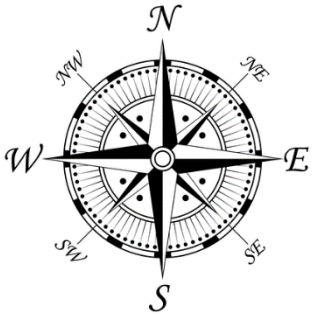
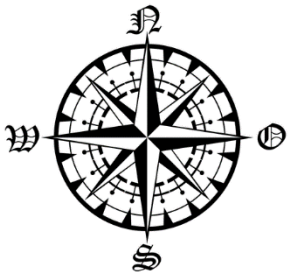


Getting Attribution Right: An Exploration and Best Practices for Television Data Inputs in Attribution Modeling

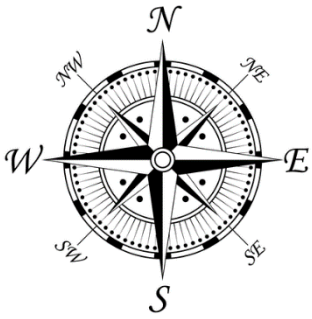
September 2020

Table of Contents



Part 1. Deep Dive Into Occurrence Data in Attribution

Part 2. Comparing Exposure Data

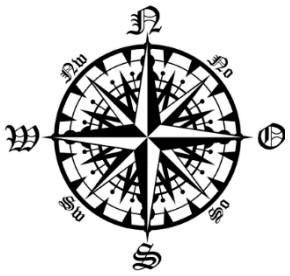


Introduction

Attribution providers offer many different approaches, including relying on different modeling techniques and data sources. These different approaches frequently lead to different results and business decisions. CIMM sought to unpack this issue and learn a bit more about what drives the difference in attribution results. And share best practices in data inputs, where appropriate.

Accurate television attribution depends on a host of variables, starting with accurate inputs—schedules and accurate identification of campaign spots. It is also dependent on ad exposure measurement, which, in television, is measured with Gross Rating Points (GRPs), Average Ratings, Reach, and Frequency. Outcome variables such as web visits, retail traffic, sales, or tune-in ratings are required. Another critical input is the identity graph that links all the variables at the device or household level. The analytics for measuring incrementality is the final piece of the puzzle for accurate attribution. Each and every one of these components can impact the accuracy of television attribution results. To begin the learning process, however, we structured this study for CIMM around the first two variables: key inputs of ad schedules and ad exposures.

Questions We Set Out to Answer



How do variations in occurrence data impact TV performance in attribution models?

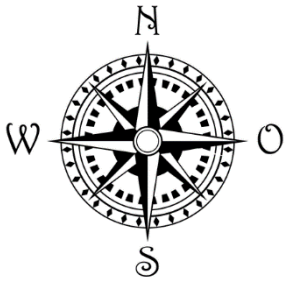
- *How different are the occurrence data? Why are they different?*
- *Do differences in occurrences make a difference in audience metrics? In outcomes?*

How do variations in exposure data impact TV in attribution models?

- *How different are the exposure data in terms of typical audience metrics and outcomes?*

How do variations in providers' combined occurrence and exposure data impact TV in attribution models?

- *How different are the occurrence and exposure data in terms of typical audience metrics and outcomes?*



Overview of Key Findings

1. Key television attribution inputs are highly inconsistent from provider to provider and across our test schedules. They may not entirely resemble the advertisers' TV buys.

2. As a result of inconsistent inputs, outcomes differ inexplicably by provider.

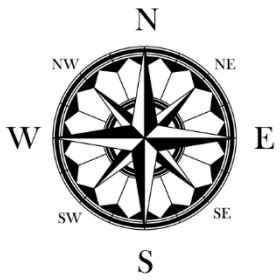
3. Provider exposure data have a bigger impact on outcome results, more than occurrence data.

4. Methodology, rather than technology, is the root cause of key differences in inputs and outcomes. Differences in underlying technology do not offer simple explanations, e.g., AI, watermarking, fingerprinting for occurrences and either ACR, STB, or both for exposures.

Attribution results won't be comparable or consistent until providers adopt more stringent media measurement standards and demonstrate competence and fluency in the media space.
It will be impossible for users to build norms across providers and extremely challenging to change providers. It will also be confusing for marketers who receive attribution studies from multiple providers, and definitely risky for media performance-based guarantees.

Importantly, television data streams also bring issues to multitouch attribution, which use the same data inputs … and may adversely impact television ROI and ROAS estimation.

Overall Observations



CIMM's experiment to detect the impact of occurrence and exposure data in television attribution highlights an array of fundamental technical and methodological issues:

- Comparisons of commercial occurrence counts across providers can be misleading. The combination of undercounts and overcounts can obscure these inaccuracies, and a simple count of spots to confirm the accuracy of third-party occurrence data will not work.

- There is no pattern of occurrence data discrepancies by schedule and no explanation of the cause of those discrepancies to be found here.

- There is no standard naming, or coding, of television networks among providers. This makes data comparisons, aggregations or meta-analyses difficult.

- There is no standard definition of a day. Some providers use a broadcast day (6:00 am to 5:59 am), whereas others used a clock day (12:00 am to 11:59 pm), again making data comparisons difficult.

- There is no standard category for commercial length among the providers. Some providers group similar lengths into larger categories. Some use the exact length upon airing, even if it was cut short. This is an obstacle to comparing findings across providers, developing normative data bases and conducting meta-analyses.

- Some spots are more difficult to discern than others. For example, a :15 and a :30 that both use the same video content may be hard to distinguish. For this reason, any provider may exhibit different levels of accuracy for different schedules.

- Network and advertiser post buy logs are not perfect sources of commercial occurrence data and should be validated.

- Underlying technology alone does not explain occurrence differences among providers. And it certainly does not explain the differences found within each provider by schedule.

- Differences in commercial occurrence data among providers result in differences in schedule GRPs and Reach that are not easily predicted by the occurrence data differences.

- Attribution differences generated by differences in commercial occurrence data inputs are directionally consistent, but can exhibit meaningful differences in magnitude. This suggests little reason to be concerned about tactical optimization applications, but significant concern with respect to the risks presented by variation in magnitude and how that will impact ROAS or ROI estimates.

- Conversely, attribution differences generated by differences in exposure data inputs varied dramatically in both direction and magnitude. Clearly, exposure data sources are a major point of origin for differences among attribution providers.

- Another view into the differences in exposure data among the providers was provided by comparison of standard media metrics: GRPs, Reach and Frequency. Differences in these metrics, for the same schedule, across providers were as high as 2:1 for Reach, 4:1 for Frequency and 6:1 for GRPs.

- Grouping providers based on their underlying technology—ACR only, set-top box only, combinations of set-top box and ACR or set-top box and panel—did not explain those inconsistences. Providers who integrate STB and ACR were somewhat more in line with Nielsen benchmarks, but those findings are not consistent across schedules.

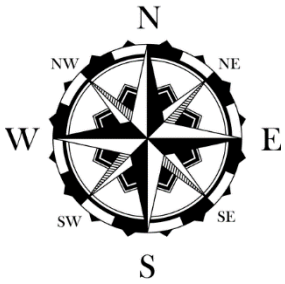
- These findings, together with a review of each providers' current procedures, leads us to conclude that methodology, rather than underlying technology, drives results. In methodology, we include providers' governing reporting rules such as how exposure is qualified (number of seconds) and how reporting panels are managed, edited and weighted to reflect the total US.

While useful learning, none of these findings tell us how to solve the problem. There is no pattern to the discrepancies we found, no "right provider" versus "wrong provider." Everyone was right sometimes and wrong sometimes; even our "truth set." There is no better or worse technology. Hence, no simple answer. But we have seen this before in media research. It is clear to us that what is needed is standardization of naming, definitions and categorization, and more careful quality assurance procedures. This study brought home the challenges presented by shortfalls on both of these fronts. Looking back over this experience, it becomes clear that this was not just the context of our work—it is the key finding. This leads us to three recommendations:

- **For the industry**: Establish standards to ensure that data is organized comparably across providers.

- **For users**: Before proceeding with an attribution study, make sure your occurrence data are validated. This might require using a specific validation study to ensure your data inputs are accurate. Otherwise, while directional guidance for tactical optimization may be trustworthy, ROAS or ROI estimates may be risky.

- **For providers:** Test your QA procedures to ensure accuracy and be prepared to adopt industry standards as they are developed.



