the numbers are consistent.

Some of the measurement problems – for example, how long we need to look at an advertisement for it to count – are subjective. But critics point out that numbers are not a basis for decision-making unless reporting is audited and transparent. Sue Unerman, chief strategy officer at MediaCom UK, has been “harping on about it for a number of years”.

“Some people count one-second views as a ‘view’,” Unerman says, “Our own data shows that, for every 20 ads served in [Facebook’s] news feed, only one is watched for more than 10 seconds.”

This research comes from GroupM’s article ‘Video: the battle for the billions’, published in its in-house magazine, *Interaction*, in February 2017. It calls the way in which video views are measured “nonsensical”. Unerman agrees with this criticism.

“If you argue that brand advertising isn’t working on digital platforms, then the first thing you need to do is poke a stick at the measurement system, rather than just focus on direct-selling ads, because they seem to work better.

The industry is at fault [for the poor standard of audience measurement]. The lack of a joint stance on this – which has continued to be the norm since the day I started out in this so-called industry – needs wholesale rethinking. I expect some [of the audience figures I see] are over-reported and some are under-reported, but I just don’t know. The problem is the lack of accuracy.” However, Unerman has praise for the way in which broadcasters are working to make TV ratings accurate and transparent (See box, Barb Dovetail Fusion).

Lynne Robinson, research director of the Institute of Practitioners in Advertising (IPA), emphasises the need for transparency to create consistency, but says it won’t be a quick fix. “Whenever anything new begins, you have to start to develop standards,” she says. “We are in the process of going through those standards, of agreeing to those standards. In terms of video impact, we’ve worked with Barb [the Broadcasters’ Audience Research Board] and Jicwebs [the joint industry committee for web standards] to develop a whole set of definitions that are available for everybody to use. But everybody ▶

THE MISTAKES YOU NEVER MAKE

Pre-testing has well-known problems, not least that it has a bias for short-term campaigns – so pre-tested campaigns underperform the average sales effect by 7% after one year, and by 10% after two, according to the IPA’s report *The Long and The Short of It*. The authors conclude that “emotionally focused pre-testing techniques should overcome this problem, but the onus should be on the suppliers to prove this”.

Realeyes is one of those suppliers. It uses facial recognition to track respondents’ emotional response to advertising, which might not be captured in a conventional survey. It claims 75% accuracy in identifying TV ads with high or

low sales lift, and 67% accuracy in identifying charity ads that will encourage high donations.

“The beauty is that we know exactly how the respondents are feeling at that moment,” says Peter Haslett, director of customer development at Realeyes. “We can say to the client, here’s what did it, here’s why, and who.”

For example, the most-watched advertisement of the 2015 Superbowl was Budweiser’s ‘Lost dog’, in which a cute puppy is saved from a wolf by horses. With each slot selling for around \$4m, this is a big bet. “It’s adorable,” said *USA Today*. “Budweiser’s little pup has won over the world’s hearts,” added *Business Insider*, based on the explosive social

media reaction. But the little dog was not seen again at the next Superbowl, having been judged a commercial flop.

Why? Haslett fires up the ad in the Realeyes dashboard: using a Realeyes panel it scores 10, 10 and 10 for Retention, Engagement and Impact with female viewers (with a skew to over-35s), and 3, 5 and 3 among males along the same dimensions. Pre-testing using emotional response would have spotted that the people who statistically are more likely to share videos about a dog loved it, but those who buy Budweiser instinctively hated it.

Realeyes does testing by allowing advertisers and researchers to self-serve. “They

take the video, and drag and drop it into our dashboard,” he says. “They specify the location, the audience they want to use, and any category needs.”

One of the effects of the Realeyes method is that it can aid performance measurement by predicting the likely effect on specific KPIs early in the campaign, or create audience response data in custom segments. In this way it helps to spot diminishing returns, or optimise media buying.

“We can run it during a campaign and say: ‘Here’s where you were earlier, here is where you are now,’” Haslett says, “In the past, I’d only run into the media people by accident. Now they’re there, in the room.”



has their own first-party digital data. It's really trying to get some harmonisation on the data definitions. The underlying issue is: are they transparent about how they counted something?"

Robinson references a recent speech by Marc Pritchard, chief brand officer at P&G, at the Interactive Advertising Bureau's (IAB's) annual leadership forum. He criticised the media supply chain – which he called “murky at best, fraudulent at worst” – and the disparate standards of self-reporting on digital platforms.

