

# CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

*A White Paper For  
The Coalition for Innovative  
Media Measurement  
and The 4A's  
February, 2017*

## Introduction

*"We are looking into adapting marketing mix modeling into something more dynamic and useful – more like attribution."*

And there you have it. The reason for all the interest in new methods of ROI analysis – "more dynamic and useful" -- moving past the annual review of marketing ROI and reallocation of budgets to the "near real-time" tactical guidance to yield better ROIs during the campaign.

But the requirements for success are high and well, in truth, the industry isn't there yet ... despite the swift and sweeping change in what you hear and what people are talking about. You wouldn't know that the industry isn't there from looking at claims providers are making. As in any early-stage market, there's a lot of hyperbole and over-promising. (Can everyone be using "best in class" data and analytics?) We are very excited about the pace of change, but it's important every once in a while to stop and see where we are, where we're going and what it's going to take to get there.

As few as four years ago, marketing mix models ruled media and marketing budget allocation and the new approach, attribution, was purely about digital paths and allocation. The gulf between the two systems that have similar outcomes (contribution to sales) was wide.

Today, the two are coming closer together. Mix models can now incorporate attribution and attribution is striving to incorporate media and marketing events beyond digital. But it's a confusing time – there are many vendors, many promises, and many different approaches emerging in ROI evaluation.

## Research Process

We were asked by the Coalition for Innovative Media Measurement and the 4A's Media Measurement Task Force to articulate the current state of marketing/media ROI analysis – and marketing mix models and attribution in particular. We started off by interviewing seven industry experts and with insights from that phase, developed a Request for Information from a variety of research providers throughout the industry. Twenty-one firms replied to the RFI. It is on this foundation that we address current practices in Marketing ROI measurement.

Before we get too far into our study, what is Attribution, and what are Marketing Mix Models?

## Definitions - Attribution

Attribution is simply the process of assigning credit to the marketing stimuli consumers encounter along the path to "conversion" – taking action, sales etc. Sometimes driven by "rules" or algorithms, sometimes by statistical models, attribution assigns value to each element's ability to drive traffic and sales. A



## CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

distinguishing characteristic of attribution is that it operates at a micro level – individual persons, households, devices and events.

The definition gets more complicated because there are a number of different ways of talking about attribution:

- Digital attribution – generally includes one or more trackable online element on both computers and mobile – display, paid search, organic search, email, etc.
- Television-Attribution – studies the impact of television advertising on digital outcomes such as website or app traffic
- Multi-Touch Attribution “MTA” – incorporates an important distinction, promising that *all* digital touchpoints, including social, are included in the evaluation.
- Cross-Channel Attribution “CCA” – the integration of offline media like television, magazines, radio and outdoor. This is clearly the ideal that marketers seek. It is not often found among the offerings we analyzed for this study. When it is, there are often many questions about the data used to feed the process. It’s a goal but the lack of addressable/programmatic inventory in offline media precludes this from being a significant factor today.

Attribution is attractive because intrinsic in the analysis is an understanding of the consumer journey, or path. Digital data are very discrete and provided in “near real time” data which allows for mid-flight optimization.

Keen to apply the same approach to non-digital media, they are experimenting with new data. Unfortunately, as we will see, there are few “closed loop systems” in other media like those that have driven the success of digital attribution.

### Definitions – Marketing Mix Models (MMM)

Marketing Mix Modeling, essentially, has meant the same thing for the past 30 or so years. Mix models involve the application of regression and other statistical approaches to estimate the impact of various marketing elements on incremental sales. They assess the effectiveness of spending by channel over and above a baseline of sales that would have occurred without any marketing efforts. MMM typically include other explanatory factors like seasonality, competitive activities and trade and consumer promotion.

Marketing mix modeling is attractive because most marketers have years of experience with it and rely on it for broad strategic analysis of the effectiveness of marketing investments. Generally, the Finance department and CFOs have conformed their systems to MMM and rely on the performance-based results for annual budget setting insights.



**Sequent Partners recommends that potential buyers and users ask questions and clarify how providers are using these modeling terms and what they mean. A common understanding of terms will minimize misunderstanding and disappointment.**

### Current Uses

A recent study by the Mobile Marketing Association showed that 39% of marketers currently use multi-touch attribution and that number is expected to grow to 75% in the next two years.

Attribution provides device or persons-level insights about the full range of digital options.

## CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

This allows marketers to optimize at a very tactical level to improve performance and return. Guiding tactical decision-making is one of attribution's strongest applications.

By comparison, marketing mix models provide market-level, store-level insights and are backed data from strong syndicated sources on both the input and output side of the equation. The ability to inform channel-level strategies is marketing mix modeling's greatest application and its prevalence in the industry demonstrates its reliability. Once the purview of packaged goods advertisers, marketing mix modeling are now used in virtually all product and service categories -- financial services, travel, pharmaceuticals, retail, etc.