Observations on Attribution Data Providers

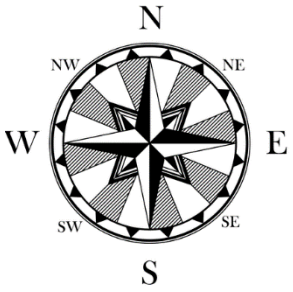
All providers were very cooperative through this process, and CIMM is appreciative of their contributions.

Providers are blessed with great, large data sets, cutting-edge technologies and experienced data scientists. We weren't looking for perfection. Perfection is not the objective, especially in this early stage of development. Besides, CIMM fosters, rather than stifles, media industry innovation.

That said, we observed different levels of traditional media expertise among the providers, different levels of interest in aligning with industry methodological standards and, at times, a lack of attention to rigorous cleansing, proofing and editing data streams.

We see the need for methodological standardization and quality control, while allowing providers' points of difference to remain intact.

Observations for Users of TV Attribution



Expect and plan for differences from provider to provider. Users should not expect that two providers measuring the ROI/ROAS of the same campaign will yield similar results. If users are planning to test or switch to a new provider, then it would be beneficial to benchmark some historical schedules and key inputs to understand differences between the current and new provider.

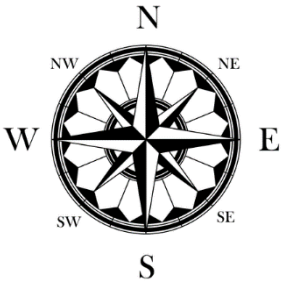
Attribution will likely require "As Run" schedules. Care must be taken to ensure that accurate occurrence data is used for ROI/ROAS studies. Our study found that advertiser post logs and network post logs are not perfect and must be audited. A process of reconciling provider logs with advertiser/network logs must be implemented prior to executing the final analysis.

Confirm detection and categorization of occurrences by commercial length and daypart, if relevant. Part of the audit needs to include analysis by commercial length and daypart, if measuring conversion by those attributes is an important part of the study being executed.

Before running attribution, make sure the data inputs match the world as you know it:

- GRPs by week
- Reach and Average Frequency

It will be valuable to produce comparisons of key exposure data elements, compared to benchmarks such as Nielsen, prior to running the actual ROI/ROAS calculation, so differences can be identified and explained.



Observations for TV Attribution Providers

Having many different TV attribution providers in the market, with different underlying occurrence and exposure data sets, provides challenges for all users of TV attribution work—advertisers, agencies and media companies. It would be preferable for the market to have more consistency in inputs, and then the differences between providers would be primarily driven by the quality of the actual ROI/ROAS calculation. Television attribution results will be more consistent and reliable when providers adopt more stringent media measurement standards:

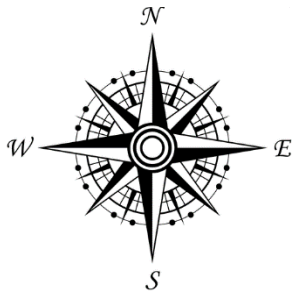
- **Weighting.** Consider implementing a robust panel weighting scheme that addresses key variables known to align with TV viewing: DMA, HH size, Presence of Children, Income, Education and Occupation.

- **Unification.** Consider creating a standard process for unifying your database for ROI measurement, and provide a common base of people with opportunity for exposure and opportunity for response.

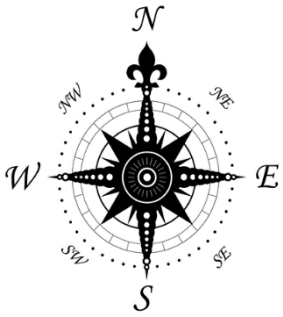
- **Reach.** Conduct evaluation of reach reporting from your exposure data across schedules. Compare to industry norms at different GRP levels (i.e. reach of prime time TV schedule at 300 GRPs). Consider using Reach as a weighting or calibration variable.

- **Exposure qualification.** Having many different measures of viewing time required for exposure in the market creates another source of confusion and differences between providers. If the market will not settle on one standard, then potentially report ROI/ROAS based on multiple exposure measures to allow for cleaner comparisons across providers.

Agenda for Future Studies



- Investigate occurrence differences across providers and advertiser/network logs to identify quality control solutions.
- Evaluate how choices of methodology (e.g. exposure qualification and panel weighting, unification and management) impact:
    - Average Rating
    - GRPs
    - Reach
    - Average Frequency
    - ROI/ROAS
- Examine the impact of clock drift and signal latency impact on those same measures.
- Evaluate digital video/display/CTV exposure data inputs and how those exposures are linked to linear TV exposures.
    - Include an analysis of Identity graphs and evaluate what steps, if any, are taken to account for issues of non-matching.
- Evaluate how results differ with various Modeling approaches (e.g. attribution window, adstock, baseline, etc.).

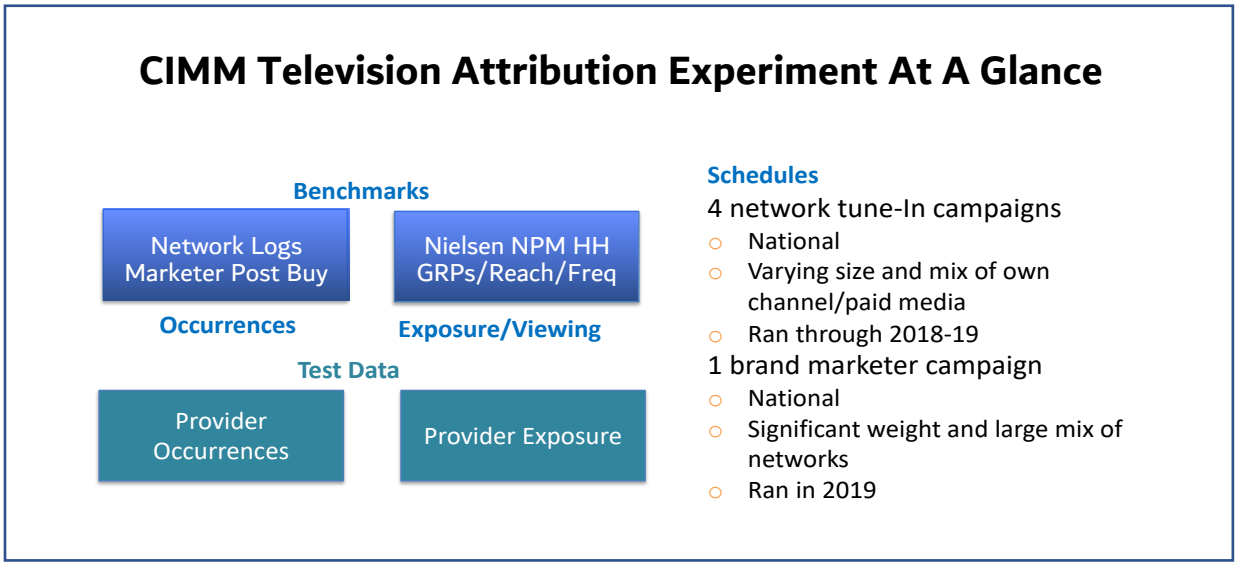


The Study Approach

CIMM conducted an experiment to compare two key components of television attribution: ad occurrences (schedules) and exposure data (GRPs/Reach/Frequency). This experiment was designed to understand whether the different sources modelers use for these data generally over- or undercount occurrences and exposure and, in turn, the degree to which they impact model results and decisions marketers subsequently make.

It's important to keep in mind that this is an experiment, not an assessment of the offerings from television attribution providers. We were only interested in their standard occurrence and exposure data, which we compared to logs and Nielsen national TV benchmarks. We also created a performance measurement (lift-like, measuring change in rating for promoted program compared to 4-week time period average, for exposed minus not-exposed) to compare the impact of these data services. Importantly, we are not advocating providers use our methodology—it simply provided a method with which to make common comparisons.

- For the test, we secured television tune-in campaign schedules from three networks. Why tune-in? Television tune-in campaigns offer a relatively straightforward test scenario given that the input (advertising weight) and outcome measures (viewership to the promoted program) operate in a closed loop within the same data set.
- We also studied **one brand advertiser** campaign.
- All the campaigns ran nationally on linear television within the past 2 years.