“The gig is up,” he said, before adding that, by the end of 2017, P&G would insist on external verification of audience numbers using MRC standards, or it would pull its advertising.

“We've come to our senses. We

“ It's about agreeing a transparent way to trade. If media owners do not do that, they may lose revenue ”

realise there is no sustainable advantage in a complicated, non-transparent, inefficient and fraudulent media supply chain,” Pritchard concluded.

Robinson has sympathy with his view: “It's up to advertisers and agencies to demand transparency and, hopefully, harmonisation of measurement. I think we want transparency. Harmony would be wonderful, but the initial phase is transparency on what you are counting. It isn't too hard to do; it's a process of industries maturing and agreeing a transparent and efficient way to trade. If media owners do not want to do that, then they may lose revenue on the back of it.”

Where there is an information gap, others move to fill it (see box-out, ‘Econometrics in the real world’). In

the UK, Unerman says that MediaCom's econometricians create a sustainable advantage by filling this transparency gap, and finding where there is advantage for advertisers either through inaccurate reporting or erroneous interpretation of what the numbers mean in reality.

“At MediaCom, we make evidence-based planning decisions,” she says. “We have our own data, and we work on the basis that audience research from the media bodies serves as a trading currency. We can deliver exponential improvements by using this to drive our planning, rather than relying purely on industry research.” But Unerman adds that she would be willing to swap some of this competitive advantage for a higher, common standard of measurement.

“If industry measurement improved, everyone would benefit. In the long term, advertising is a very important revenue stream for media owners. They are struggling to maintain quality and, if the industry as a whole was working in a positive spiral, that would benefit the media owners, as well as our clients.”

Gen Z knows what it likes

In the absence of detailed reporting on how people consume advertising, there is the potential – as Unerman warns – for media owners to maximise short-term revenues at the expense of irritating the rest of us. Duncan Southgate, global brand director, media & digital at Millward Brown, runs the yearly AdReaction survey, which this year focused on the habits and preferences of the emerging consumers in ‘Generation Z’ – and how they differ from millennials and generations X and Y.

“Even if there are flaws in the way that we measure audiences, or if the currencies are still flawed, then – using this – at least we can think about that quality decision from day one,” he says.





Millward Brown has a global reputation for measuring the impact of advertising, but this project aims to investigate the global trends that underpin those responses. From its inception 16 years ago, it has focused on the changing responses to digital, and how it compares to what we still call 'traditional' formats.

The long-run view of AdReaction shows that digital advertising is too often an irritant. In 2001, the survey was focused primarily on pop-up ads. Hard to imagine, but – at the time – they were as popular with consumers as TV ads, and TV ads were less popular than billboards, radio and print. But, by 2004, digital had dropped behind the other formats, and has not closed the gap since. “By 2004 we were already seeing massive variation in attitudes to online ads,” Southgate points out. This was partly a reaction to the appearance of intrusive formats such as pop-unders. “While the industry has generally moved away from the worst offenders, other intrusions – such as non-skippable pre-rolls – have taken their place.”

In the latest survey, which now covers 39 countries and 24,000 interviews, there’s a nasty surprise about the digital generation. “We expected Generation Z to be more accommodating to mobile ads,” says Southgate. “But they’re not. Despite spending more time on mobile devices, they are actually more negative.”

Southgate welcomes Google’s decision, announced in February, to ban 30-second, non-skippable pre-rolls from YouTube. “We’re in danger of undermining the people’s acceptance of advertising. Gen Z are more discriminating.

Mobile and digital aren’t just one thing, but a massive variety of skippable ▶

► and non-skippable video, display, and streaming, in page, and so on. What came through really clearly was that gap between how positive they [Gen Z] are towards the best formats, and how negative they are towards the worst.”

If the effectiveness of advertising is misrepresented – by accident or design – Millward Brown’s research implies that Generation Z will become increasingly resistant – for example, by using ad-blocking software. In February 2017, 22% of consumers used ad blockers, according to research conducted by YouGov for the IAB. While the IAB reports that ad-blocking levels have stabilised, it has called for a “better, lighter and more considerate online advertising experience” on the back of this research.