### Strengths and Weaknesses

Attribution and marketing mix are different analyses that answer essentially the same fundamental questions: is the budget right? Is the media mix paying off? Where should I invest more or less?

#### Marketing Mix Model Strengths:

- A complete model - explains most of the variance in sales
- Statistical techniques proven over time; understood by management
- Accounts for all major sales drivers; less likely to misattribute cross-media contributions
- No privacy concerns with the data

#### Marketing Mix Model Weaknesses:

- Mix models are considered too slow, too macro and too backwards-thinking
- Typically conducted annually
- Insufficient granularity and timing to drive tactical decisions and optimization
- Dissatisfaction emerging

Today, marketers are pushing for more frequent, timely, granular and integrated “top down/bottom up” perspectives.

One of the marketers told us “*We are two years into the journey still doing the same measurement and MMM we did then. We are just now starting to trust some of the new techniques. 90% of the stuff doesn't work. The techniques have more error than anyone wants.*”

When attribution began, simple algorithms were used – and data were fit to a *a priori* pre-set solutions. Today, more complex statistical methods are used to infer causality rather than “Last Click,” “First Click” “U”, “W” or “M” rules and configurations.

#### Attribution Strengths:

- Examines creative executions, media placements and the synergistic combinations of the two
- Provides outputs at a level of granularity to impact timing and tactical decisions
  - Models can be updated weekly; digital data ingested daily; limited analytic feeds to television programmatic
- Covers new ground with novel techniques like AI and game theory
- Most advanced providers link to DMPs, programmatic buying/selling platforms and dynamic ad targeting

#### Attribution Weaknesses:

- Digital can get too much credit if other factors aren't accounted for
- Most systems do not account for a baseline of sales, or prior consumer purchase habits
- HH-level data quality can be questionable and timing slow
- Involves untested data and techniques; too little experience to know what works best for specific purposes

# CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

- Depends on imputation of offline media, may/may not consider price, promotion, brand momentum or factors like the weather and economy
- Fair amount of salesmanship in this arena; lack of transparency in the techniques

And in this transitional period, the industry is wrestling with the right model for the right application. When attribution and marketing mix modeling tackle the same questions, results rarely agree and marketers don't know which results to follow.

The real challenge facing the industry is how to integrate the best of both approaches. Simply bolting the two together has led to varying degrees of success.



**Sequent Partners recommends asking if you understand every step of the process from data inputs to modeling and model outputs; does it make sense and does it accurately reflect your business in all its nuances? Can the provider prove that the model is accurate overall and for every marketing element in the model? Will the model help improve my business results?**

## Approaches in Detail

Having seen what the approaches are, and their strengths and weaknesses, now we will address the modeling approaches in much greater detail. We will focus on techniques typically used by providers across the industry today: marketing mix models plus attribution, attribution alone and single source.

## 1. Integrated MMM Plus Attribution

Top-down and bottom-up, integrated marketing mix modeling plus attribution would appear to offer the best of both techniques. This approach can offer a full revenue attribution of offline and online marketing elements together. MMM is used to determine ROIs and contributions of each medium in the context of the complete marketing mix. It is a complete model that explains most of the variance in sales.

Attribution optimizes digital tactics and the digital path to purchase. In this approach, attribution is constrained by the MMM parameters, i.e., sets levels -- not allowing historical search or other digital elements to account for more than x% of sales, as determined by the marketing mix model.

Several marketers talked about this two-stage model -- running a "normal" MMM then using attribution to break down the results into networks, genres, and major sponsorships. For digital, they look at paid vs. organic search, all types of display. Sometimes, they can go to a message or campaign level. This is predicated on the fact that they don't expect one tool to do everything and joining attribution to marketing mix and calibrating the attribution outputs provides a workable solution.

This unified approach links market-level learning to consumers using automated data collection and machine learning for fast, granular in-flight insights. Right now, though, the linking techniques have not been validated.



**Sequent Partners recommends asking exactly how the macro level mix model and the micro level attribution model are linked and how your provider knows it works?**

## 2. Attribution Alone

In this approach, return on investment and individual media contributions are determined directly with household or device-level data and modeling approach. All media are handled similarly but this analysis may require multi-level models that treat some media with greater granularity and a faster tempo than others.

Attribution is flexible and can measure a variety of outcomes from search terms, website visits, ecommerce shopping or sales, to offline retail visits and sales. In this study, we found it rare that attribution ventured beyond the digital marketing sphere to the offline marketing world. The great risk in that is that sales or other outcomes that were produced by offline media, or other factors may be attributed, erroneously, to digital media.

By its nature, attribution can examine individual creative executions, individual media placements and the potentially synergistic combinations of the two. It can consider the responsiveness of various consumer segments and place the right ad in front of the right consumer at the right time.