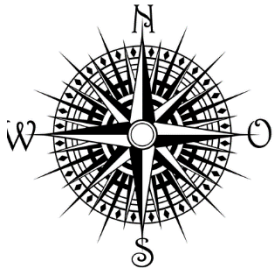
We are very grateful for the participation of the following providers: **605, Alphonso, Ampersand, Comscore, iSpot, NCS, Samba TV, TVSquared and VideoAmp. We are also grateful for the participation of Nielsen in providing the benchmarks this study utilized.**

These providers utilized data from occurrence providers such as iSpot, Hive, Kantar, and Nielsen as well as their own proprietary approaches. They sourced viewing/exposure data from **set-top box providers, Automatic Content Recognition (ACR) systems and Nielsen.**

Providers were compared on schedule accuracy, GRPs, Reach and Frequency generated by the exposure data and Incremental Actual Rating Point Change, the outcome variable we created to compare television tune-in campaigns. All of the providers' results were blinded because the point of the study was to compare input variables, not to evaluate the providers' offerings. Importantly, we do not have all data from all providers. Providers C and I only provided occurrences. Provider F only provided data for 3 schedules.

We believe this array of providers, sources and measures provides a solid representation of the television attribution practices in place today.

Part 1. Deep Dive Into Occurrence Data in Attribution



Advertising occurrence data is the entry point for television (or any video) into TV attribution, multitouch attribution, or any device-level measure of television advertising performance. Advertisers run ads, viewers are exposed to them, and viewers either respond in the marketplace or not. Attribution models connect the dots from the ads that are run to the marketplace response they generate.

Users of attribution assume the ad occurrence data used for television attribution are accurate. This phase of the study tests that assumption. Perfection is never the goal, so the question is really, are the television occurrence data in use for attribution sufficiently accurate to enable useful television attribution findings?

Objective

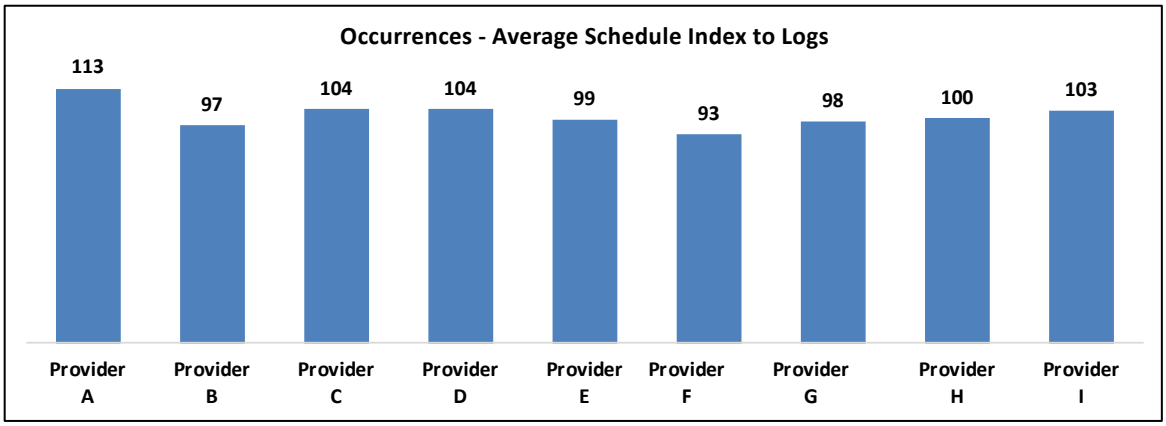
The objective of this phase of the study is to reveal how different TV occurrence data sets may be among a cross section of leading data providers and to what degree do those differences result in differing attribution results. As a reminder, the "As Run" network or advertiser post logs were used as benchmarks.

Comparing Occurrence Counts

The analysis plan was to start at the most summarized level, and then to peel away successive layers in search of data patterns that explain any discrepancies in the data and might indicate best practices.

At the most summarized level, our findings were reassuring. When we compare the count of total occurrences for each campaign from the logs to the counts provided by each provider, we find a 101 index, only a 1% difference. Of course, each provider's index is higher or lower, but overall, there is no tendency to under- or overstate the total number of occurrences.

The chart below displays occurrence count indices for each provider. These compare the differences in occurrence counts between the provider's data and the log data, by campaign, averaged among campaigns.



All but one provider's results are within 10% of the logs; five are within 5%.

Next, we took a deeper look by computing match rates. Individual occurrences from the logs were matched to the individual occurrences from each provider. Spots were matched based on network, date and time, allowing for clock drift (± 5 minutes). This analysis revealed three important findings:

- There is no standard naming, or coding, of networks among providers. This makes data comparisons, aggregations or meta-analyses difficult.

- There is no standard definition of a "day." Some providers use a broadcast day (6:00 am to 5:59 am), whereas others used a clock day (12:00 am to 11:59 pm).

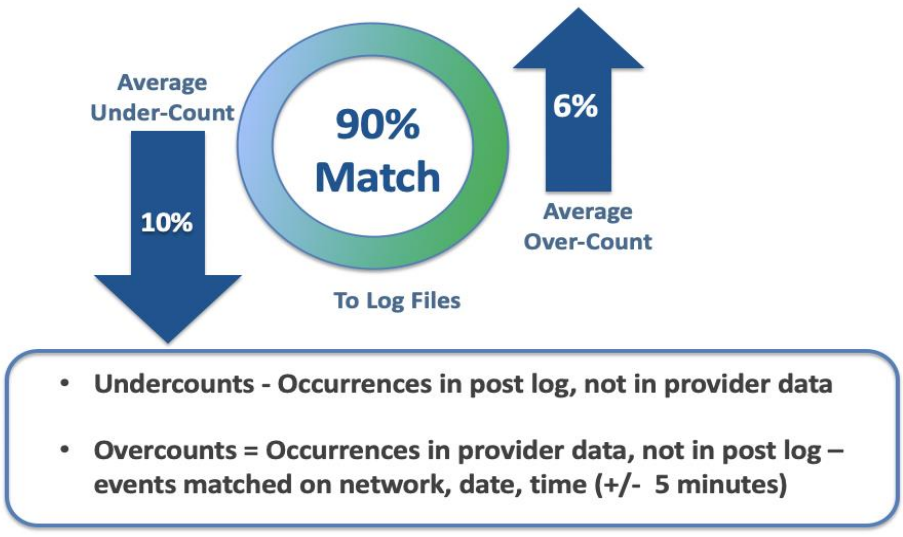
- Comparisons of commercial occurrence counts can be misleading. The following chart shows that occurrence count indices and match rates reveal different views of provider accuracy. Consider Provider E, with a 99 index on occurrences relative to the log, but only a 79.3% match rate. While this is the most extreme disconnect, it's not the only one.

	Provider A	Provider B	Provider C	Provider D	Provider E	Provider F	Provider G	Provider H	Provider I
Provider Occurrence Data Indexed to Network Logs									
Average Among All Campaigns	113	97	104	104	99	93	98	100	103
Legend:	120+	110-119	90-81	80-					
MATCH RATE									
Average Among All Campaigns	76.0%	91.0%	95.3%	95.2%	79.3%	91.1%	96.0%	96.3%	91.3%
LEGEND:	<90%	<75%	<50%						

Overall Match Rates Between Post Logs & Provider Occurrences

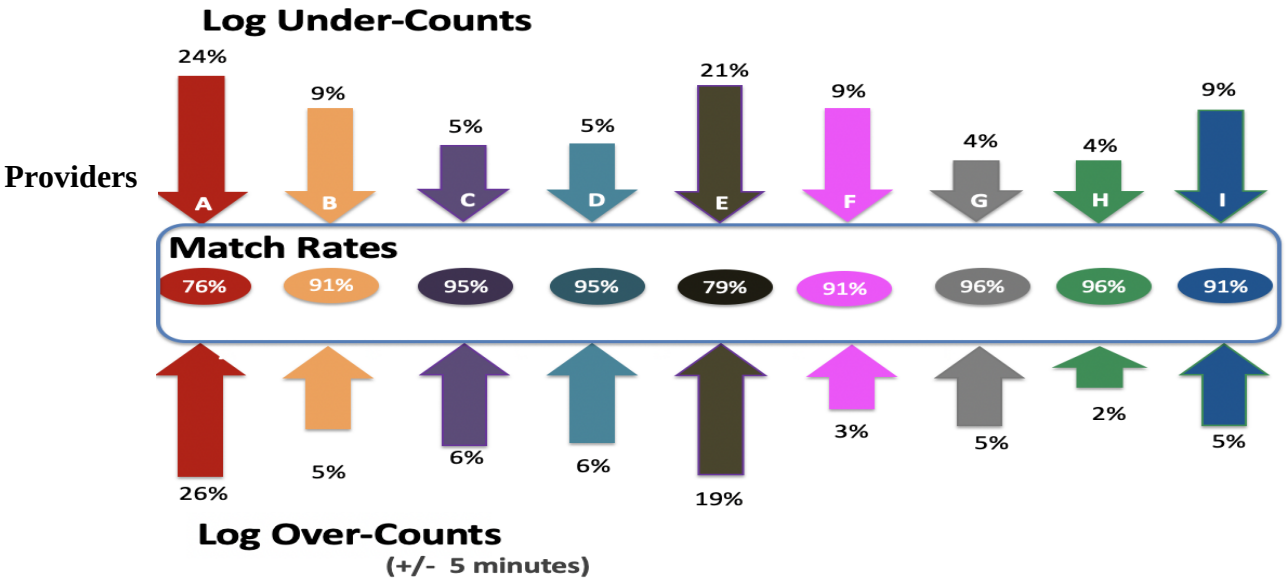
Summarizing the match rate analysis at the highest level, we find promising findings. We also discover the reason for the disconnect between match rates and occurrence count comparisons.