“Because Gen Z are used to interacting with anything and everything, having so much choice makes them very selective about what they absorb,” Southgate warns. “Millennials are actually becoming more positive to digital ads than the Gen X guys were. The gap is widening again. So it’s a perception being dragged down by a few formats that we wish weren’t in existence.”

Perverse incentives and short-term effects

Even if we measure well, we also have to interpret digital data, and it is problematic to link measured behaviour to outcomes. At the very minimum, the client needs to know which outcomes it wants to measure – and not be distracted if it doesn’t move the needle outside those KPIs. This hides the problem that a KPI has many real-world consequences.

When we measure, we incentivise the behaviour that we are measuring. Pre-testing can fail because it emphasises certain types of response (see box-out, ‘Realeyes’) and econometric measurement can give misleading results (see box-out, ‘Econometrics’). Shaw, at Ipsos,



highlights the problems of media buzz as, for example, an indicator of real-world consequences for FMCG brands. “How many people are going round having long conversations with their friends about biscuits? It just doesn’t happen. Just because you can do something doesn’t mean you should do it.”

Digital measurement also creates perverse incentives, adds Shaw. “If you’re going to optimise to click-through rate, you’ll end up serving campaigns in which the ads that get the most budget are the ones that have the best click-through. But click-through rate has a correlation of virtually zero with anything meaningful.”

Another perverse incentive: many advertisers, Ipsos finds, want viewers to watch their ads for as long as possible. In research with Google, Ipsos discovered that there was a single factor that caused skipping: a brand message – leading some clients to bury theirs. “If you’re in the entertainment business, then

great,” says Shaw, “but if you’re in the business of selling products or building brands, you need to optimise the videos that deliver the most brand impact.”

For advertisers who want to drive conversions online, the advent of retargeting and other conversion-driven technologies has been an

“How many people are going round having long conversations about biscuits? It just doesn’t happen”

advantage, simply because the conversion rate can drive the cost of advertising by design. Jon Lord, commercial director for adtech company Criteo, sees the way the industry works as creating efficiency without research. For Criteo, the results that advertisers get from their campaigns decide the price that their algorithm offers in the

BARB DOVETAIL FUSION

Barb is evolving the way it measures television audiences with Dovetail Fusion, a project to evaluate the total reach of programmes and advertising campaigns across TV sets, personal computers and tablets.

Fusion is due for delivery in March 2018.

Justin Sampson, chief executive of Barb, told

Impact about what Dovetail Fusion is designed to deliver, and the challenges to rigour and transparency that it has to solve.

What is Dovetail Fusion?

We’re collecting data from 5,100 homes every day and we’re now collecting census data from TV-player apps – so every time somebody in our panel of homes, so that we can get evidence of how people are watching on computer devices; we’ve now got meters on personal computers and tablets in about half of eligible homes. Third – which is the contract we’ve just announced [Kantar Media won this work] – we need to actually develop the fusion of the data, and

How do you do that?

The first deliverable was to generate online census data, and we can put a tick next to that one. The second was to install software meters in our panel of homes, so that we can get evidence of how people are watching on computer devices; we’ve now got meters on personal computers and tablets in about half of eligible homes. Third – which is the contract we’ve just announced [Kantar Media won this work] – we need to actually develop the fusion of the data, and

there’s now a 13-month development period before it will be ready for launch.

What will you report?

There will be two file formats. The first is equivalent to the daily data file we produce already, with audiences by programme, but we will be delivering that for all online viewing. The second is a respondent-level database, to get more flexible and detailed analysis. So if you were Unilever and looking at your Marmite campaign, you’d be able to see what the reach and frequency was for your – I hate the word traditional – linear campaign. Then you would be able to see the reach and frequency for the on-demand part, and the combined reach and frequency across both parts of the campaign.

You’re crunching two different types of data together. How transparent will you be about how you do it?

For an organisation such as Barb, objectivity and transparency is fundamental. So we have to develop a fusion that is explicable and understandable. People will not necessarily inspect the algorithms, but they can see the process and how it’s been done and know it’s not favouring one type of viewing over another.

Will this help advertisers and media buyers understand audience fragmentation?

This is an issue Barb has had for 35 years. The researcher’s answer to fragmentation is to build bigger online samples. How can we build bigger samples in a cost-effective way? Can we access return-path data from set-top boxes, and integrate that with our data? That would be a very cost-effective way of generating large samples of device data.