However, at the moment, attribution is dependent upon a series of untested data and techniques:

- Untested data inputs
- Untested media matching
- Untested offline-online matching
- Untested modeling techniques.

We see a high degree of faith and salesmanship in this arena. The promise is great, but there is work to be done before that promise is secured.

The offline media inventory that attribution can influence is limited. However, as this technique

creates value for both buyer and seller of media, we would expect to see more addressable inventory available. At that point, it will be critical for Cross-Channel Attribution to be fully functional.



**Sequent Partners recommends asking your Attribution provider to:**

- **Demonstrate that the data being used accurately reflects your business**
- **Be transparent about the way online and offline data are matched and identify any potential distortions**
- **Describe how the modeling technique reflects the way in which consumers respond to advertising**

## 3. Single Source, Panel-Based and Agent-Based Modeling

These three closely related techniques operated at the individual person or household level long before attribution modeling. We have included them in this study because, like attribution, they are all based on an integrated individual media exposure and product purchase engine. The ideal system has been in the mind's eye of marketers since the 1980s – originally based on UPC-scanned purchasing and set-metered television exposure collected from the same household. Today these single-source data analytics utilize abundant big-data matched at the household level, and can integrate television, digital, in some instances radio, magazines, and out of home. This full spectrum of media options can be analyzed with respect to their sales lift across a number of industries including, CPG, retail, auto, entertainment, pharmaceuticals, and anything purchased with a credit card.

# CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

Panel-based approaches can accomplish the same things starting with a panel of individuals to whom all of their media exposures and sales outcomes can be attached. While these panels can't match the scale of big data, they have the advantage of being persons-based, not household based, and they come with the ability to directly integrate attitudinal data via survey.

Agent based models simulate the households of single-source or the persons of panel based analytics. They describe a class of modeling approaches that can be fueled by any data source. They are especially adept at simulating future scenarios and could be combined with single-source or panel based approaches for that purpose.

## 4. One More Option - More Granular Marketing Mix Models

One additional approach surfaced during our investigation of current practices that bears mentioning. Undertaking more specific modeling within a traditional MMM framework, with daily data, finer geographies (e.g., built-up from block groups or zip codes) and more sophisticated models could yield greater granularity and the timing to drive more tactical decisions. Just because most marketing mix models are done on an annual basis, for instance, and with larger geographies, doesn't mean that's the only way to employ these models. This approach could provide better, more granular and precise tactical insights, it still may not be sufficiently granular for digital media optimization.

It's interesting to note that while both MMM and single-source analytics began focused on television and the classic CPG marketing mix, they most recently integrated digital media as an equal partner. At the same time, digital attribution has moved to integrate television and

other offline media, along with the rest of the marketing mix. Both approaches rushing towards the other's core offering in order to meet marketers' needs.

## Modeling Techniques

There are at least 26 different modeling techniques in use across the spectrum of Marketing Mix Models and Attribution, from simple multivariate regression to more sophisticated Bayesian and other multiple equation methods.

Multi-Touch Attribution and Cross-Channel Attribution do not have common technical approaches and the number of techniques in place show about how bleeding edge this work really is.

For the most part, the statistical techniques in use for MMM have proven the test of time. Because attribution works at the individual level (event, person, HH, device), it requires different techniques. Some traditional econometrics (e.g. logistic regression) are used, as are novel methods like game theory. At this point, however, across the industry there is too little experience to know what works, or even what works best for specific purposes. This topic is ripe for further study. The full list of modeling techniques is in the appendix.



**Sequent Partners recommends asking if key model inputs represent your business. Are price, promotion, distribution and other marketing variables included? Are key environmental factors that impact business results (e.g. weather, economy) accurately and completely reflected?**

## CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

This study is not large enough to quantify the precise proportions of the various techniques in use. However, the 21 responding providers represent the industry's major players and provide a good cross section of the techniques mentioned above. It is a useful sample for giving us a sense of the current landscape.

### Common Product Offerings

The most common products being offered today are integrated Marketing Mix Models and Multi-Touch Attribution with the findings applied to digital media only. About half the providers offer this solution in some form. It was surprisingly rare to find these cautious modelers venturing beyond digital to apply their findings to television.

Unsurprisingly, purveyors of MTA apply those findings to digital media. And providers of television attribution only apply their findings to television.

We see more interesting variation among Cross-Channel Attribution providers -- with some limiting application to digital, some integrating their findings for digital and television and about half promising to drive improved tactical decisions in multiple media.

These cases span a spectrum of caution to risk-taking to deliver an inversely related range of promised benefits – a classic risk-reward trade-off.