Overall, across all providers and all campaigns, the average match rate was 90%. The complement to that is that there was an average undercount of 10%. In other words, on average, our analysis found 90% of the spots from the logs in the providers' data and 10% of the spots from the logs were not found in the providers' data. But the analysis also found spots in the provider data that were not in the logs—overcounts. Overall, overcounts were equal to 6% of the number of spots in the logs, on average.



Breaking out the match rates, undercounts and overcounts by provider and averaging results across campaigns, we find the results below.

Occurrence Match Rates & Under/Over Counts Vs. Post Logs by Provider

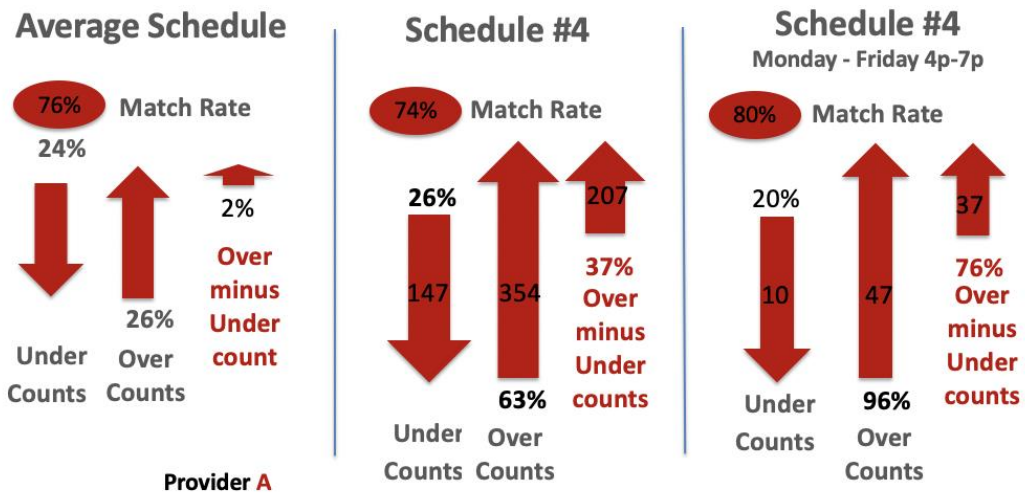


These results show us that:

- Accuracy of ad occurrence data varies by provider.
- The combination of undercounts and overcounts can obscure these inaccuracies. You can't depend upon a simple count of spots to confirm the accuracy of third-party occurrence data, especially if you are looking at daypart performance.

This results also beg the question, are at least some of the undercount and overcount spots actually the same spots, just not matched through some flaw in the analysis?

To assess this question, we dove more deeply into the Match rate details. Here's a picture of the approach:



In this example, we looked at Provider A's occurrence data. Overall, Provider A had a match rate of 76%, which left 24% undercounts. But they also had 26% overcounts. Were some of those the same spots? Looking at Provider A's results for schedule 4, we find a different relationship between undercounts and overcounts—37% different. There are not enough undercounts to match even half the overcounts, so they can't be the same spots, just mismatched. When we go a level deeper, looking only at spots for schedule 4 in the Monday-Friday 4 pm-7 pm daypart, we see a 76% difference between undercounts and overcounts. Almost 80% of the overcounts could not be matched to the undercounts—there are just not enough of them.

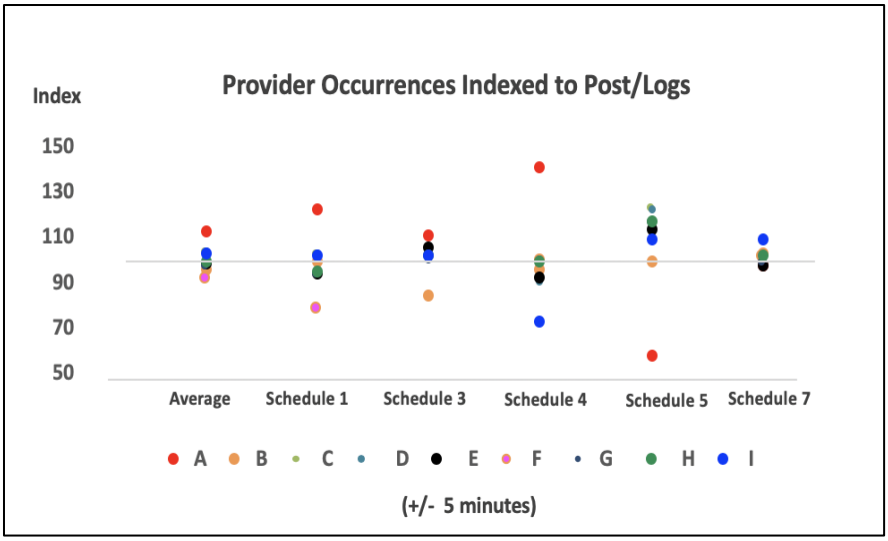
Running this analysis for all providers, schedules and dayparts, we find overcounts exceed undercounts by 10%-20%, or 20% or more in more than 50 cases and undercounts exceed overcounts by 10%-20%, or 20% or more in another 50+ cases.