► auctions for inventory. “Our technology is designed to generate clicks that lead to conversions,” Lord says. “Our ROI is very focused on driving conversions, whereas other providers may work on a CPM [cost per thousand] basis. We’ll always work with an ROI goal. The beauty is that – because it’s an auction model – we will lower our price if our advertisers don’t hit their target.”

Criteo is also working to be as transparent as possible about how it achieves its results. “A lot of our

advertisers now have a much stronger view on measurement. They are more focused on generating performance from all their channels, and it sets us up to take more share of wallet than we had historically,” says Lord. “In many meetings I have now, we discuss with advertisers what data they want to share, to create key insights to help them understand customer behaviour within their attribution platforms.”

Adtech at its best has become an extremely efficient way to maximise

ROI for one type of advertising. But the IPA’s report, *The Long and The Short of It* – written by Les Binet and Peter Field – uses long-term data to recommend a 60–40 split between creating desire and satisfying demand. It points out that too much of one undermines the other, and observes a drift towards what it calls ‘rational’ campaigns.

Unerman, at MediaCom, argues that some of the growth in response advertising is an unintended consequence of it being easier to measure with certainty. Advertising that creates desire and advertising that satisfies demand can be of equal quality if they are doing what they are designed to do, she says.

“You can very quickly get bedazzled by rational campaigns,” Shaw warns. “I think, now, many advertisers are saying: ‘Well, that’s fine, but what did it do for my brand? Am I building long-term brand equity? Will it increase sales in the future?’ Those are questions that most of the digital metrics can’t answer.” ■

ECONOMETRICS IN THE REAL WORLD

If you want to understand the value of a campaign, then a sophisticated, econometric market-mix model is a powerful – and pervasive – tool. But Pointlogic offers the opportunity to model at the respondent level instead, and aggregates what it learns to create a bigger picture. This means that Pointlogic’s models can forecast the effect of advertising on brand KPIs such as consideration, and test and optimise media plans against these KPIs.

The company was founded 25 years ago, by a mathematician in the Netherlands, and was acquired by Nielsen in 2016. “Being part of Nielsen opens up many opportunities for us,” says Phil Spencer, business director UK at Pointlogic, “It has some huge data sets, and those are untapped resources for the sort of work we do.”

The two key features of Pointlogic’s approach are that it

uses respondent-level data rather than aggregate spend on each channel, and it uses Bayesian analysis – in which data is used to refine existing beliefs – to create more appropriate estimations.

Spencer explains the attraction of the respondent-level approach: “The standard market mix model running over three years has only about 150 data points, if you collect data once a week,” he says. “But if you have 1,500 respondents each week, you have thousands of data points to model.” A small number of data points, modelling many potential causes, will mean that trying to establish which actions caused which effects may lead to large errors.

This is the approach that Pointlogic employs with ESPN, which uses its Valuepoint product to model the impact of advertising across all of its vehicles for its top 26 advertisers. “ESPN doesn’t have the sales

data for its advertisers, but it wants to be able to have informed conversations,” Spencer says.

Valuepoint takes the individual responses from panellists on media consumption – and measurement from digital sources – and matches that with the media schedules of advertisers. It combines those with KPI scores on measures, such as awareness, that are captured in the tracker. It can then estimate how exposure to a campaign on each platform contributes to an awareness uplift, and when there are diminishing returns.

Finally, it aggregates these results across platforms to find not only the ROI in terms of effect on KPIs for past campaigns, but a predictive tool that allows planners to test ‘what if’ scenarios: for example, with a fixed TV spend, what mix of other platforms would achieve the highest uplift. ESPN, for example, can show its advertisers

hypotheticals: not only plotting the speed of change in awareness against exposure across all its channels, but also how one platform complements or drives the other in detail.

Spencer argues that conventional media-mix models often fail to capture nuances in how advertising performs, not least because the attribution models use some brave assumptions. “In some analysis, anything that can be connected back to a sale online is assumed to be causal. If there’s a cookie, and somebody buys it, it gets attributed to that platform. And just because someone does a transaction on a website, the industry often attributes it to digital advertising. That’s a fallacy, but it persists,” he says.

“The problem, for econometrics in general, is not just to stare at columns of numbers, but to think about what it means in the real world.”