### Specific Concerns About Media Data

Media data are necessary model fuel. They are also a significant investment on the part of the marketers and modelers. But right now, there are data quality issues that may

create bias and potentially misleading findings in the short term. MRC-audited data sources are highly preferable, when available.

Incomplete media data is a systemic issue - not associated with only one type of modeling or model provider. As an example, we often hear worries about incomplete digital data painting incomplete paths, which can be very misleading. The blind spots created by walled gardens, such as Facebook and Google, are especially damaging. For the most part, household-level data is problematic right now. In digital, the unit of measure is devices, not people. In television, it's sets, not people. The practice of marketing to households was deemed invalid more than half a century ago. It is foolish to step back in time because we have a data problem. And linking specific media data to households for a complete household level sample is complicated and in many cases untrustworthy.

Here is a list of common issues associated with current media data.

- Ad Servers: Data from all ad servers may not be included
- Cookies: Not one-to-one with consumers -- users delete cookies frequently, use multiple browsers, multiple devices, multiple users on the same browser
- Tags: Not all sites accept third party tags (e.g. Facebook, YouTube)
- SDK: Need to be installed in all relevant apps, leading over-the-top players don't cooperate (e.g. Comcast)
- TV Set Top Box: Data sharing limitations, convenience samples across geographies, incomplete data (multiple data sets sometimes stitched together to simulate national sample.)
- Smart TV data: Data sharing and privacy limitations, convenience samples by

manufacturer, and limited number of sets per HH

- Subscriber data for magazines, not readers – (Reader-per-Copy estimate sometimes applied)
- Digital radio only, not terrestrial
- OOH – possibly integrated with location latitude/longitudinal data, but often based on home location, not trips

Why do these data issues matter? Why isn't a system in development, albeit with incomplete data, better than no system at all?

We heard the notion several times – it's better not to let *"perfect be the enemy of good."* There's truth to that. But do we know if "good" is "good enough" or just wrong? Incomplete and erroneously linked data can result in biased and imprecise information. The most advanced attribution systems link their results to DMPs and buying/selling platforms, and their impact is felt quite immediately. There is little room for error. In the fully automated system of evaluation, optimization and programmatic buying, standardized media data would go a long way towards increasing confidence in the system.



**Sequent Partners recommends asking these questions about every stream of media data in your model:**

- **Does it match currency levels?**
- **Are the geographies covered fully representative of the brand's business situations?**
- **Are individual ads and media placements identified in the data?**

### Concerns About Data and Device Matching

Attribution requires detailed data streams for each household that include every causal factor:

- All the digital data must be integrated through device matching
- Television's set-top box data must be matched at the household level with all of the household's devices
- Offline purchases must be matched to the household
- Offline marketing activities must be matched to the households they may have influenced
- Environmental factors such as weather must be matched to households
- And each household's purchase history with the brand needs to be captured to provide a baseline.

In the short-term, cross-platform attribution will suffer from imperfect cross-device matching systems.

This area is ripe for development. Amazon, Google and Facebook have a "goldmine" of consumer sign-on data across multiple devices data, but they do not share it. Device graphs, which link devices to households, are works in progress. There are several probabilistic solutions available, but reliability is uncertain. As one marketer told us, *"In the land of the blind, the one-eyed man is king. This limitation isn't stopping us. There will be a solution down the road."*

Consumer privacy constraints and other factors will limit the ability to match with certainty, but several companies are developing solutions and device graph roll-ups.

There will be an intrinsic bias in device matching due to the lack of full participation by consumers, but the industry will likely develop a gold

standard or a single provider that will eventually standardize this aspect of data inputs.

Another key methodological area is the accuracy of linking separate streams of data, like offline media, at the individual level. There are a number of unsubstantiated claims and broad descriptions, lacking relevant details, associated with these approaches.

Data fusion methodologies, involving tapping into match-back technologies and proprietary algorithms, can extract multiple variables from different data sources for matching and de-duping. There are providers that use match-backs with blinded personal individual information and in theory, these approaches should include all marketing tactics and sales across all channels. Despite claims of probabilistic and deterministic methods for matching being used, the success and quality of the match depends on the data available.

When marketers have addressable digital data, they develop a matchable clickstream across devices with links back to offline purchases. To assess cross-device interactions, when unified mobile/cross-channel IDs are available, they can be directly incorporated. When such data isn't available, there is a search for a method based on the best fit with available data.

Theoretically, data scientists and experts should be able to provide good matches, but there is no common approach or match review process currently used by the industry. The concern is that this can introduce significant noise, and lessens the precision of an approach built on granularity and precision.

CIMM's "Best Practices in Cross-Device and Cross-Identity Measurement" white paper delves into this topic in great detail. The paper is available on the website [cimm-us.org](http://cimm-us.org).