	PROVIDER								
	A	B	C	D	E	F	G	H	I
Campaign # 1									
Monday - Friday 6a-9a	2%	1%	1%	1%	8%	-14%	-1%	-1%	1%
Monday - Friday 9a-12n	16%	1%	1%	1%	-6%	-15%	-1%	-1%	1%
Monday - Friday 12n-4p	14%	2%	2%	0%	-10%	-20%	-1%	-1%	1%
Monday - Friday 4p-7p	26%	1%	2%	2%	15%	-16%	-3%	-2%	5%
Saturday - Sunday 6a-12n	-3%	0%	0%	0%	-13%	-18%	-2%	-2%	0%
Saturday - Sunday 12n-7p	2%	-1%	1%	1%	-9%	-20%	-8%	-9%	1%
Monday - Sunday 7p-11p	2%	-1%	1%	1%	-15%	-20%	-5%	-5%	1%
Monday - Sunday 11p-1a	17%	0%	3%	5%	-9%	-16%	-1%	2%	1%
Monday - Sunday 1a-6a	30%	0%	2%	2%	4%	-15%	0%	1%	2%
Campaign #3									
Monday - Friday 6a-9a	46%	-12%	0%	0%	-1%	-1%	-1%	-5%	-5%
Monday - Friday 9a-12n	9%	-12%	1%	1%	1%	0%	0%	-5%	-5%
Monday - Friday 12n-4p	-1%	-16%	-5%	-4%	0%	-6%	-5%	-5%	-5%
Monday - Friday 4p-7p	4%	-15%	-2%	-1%	-1%	-5%	-3%	-4%	-4%
Saturday - Sunday 6a-12n	9%	-11%	3%	4%	-1%	1%	1%	2%	2%
Saturday - Sunday 12n-7p	0%	-17%	2%	0%	0%	-2%	-2%	1%	1%
Monday - Sunday 7p-11p	-1%	-18%	-4%	-4%	-1%	-6%	-5%	-2%	-2%
Monday - Sunday 11p-1a	9%	-21%	-4%	-4%	-9%	-2%	-7%	-1%	-1%
Monday - Sunday 1a-6a	1%	-15%	0%	0%	0%	-5%	0%	2%	2%
Campaign #4									
Monday - Friday 6a-9a	44%	0%	-7%	-7%	-2%	-2%	-2%	-2%	-20%
Monday - Friday 9a-12n	4%	-6%	-10%	-12%	-3%	1%	1%	1%	-21%
Monday - Friday 12n-4p	41%	-3%	-14%	-11%	-3%	0%	0%	0%	-25%
Monday - Friday 4p-7p	76%	-2%	-14%	-14%	-10%	-2%	-2%	-2%	-16%
Saturday - Sunday 6a-12n	31%	-12%	-12%	-15%	-8%	4%	4%	4%	-31%
Saturday - Sunday 12n-7p	36%	-3%	-3%	3%	-3%	0%	0%	0%	-28%
Monday - Sunday 7p-11p	23%	-6%	-6%	-11%	-14%	-1%	-1%	-4%	-26%
Monday - Sunday 11p-1a	35%	2%	-4%	13%	2%	5%	11%	11%	-40%
Monday - Sunday 1a-6a	51%	-2%	-3%	-8%	-3%	1%	-2%	1%	-29%
Campaign # 5									
Monday - Friday 6a-9a	-37%	-16%	7%	7%	2%	0%	4%	4%	-35%
Monday - Friday 9a-12n	-52%	-9%	8%	8%	1%	0%	4%	4%	1%
Monday - Friday 12n-4p	-42%	-7%	10%	10%	7%	0%	7%	7%	5%
Monday - Friday 4p-7p	-30%	6%	22%	22%	18%	0%	18%	18%	16%
Saturday - Sunday 6a-12n	-53%	16%	17%	17%	16%	0%	16%	16%	16%
Saturday - Sunday 12n-7p	-51%	0%	16%	16%	9%	0%	12%	12%	8%
Monday - Sunday 7p-11p	-34%	-2%	19%	19%	7%	0%	12%	8%	9%
Monday - Sunday 11p-1a	-34%	8%	17%	17%	13%	0%	18%	23%	11%
Monday - Sunday 1a-6a	-39%	-1%	10%	10%	7%	0%	9%	11%	-9%
Campaign # 7									
Monday - Friday 6a-9a	30%	-7%	-1%	-1%	-14%	-4%	-4%	-4%	1%
Monday - Friday 9a-12n	3%	-6%	-3%	-3%	-11%	-3%	-3%	-2%	1%
Monday - Friday 12n-4p	2%	8%	-2%	-2%	4%	-2%	-2%	-2%	10%
Monday - Friday 4p-7p	32%	24%	22%	22%	55%	10%	10%	10%	34%
Saturday - Sunday 6a-12n	-12%	0%	-2%	-2%	-15%	-5%	-5%	-5%	-1%
Saturday - Sunday 12n-7p	-6%	3%	0%	0%	4%	3%	-4%	3%	4%
Monday - Sunday 7p-11p	-12%	9%	1%	1%	3%	2%	1%	-1%	15%
Monday - Sunday 11p-1a	-14%	-7%	-10%	-10%	-14%	-7%	-4%	-2%	-3%
Monday - Sunday 1a-6a	-9%	2%	0%	0%	-6%	2%	-2%	2%	3%

If overcounts were simply the unmatched undercounts, then the difference in each cell of this table would be close to zero. In contrast, we see the lack of matches and the extreme in either overcounts or undercounts in the data for every provider, schedule and daypart. The factors causing these discrepancies are not isolated to a deficient provider, a problematic schedule or a difficult to measure daypart. No clear, potentially causal, pattern has yet emerged in this analysis, as frustrating as that is.

LEGEND	
Over counts exceed undercounts	
20% or more	32%
10%-20%	15%
Undercounts exceed over counts	
20% or more	-39%
10%-20%	-14%

Match Rates by Schedule



Rolling back up to the schedule level, we find no consistency within each provider. This chart is based on provider occurrence counts indexed to logs, by schedule. Provider A, for example, had the highest index on average (113), but you can see that their average is composed of high indices for schedules 1 and 4, a low index for schedule 5 and average, or close to average indices for schedules 3 and 7. This lack of consistency is more or less evident for every provider. No provider is consistently high or low on occurrences versus the post logs.

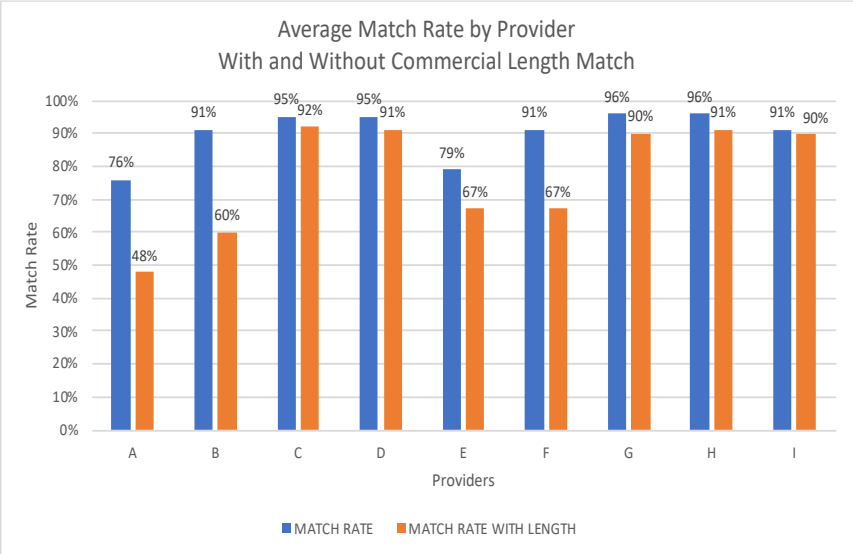
	Provider A	Provider B	Provider C	Provider D	Provider E	Provider F	Provider G	Provider H	Provider I
MATCH RATE With NO DURATION									
Average Among All Campaigns	76.0%	91.0%	95.3%	95.2%	79.3%	91.1%	96.0%	96.3%	91.3%
Campaign #1	93.8%	98.8%	99.8%	99.8%	59.3%	81.7%	96.3%	96.4%	100.0%
Campaign #3	90.8%	81.4%	94.7%	94.1%	90.2%		93.9%	94.0%	95.5%
Campaign # 4	73.9%	94.1%	89.5%	89.5%	92.5%	97.7%	97.7%	97.7%	71.0%
Campaign #5 *	39.4%	86.7%	99.2%	99.2%	94.2%		99.3%	99.3%	91.7%
Campaign #7	82.2%	93.8%	93.6%	93.5%	60.3%	94.0%	92.7%	94.0%	98.0%
LEGEND:	<90%	<75%	<50%						
OVERCOUNT RATE									
Average Among All Campaigns	26%	5%	6%	6%	19%	3%	5%	5%	5%
Campaign #1	18%	1%	2%	2%	36%	1%	1%	2%	1%
Campaign #3	15%	3%	4%	5%	8%		3%	3%	3%
Campaign # 4	63%	2%	3%	3%	2%	3%	3%	3%	3%
Campaign #5 *	20%	12%	15%	15%	14%		12%	12%	12%
Campaign #7	14%	8%	6%	6%	36%	6%	5%	6%	8%
LEGEND:	10% -19%	20%-49%	50%+						
UNDERCOUNT RATE									
Average Among All Campaigns	24%	9%	5%	5%	21%	9%	4%	4%	9%
Campaign #1	6%	1%	0%	0%	41%	18%	4%	4%	0%
Campaign #3	9%	19%	5%	6%	10%		6%	6%	4%
Campaign # 4	26%	6%	10%	10%	7%	2%	2%	2%	29%
Campaign #5 *	61%	13%	1%	1%	6%		1%	1%	8%
Campaign #7	18%	6%	6%	6%	40%	6%	7%	6%	2%
LEGEND:	10% -19%	20%-49%	50%+						

All but two of the providers exhibited low match rates for at least one schedule, despite average match rates on average among all schedules. We also find high overcount and undercount rates for all providers for at least one schedule. This analysis reaffirms that discrepancy rates vary among providers and within provider by schedule. Once again, there is no clear pattern.