**Sequent Partners recommends asking how accurately the separate streams of data are linked at the individual level.**

### Validation and Transparency Concerns

There's a belief, generally, that model providers are not transparent enough about their approaches. Opaqueness does not serve the industry well – black boxes breed distrust and skepticism. Transparency should be the price of entry – but not all modelers are as open about their techniques as they need to be. In fact, several providers skipped the question altogether in our RFI, ignoring the issue. One of the marketers we talked to said, *"Transparency runs the gamut from white box to black box."*

Some modelers, particularly the market mix modelers, provide transparency in the model selection and specification, pros and cons of various modeling techniques, model outputs such as coefficients, confidence intervals, model fits and other model diagnostic statistics. But not everyone does this.

Internal validation consists of standard statistical approaches to ensure that the model is well built. Most modelers are very open about this, using holdout samples and goodness of fit statistics. In the RFIs, little was said about how the models perform over time, although a few providers suggested that built-in tests were the gold standard. This topic is worth further exploration.

The issues of transparency and model validation are important because we need to be able to determine whether the relationships addressed in the model are coincidental, judgmental or really causal. Are they flukes or facts? These

issues are so important – and complex - that it makes sense for a top-flight industry and academic working committee to explore them in greater depth.

### Organizational and Training Needs

Clearly, attribution modeling and marketing mix modeling is complicated stuff. It's no wonder so many questions are being raised throughout the industry. The media world tends not to be comfortable simply deferring to an outside research provider, without any ability to separate hype from truth.

There is a keen need for people to holistically understand what models can and cannot provide. They need to understand specific data sources and how to be smart buyers and users of models.

On the front end, marketers and their teams should be able to articulate what business objectives are meant to be addressed with media, which will help guide the analysis of its impact.

The industry will need to begin training people in interpreting the data and telling stories with these modeling insights. In addition, there will be a need for guidance on extracting meaningful action from modeling results. Some marketers told us they didn't know what to do with the results, or how to reconcile them with things they know about their businesses.

And finally, there is a fairly significant organizational structure issue facing the industry: how do you apply cross-platform insights in organizations that are not cross-platform?



**Sequent Partners recommends including organizational considerations when planning adoption of attribution modeling. It needs to be understood and applied if it's going to generate value**

### Completing The Picture: What The Models Need to Do

The rush to apply digital-like attribution metrics to the rest of the marketing mix is definitely producing some trade-offs.

But we should not relinquish everything we know about the way advertising and media planning work in our exuberance to embrace methodological change. Advertising has short-term effects. Advertising lingers. Advertising effects decay over time. Advertising works at certain thresholds at the low end – and ceases to work at certain thresholds at the high end. And media, through planning and scheduling, interact with each other and with other marketing elements – synergistically – producing an effect greater than the sum of the parts.

Many modelers incorporate ad dynamics – diminishing returns, saturation/threshold, and adstock/decay into their analysis. While most modelers fit these variables, some make assumptions about the shape of the curves rather than fit it with data. Most modelers fit these effects at the market level, and some even at the household level. In some cases, these factors are not mentioned – possibly not understood.

Beyond paid media, owned and earned media need be incorporated. In addition to pulling in the right data, this process requires modeling the complex interactions, halo and network effects that characterize the impact of these media.

## CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

Long-term effects are often handled in the models, typically in a multi-level model with key performance indicators such as brand awareness, brand health metrics, social media engagement, word of mouth, referrals and recommendations, and penetration. Customer lifetime value, a measure of long-term effects, is available in some industries. The Mela model of evolving baseline is one approach to long-term, while others handle this with long-term adstocks (weeks) or repeat-purchase patterns.



**Sequent Partners recommends asking if the model includes all of the marketing drivers you know to be important to the business and does it reflect the well-known dynamics of consumer response to advertising?**

This is a time of great experimentation and building. We observed very few systems that are start-to-finish, integrated hiccup-free looks at consumer behavior cross-platform. In order to have a truly complete picture, the models should include:

- All online and offline media
- The rest of the marketing mix
- Important environmental factors like weather
- All offline and online sales, or other conversion behaviors
- Some measure of brand health/strength
- Baseline sales must be identified in order to discriminate incremental sales

The last point merits elaboration. A challenge for many attribution models is replicating baseline sales, which are so integral to the ability to differentiate incremental sales from those that would have occurred naturally. Attribution

methods miss on the baseline because it is hard to put in the model and they lack the necessary data.

With the focus on short-term incremental sales in attribution, the question of what happens to the brand is very relevant. The integration of brand tracking data is essential to determine how well digital or traditional media drive the brand.

Attitudinal surveys have a place in attribution and marketing mix models, but sample sizes tend to be significantly smaller than what people are used to dealing with relative to behavioral data. And there's a distrust associated with attitudinal rather than behavioral data these days. But many modelers do include brand and other macro-economic factors in attribution models.

### Current Best-Practice

Given the early development stage of Cross-Platform Attribution, anointing a "best practice" now is likely to thwart development. However, the characteristics of a best-practice solution are clear. It should:

- Operate at a level of granularity that enables marketing plans to be optimized mid-course.
- Use a statistical model to infer causality for all the elements of the marketing plan.
- Incorporate a baseline and all the drivers of consumer purchase behavior that account for most of the variance in sales or other outcome variables.
- The data used for all variables must be representative of the brand's business, especially after being integrated through matching or imputation.

We foresee models that deliver all these characteristics. But we are skeptical because of the current state of data and data integration. For now, we believe both marketing

## CURRENT PRACTICES IN ATTRIBUTION AND ROI ANALYSIS

mix modeling and attribution modeling are necessary.

Marketing mix models are the most reliable way to evaluate the sales contribution and ROI of each element of the marketing mix. They are the ROI best-practice. Attribution models (when they take into account all sales drivers), are able to optimize marketing tactics mid-campaign. Attribution modeling is the activation best-practice. Linking these two appears to offer the best of both, but the techniques for linking have not yet been validated.

### Next Steps:

With sufficient resources the industry could tackle three development streams concurrently:

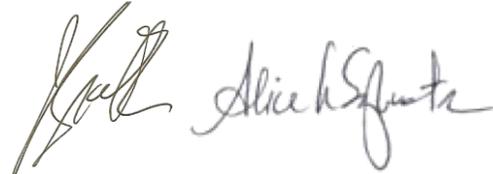
- Data sources and integration
- Modeling techniques
- Training and organizational best-practices

A recommended, detailed agenda for each is outlined in Appendix 2.

Broadly speaking data sources and linkage techniques need to be assessed and improvements recommended and implemented. Modeling techniques need to be more transparent, better understood and validated. Users need to deepen their knowledge and organizations need to identify and adopt best-

practices.

These steps should propel the industry forward and ensure event greater success. The coming years will be ones of great excitement as the industry leaps to the next generation of analytics and ROI measurement.



Sequent Partners  
2/1/17



380 Lexington Avenue  
New York, NY 10168  
914.862.0054  
[sequentpartners.com](http://sequentpartners.com)

## Appendix 1 Summarizing Key Differences In The Techniques

### Attribution Modeling

- Device or persons-level insights about the full range of digital/mobile options
- Allows for optimization of tactics to improve performance and return; guides tactical decision-making
- Provides trackable “paths” to conversions/outcomes
- Best for digital; providers are scrambling to align household-level data from other media
- Currently, resides in the digital media domain

### Marketing Mix Modeling

- Market-level, store-level insights backed by years strong syndicated data on inputs and outputs
- Reliably informs channel-level strategies
- Attractive because most marketers have years of experience with it and rely on it for broad strategic analysis of the effectiveness of marketing investments
- Has CFO buy-in

### Single Source, Panel-Based and Agent-Based Modeling

- Utilize abundant big-data matched at the household level,
- Integrate television, digital, in some instances radio, magazines, and out of home
- This full spectrum of media options can be analyzed with respect to their sales lift
- Available in a large number of industries including, CPG, retail, auto, entertainment, pharmaceuticals, and anything purchased with a credit card.

### More Granular Marketing Mix Models

- More specific modeling within a traditional MMM framework
- Utilizing daily data, finer geographies (e.g., built-up from block groups or zip codes) and more sophisticated models
- Could provide better, more granular and precise tactical insights
  - Though still may not be sufficiently granular for digital media optimization

## Appendix 2

### Next Steps: A Proposed Industry Agenda

With sufficient resources the industry could tackle three development streams concurrently:

- Data sources and integration
- Modeling techniques
- Training and organizational best-practices

### Data Issues: Completing Digital Attribution

#### 1. Data Quality

Complete, unbiased, accurate and precise data for every step in consumers' digital journey; either from a true census, or a representative sample.

##### Issues

- Walled gardens – the black hole of the Web
- Tags – not all sites accept third party tags
- Ad-Server data – potentially incompletely captured
- Cookies – not one-to-one with consumers (users delete cookies, use multiple browsers, multiple devices, multiple users on the same browser)
- SDKs – need to be installed in all relevant apps; not all players cooperate

#### 2. Data Linkage

The ability to link all digital devices for a single consumer for all, or a large representative sample of, relevant consumers.