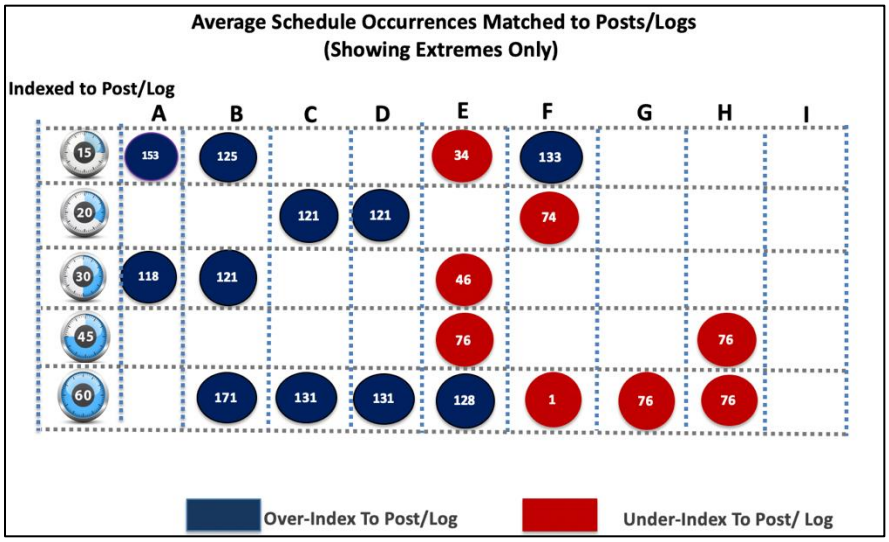
We found no clear explanation for these discrepancies and conclude that there is simply no pattern of occurrence data discrepancies by schedule—and no explanation of the cause of those discrepancies to be found here. Issues are found with data from all providers.

Match Rates by Commercial Length

Continuing to search for the source of commercial occurrence data discrepancies, we turned to commercial length. The match rate analysis was re-run with commercial length as an added criterion. All match rates dropped, sometimes precipitously. For example, Provider A's match rate dropped from 76% to 48% and Provider F's match rates dropped from 91% to 67%. Others showed modest declines. In every case, there was disagreement between the logs and the provider's occurrence data. In some cases, this disagreement was extreme.



Looking at the commercial length issue from the perspective of overcounts and undercounts, we see that all but one provider had difficulty with commercial length. This chart shows the index of overcounts and undercounts by provider, by selected commercial lengths, on average across all campaigns. Only one provider did not exhibit extreme undercounts or overcounts. But even that provider had difficulty with one campaign.



Discussing the length issue with the providers we learned two important findings:

- There are no standard categories for commercial length among the providers. Some group similar lengths into larger categories. Some report the exact number of seconds the spot ran, even if it was

a cut-short, reporting a 30-second spot as a :25, for example. This is clearly an obstacle to comparing findings across providers, developing normative data bases and conducting meta-analyses.

- Some spots are inherently more difficult to discern than others. For example, a :15 and a :30 that both use the same video content may be hard to distinguish. For this reason, any provider may exhibit different levels of accuracy for different schedules.

Accuracy of Network Logs and Post-Buys

Throughout these analyses, we found the number of overcounts puzzling. We investigated by matching overcounts across providers. For example, if Provider A found a spot on ABC at 9:05 pm on 3/17/20, which was not in the network/advertiser log, then how many other providers also found that same overcount spot? A significant number of overcounts were identified by as many as seven providers. This finding supported the theory that the network/advertiser logs were not a perfect truth set. A threshold of at least three providers was set and the number of "common overcounts" was tabulated for each schedule. The range of common overcounts, potentially actual spots missing from the log, ranged from 1% to 10%.

Over Counts Identified By 3 Or More Providers		
Schedule	# of Over Counts Found by 3+ Providers	% of # Spots In Log
1	89	1%
3	134	2%
4	17	3%
5	200	10%
7	77	3%

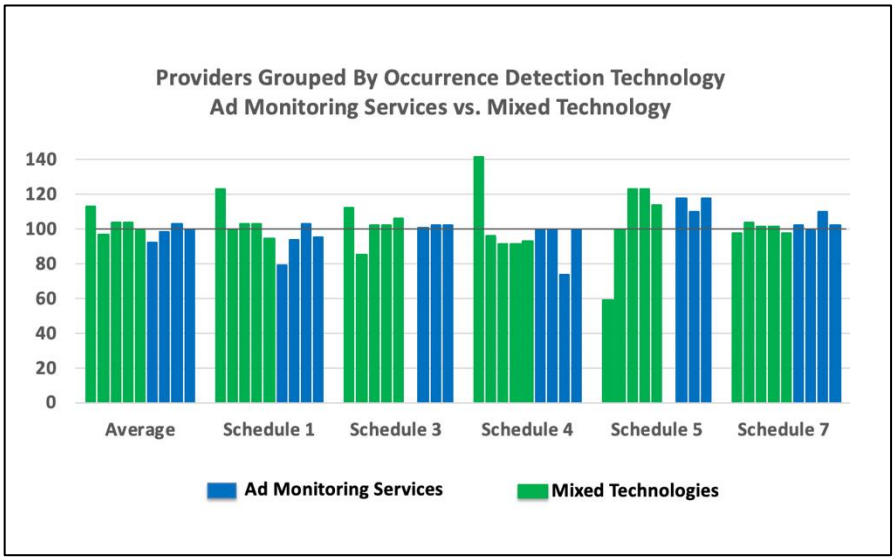
This reminds us that, as is so often the case in our big data world, we are using data for measurement purposes that were not designed for measurement. Network logs and advertiser logs may be fine for their intended purpose, but are imperfect for attribution. This gives us our surprising finding that:

- Network and advertiser logs are not perfect sources of commercial occurrence data and should be validated.

Are Data Differences Driven by an Underlying Technology Bias?

One common hypothesis has been that the underlying data-gathering technologies introduce bias into the measurement system. The fact that we found no consistent patterns by provider appeared to belie that theory, but we were able to consider it directly.

This chart compares each providers' index to the logs for each schedule and on average among all schedules. The blue bars represent traditional fingerprint-based monitoring services with monitoring stations in all markets. The green bars represent a mixture of ACR and AI–based technologies. As is visually evident, there is as much variation within a technology-based provider group as across groups. We concluded that:



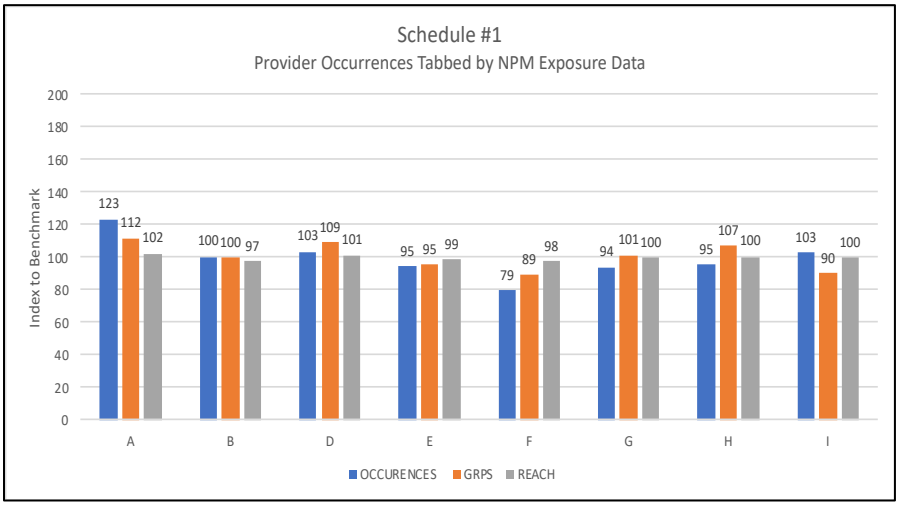
- Ad detection technology alone does not explain occurrence differences among providers. And it certainly does not explain the differences found within each provider by schedule.

Do Data Discrepancies Matter?

This phase of the study has revealed substantial discrepancies among provider occurrence data and between provider occurrence data and the post log data. These findings beg the question: Do these discrepancies matter? This question is addressed in two ways. First, through the lens of standard media metrics—GRPs and Reach. Second, through the lens of attribution, using a common "lift" calculation.

Attribution providers are not in the business of providing their clients with GRP or Reach metrics. But these are the most common ways of dimensionalizing a television schedule. They provide a measure of the schedule that is highly relevant to attribution—how many households were reached by the schedule and how many individual impacts occurred. If more or fewer households were reached, if there are more or fewer GRPs, then the schedule will have greater or lesser impact on market performance. We also looked at Average Rating and Average Frequency and could have looked at Impressions, but those measures are themselves defined by occurrences, GRPs and Reach, so they do not add new information to this analysis.