##### Issues

- No provider can meet this need today; some users integrate three or more sources
- No transparency about the quality of the matching and the resulting usable sample
- No identification of consumers

#### 3. Complete Data Sets

Incorporate all the relevant data

##### Issues

- Include all offline and online sales and all of their drivers:
  - Baseline sales
  - All online and offline media
  - The rest of the marketing mix
  - Important environmental factors like weather
  - Some measure of brand health
- Ideally all offline media should be included as well, but given the way those media work, the chance of misattribution is not as great

## Appendix 2 Next Steps: A Proposed Industry Agenda (Continued)

There are similar issues in completing cross-channel attribution in off-line media:

- Television
- Magazines
- Radio
- Outdoor

### Data Issues: Cross-Channel Attribution – Adding Television

#### 1. Data Quality

Complete, unbiased, accurate and precise television data at the persons (or household) level; either from a true census, or representative sample.

##### Issues

- STB data
  - No complete source of national coverage
  - Growing over the air and over the top segments
  - “STB on” does not equate to “television on”
  - No persons data
  - Privacy issues can impede granularity
- SmartTV data
  - Smart TV distribution is far from universal and likely biased
  - No complete source of national coverage
  - Not all sets in the house are smart
  - No persons data
- ACR data
  - Not sufficiently deployed for attribution

#### 2. Data Linkage

Persons or household television exposure data must be matched to other offline and online media, marketing and outcomes data.

##### Issues

- Household matching
  - How accurate is the match?
  - How representative is the resulting matched sample?
- Probabilistic imputation
  - How accurate is the imputation?
  - Cross-Channel Attribution:  
Adding Magazines

## Appendix 2

### Next Steps: A Proposed Industry Agenda (Continued)

#### Data Issues: Cross-Channel Attribution – Adding Magazines

##### Data Quality & Linkage

Representative magazine readership at the persons-level; matched to other offline and online media, marketing and outcomes data.

##### Issues

- Subscriber Databases
  - Subscribers may not read an issue
  - Pass-a-long, non-subscribers, are the largest share of any magazine’s readers
- Household matching
  - How accurate is the match?
  - How representative is the resulting matched sample?
- Probabilistic imputation
  - How accurate is the imputation?

#### Data Issues: Cross-Channel Attribution – Adding Radio

##### Data Quality & Linkage

Representative radio audience data at the persons level; matched to other offline and online media, marketing and outcomes data

##### Issues

- PPM data (Nielsen Audio)
  - Only measures 48 markets
  - Limited sample in a big data setting
  - RLD only available inside of Nielsen
- Household matching
  - How accurate is the match?
  - How representative is the resulting matched sample?
- Probabilistic imputation
  - How accurate is the imputation?

#### Data Issues: Cross-Channel Attribution – Adding Out of Home

##### Data Quality & Linkage

Representative out of home audience data at the persons-level; matched to other offline and online media, marketing and outcomes data.

## Appendix 2 Next Steps: A Proposed Industry Agenda (Continued)

### Issues

- GeoPath data
  - Has provided only an annual audience estimate with no temporal variability
  - No persons or household data, market level only
  - Is in the process of integrating GPS data
- Probabilistic imputation
  - How accurate is the imputation?

In general, this situation is challenging because no media data source for attribution is 100% correct. Data may be “correct enough” for a given brand’s model. Meaning reasonably representative of the brand’s business, but users should vet the data against key criteria every time.

### Industry Next Steps - Modeling Issues

This study found 26 different analytical techniques in use and an unsettling lack of transparency. Some of the methods can be considered tried and true while others are innovative and unproven. At least publicly unproven. Choice of technique is heavily dependent upon the preference and experience of the practitioner; we do not recommend being prescriptive. However, the industry needs to know how each technique works, what it can and can’t do and some sense of how well it performs.

### Modeling Issues: Layman’s Guide

Developing a laymen’s guide to modeling issues is important. It would take the form of a toolkit to better inform the non-technical user of attribution modeling.

- A detailed exposition of each technique, its capabilities strengths and weaknesses
- Fit for purpose guidelines – specific modeling techniques recommended for particular use cases
- Guidelines for how to ensure that your model is valid for your purposes. How to use goodness of fit statistics, employ hold-outs, conduct tests, etc.

### Modeling Issues: Validation

Validation will be difficult due to the confidential nature of models and the proprietary nature of modeling techniques, but it would provide the facts to support smarter use of attribution

- Conduct confidential audits of existing modeling cases to validate the performance of specific modeling techniques for specific situations.
- Make the aggregate meta-learning public
- Identify the need for further development and innovation

## Appendix 2 Next Steps: A Proposed Industry Agenda (Continued)

### Organizational Issues

This study uncovered a need for training for buyers and users of attribution services. There were also no accepted best practices in organizational adoption of attribution.

### Organizational Issues: Training

Develop a training program for users to understand language, modeling strengths, weaknesses, data issues and appropriate applications. This program could be distributed through industry trade groups.

### Organizational Issues: Best Practices

Benchmark best practices in advertiser/agency organizational adoption and application of modeling to drive more profitable marketing/media tactical decisions. Once developed this could be distributed as a report, presentation and/or workshops through the 4As and ANA.