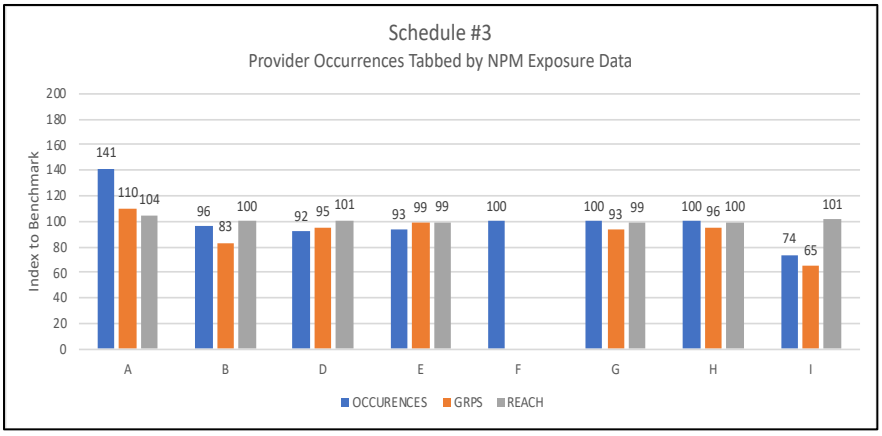
In the following charts we see the differences in occurrences (blue bar), GRPs (orange bar) and Reach (grey bar) by provider for each schedule. For this analysis, we tabulated GRPs and Reach using provider occurrences and Nielsen National People Meter exposure data. Using a common source of exposure data enables us to isolate the effect of the exposure differences, expressing them in terms of exposure metrics.

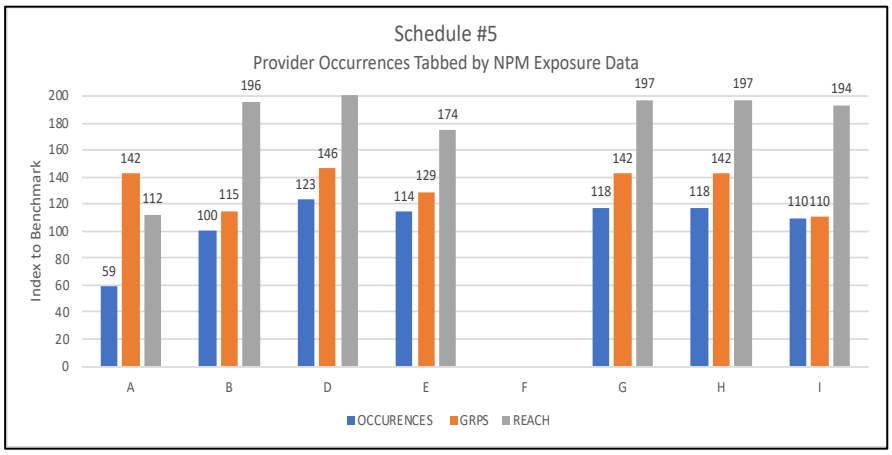


It's easy to see that the providers' differences in occurrences also result in differences in GRPs for each schedule. But these differences are not the same. For example, for schedule 1, Provider A, occurrences were overstated by 23%, but GRPs were only overstated by 12%.

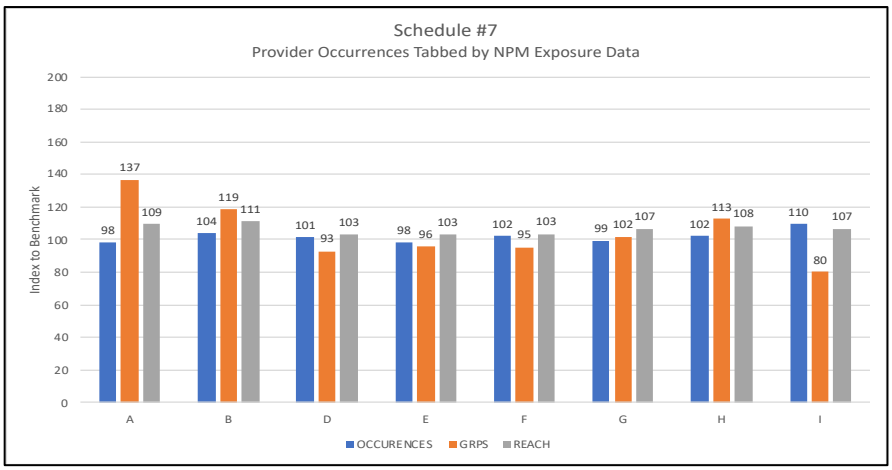
In some instances, the GRP indices are more extreme than the occurrence indices, and in other cases less extreme.

This tells us that the schedule occurrences are varied among providers and result in higher or lower rated spots being included or excluded. Differences in Reach are less extreme because it is bounded by 100% at the high end, unlike GRPs. But we still find differences in Reach generated by differences in occurrences that may be more or less extreme than the occurrence differences themselves.





Note that the extreme indices for schedule 5 are due to a significant number of spots missing from the network log, which was used to create the benchmark.



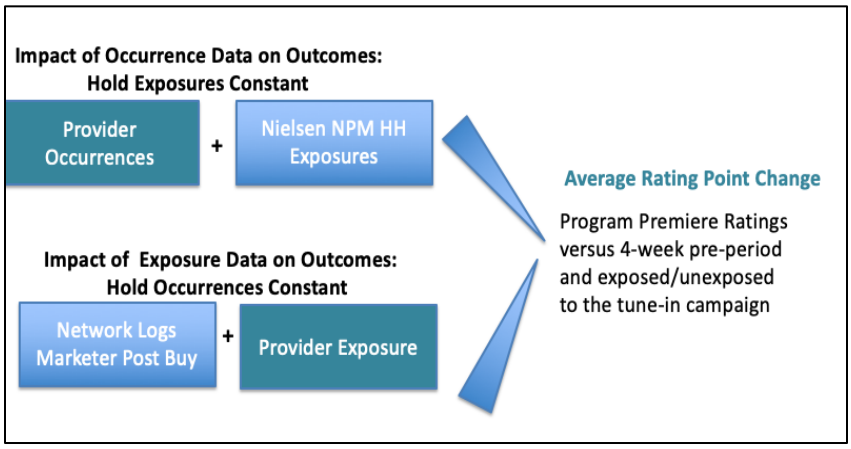
From this analysis we conclude:

- Differences in commercial occurrence data among providers result in differences in schedule GRPs and Reach that are not predictable or comparable to the occurrence data differences.

Assessing the Impact of Occurrence Data Using Custom "Lift" Method

The final answer to the "so what" question is to evaluate the impact of commercial occurrence data discrepancies on attribution itself. In keeping all other things equal, we must also keep the attribution model, or algorithm, equal across providers. So we created a custom "lift-like" measure of impact.

To be perfectly clear, this study is not a comparison of the modelers' attribution models, only of their occurrence and exposure data. There are many differences between our approach and what attribution providers offer, but this suited our experimental purposes. Keep in mind we studied only the television tune-in campaigns because we were able to see the impact of the campaigns on ratings to the promoted program in the same data set. Working at the aggregate level and comparing ratings, not individual household viewing behavior, we used the Difference of Differences method with the network, day of week and time of day as the unit of analysis. A 4-week pre-period was defined and its average rating was obtained for the exact network, day of week and time of day as the promoted program.



The difference or "lift" for the promoted program over the same network/time-period during the pre-period was calculated separately among households exposed and households not exposed to the schedule. Then the exposed-unexposed difference was taken.

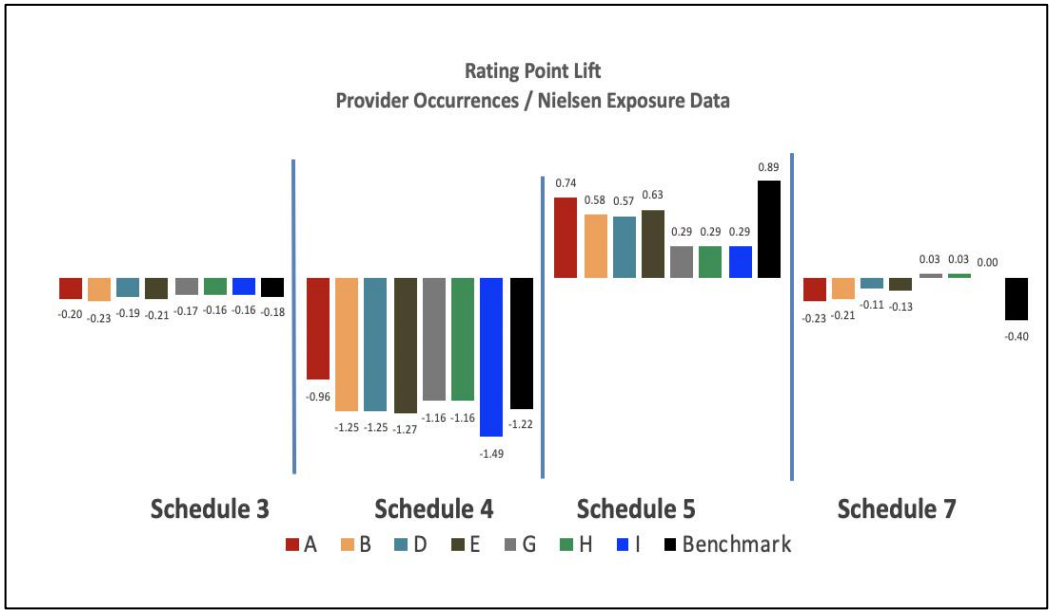
The result is an estimate of rating point change, or "lift over the pre-period," accounting for baseline viewing pattern of the unexposed households. Please note the data providers strenuously objected to our use of the word "lift" to describe the rating point changes from the exposed/unexposed pre/post periods. They do not want there to be any confusion between their lift calculations, which tend to be more intricate, operating at the individual TV set or household level, and this rather straightforward aggregate approach. We respect that.