## Appendix 3 Questions Buyers and Users Should Ask

### The Big Three Questions:

1. Do I understand every step of the process from data inputs to modeling and model outputs; does it make sense to me and do I believe that accurately reflects my business in all its nuances?
2. Is the model accurate with respect to both overall fit/prediction and the effectiveness of every element of my marketing plan, competitive actions, and environmental factors?
3. Does the model help me improve my business results?

### The Details:

- Is the model input-data representative of my business?
  - Does it reflect the factors that influence consumer response to my marketing, as we see them nationally? (These questions will need to be adapted for different categories)
  - Do the sales or other outcome variables (if using third-party data) match what we know to be true:
    - Trend: Weekly sales volume for the past three years
    - Purchase Dynamics: Brand shares, penetration rates, purchase frequency
    - Coverage: by geography and key-distribution channel
    - Buyer profile: demo, psychographics, other useful segmentations
  - Media
    - Completeness: Are all the key media being used measured?
    - Do the reported levels match our reference measures: overall usage, media type usage by time of day and day of week
    - Coverage: are the measured geographies representative of the range of the brand's business situations?
    - Consistency: Are the media all measured at the same level of granularity? If not how is that handled and what deficits/risks result?
    - Granularity: are individual ads and media placements identified in the data?
  - Marketing
    - Price: does the brand's and competitors' prices reflect that same range as seen nationally?
    - Promotion: are the same offers available as nationally?
    - Distribution: Are distribution levels through key channels the same as national?
    - Other key marketing variables
  - Are key environmental factors accurately and completely reflected? (Weather, economy, commodity prices, interest rates, ... whatever has been found to impact business results)

## Appendix 3 Questions Buyers and Users Should Ask (Continued)

- Data linkage: how accurately are the separate streams of data linked at the individual level?
  - Digital devices
  - Digital to households
  - Offline media to households
  - Offline marketing to households
  - Environmental factors to households
  - Offline sales to households
  
- Can the modeling technique reflect the way in which my marketing elements motivate consumers and interact with each other?
  - Advertising: diminishing returns, adstock, halo effects
  - Social media: amplification effects
  - Granularity: are individual ads and media placements evaluated
  - Interactions: e.g. advertising potentiating promotions and dampening competitive pricing and promotional activity
  - Long-term effects: can slower-moving long-term brand effects be captured and reflected?
  - Brand momentum: is there a sales baseline?
  - Consumer segments: can various consumer segments be represented?
  - Evolving effectiveness: are factors like wear-in & wear-out captured?
  
- Other modeling issues
  - Colinearity: how is this overcome?
  - Validity: how is this proven?
  - Causality: Is it clear that the model captures causality, not coincidence, or some a priori idea of the role of each marketing element?
  - Multi-level factors: how are the macro (market level) and micro (household level) models linked?
  - Transparency: can the modeler explain how the model works in terms that give me confidence?
  
- Are model outputs suited to drive better tactical media and creative decisions?
  - Granularity: are individual ad units and their specific placements identified?
  - Tempo: Are results provided quickly enough to make mid-campaign improvements?
  - Metrics: Are results denominated in metrics that relate directly to your business objectives and decision-making process?

## Appendix 4 Study Participants

### Phase One Interviews

1. ABC
2. Kraft-Heinz
3. AT&T
4. Coke
5. CitiCorp
6. Digitas
7. Brand Science
8. Havas

### Phase Two RFI Responses Received

1. Convertro
2. C3 Metrics
3. Conversion Logic
4. Visual IQ
5. WyWy
6. Analytic Partners
7. Neustar/Marketshare
8. IRI
9. Millward-Brown
10. Thinkvine
11. Concentric
12. In4mation Insights
13. TiVo Research
14. Mphasize/Sapient
15. Merkle
16. Nielsen
17. Oracle/ Data Logix
18. Marketing Evolution
19. Adometry
20. TV Squared
21. Nielsen Catalina Solutions

## Appendix 5 Modeling Techniques in Use Today

- **Attribution**
  - Machine learning
  - Discrete Choice
  - Logistic regression
  - Unified logit
  - Hidden-stage Markov models
  - Markov chain
  - Game theory
  - Customer propensity models
  - ANCOVA
  - A/B testing
  - Test vs. control designed experiments
  - Ridge regression
  - Lasso regression
  - Agent based models with Monte Carlo simulation
  - Proportional hazards modeling
  - Elastic Net
  - Ensemble learning framework
- **Multi-level models**
  - Structural Equation models
  - Hierarchical Bayesian models
  - Vector Auto-Regression
- **Econometric models**
  - Multiple regression with non-linear transformations
  - Bass diffusion model
  - Granger causality
  - Seemingly unrelated regressions
  - State space time-series models
  - Anomaly detection