And we acknowledge there are some flaws in this approach. Most notably, there were situations in which the pre-period—intended to represent the usual viewing levels for the network and time period—were unusual with major sporting events airing. We could have tailored the pre-period data to avoid these issues. If our goal was the best measure of rating point change or lift, then we would have made adjustments. But our goal is comparability, so we didn't. As a result, you will see negative rating point

changes, which is unusual for "lifts." What is important is, whether all the providers' occurrence data produce the same changes … positive or negative.

We gave considerable thought to the possibility that different rating point change/lift calculations would have changed our findings. In particular, would a method using individual household level data with test and control cells balanced for propensity to view have given us different results? Yes, of course it would. But if we used that same method with each providers' occurrence data as the only difference, would we obtain the same findings with respect to differences among providers? We believe so.

That said, this chart shows the rating point differences we derived for each schedule, using the occurrence data from each provider, tabulated against Nielsen NPM exposure data and with lifts calculated as described above. The first thing we see is good directional agreement among providers' and with the benchmark. Everyone finds schedule 5 successful and schedules 3 and 7 to have little effect. For schedule 4, all providers suffer similarly from the pre-period discontinuity mentioned above. Looking past the directional agreement among providers, however, we find material magnitude differences—as much as 2 to 1.



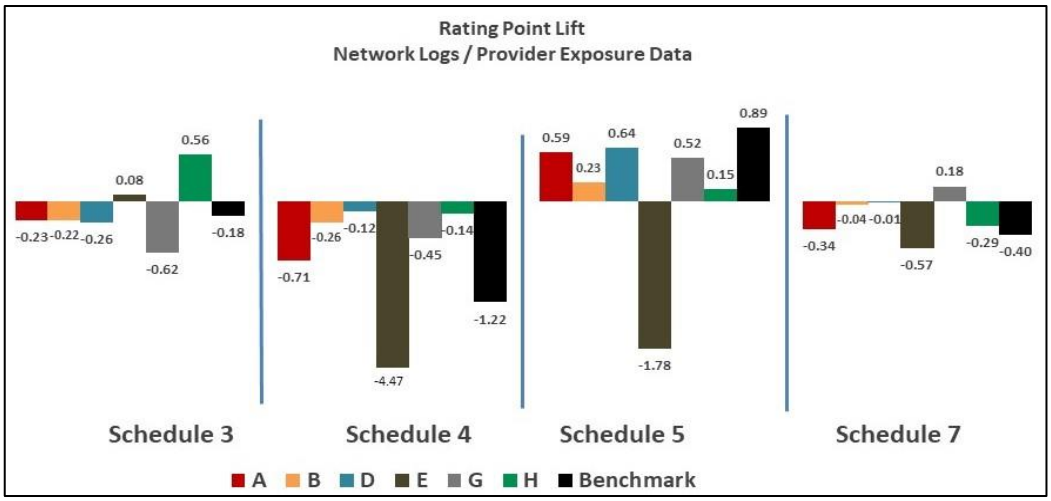
This analysis suggests that:

- Attribution results generated by differences in commercial occurrence data inputs are directionally consistent, but can exhibit meaningful differences in magnitude. This suggests little reason to be concerned about tactical optimization applications, but significant concern with respect to the risks presented by variation in magnitude and how that will impact ROAS or ROI estimates.

Comparing Lift by Provider Using Their Own Exposure Data

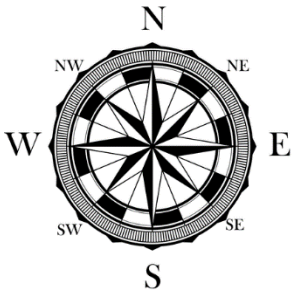
The analysis of attribution results for each of the schedules using a constant exposure data source (Nielsen) and each provider's occurrences showed consistency in lift measurement. Flipping the page, we executed that same comparison, but kept the occurrence data consistent (network/advertiser post logs) and used each provider's exposure data, not Nielsen's. This isolated the impact of exposure data variations on attribution results. In this case, we found dramatic differences among providers.

All providers agreed on the direction of rating point changes for only one campaign. The other three campaigns had at least one provider reporting a difference in the direction relative to the other providers and the Nielsen benchmark. Even in cases where the providers reported the same direction of rating point change, there were many more instances where the magnitude of that change differed dramatically from provider to provider, compared to the Nielsen benchmark, than there were instances where the rating point changes were close to other providers and the Nielsen benchmark.



A clear example of this is schedule 4. Even though all providers and Nielsen reported negative rating point changes, one provider reported dramatically larger negative impact than the others and Nielsen, and five providers reported dramatically less negative impact. None were in line with the Nielsen benchmark. Provider to provider, there is no consistency relative to the Nielsen benchmarks. Each provider reported schedules with both greater impact and less impact than the Nielsen benchmark.

Part 2. Comparing Exposure Data



Approach

The second key input in attribution is media measurement—the exposures or impressions generated by each of the occurrences in the schedule. We set out to determine how variations in exposure data across the providers also impact attribution results.

The media world uses impressions to represent the total number of people who see an ad, but impressions fail to demonstrate important dimensions of how many different people saw the ad and how many times, on average, they were exposed to the ad. Reach and Frequency metrics provide those insights.

Our analysis plan was to evaluate the overall Reach and Frequency from each provider, using the network post-log files to eliminate any discrepancies that were due to differences in occurrence data. Reach is a key metric for TV attribution studies because most vendors will multiply the incremental ROI from households or people exposed by the number of households or people exposed to calculate the incremental ROI from the campaign being measured. Frequency is also important because we know that there is a point of diminishing returns where incremental frequency does not provide incremental ROI.

In addition to evaluating Reach and Frequency, other key metrics such as Average Rating and Total Schedule Gross Rating Points (GRPs) were evaluated. Average Rating tells us, on average, whether each occurrence impacts roughly the same proportion of the population. Schedule Gross Rating Points tells us whether the entire campaign deliver as many impacts, as a percent of the population.

To provide some benchmark comparison, each provider's data was compared to Nielsen National People Meter data, evaluated against the same network post-log files with these basic media metrics.

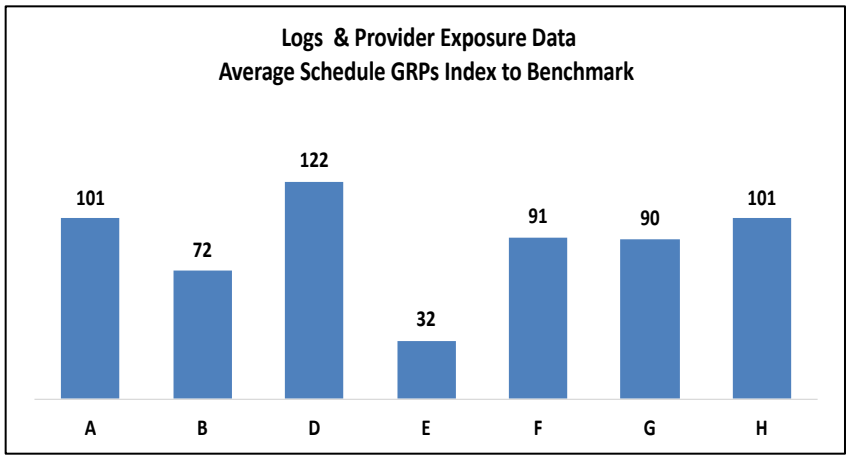
Key Findings

At the most summarized level, our findings were concerning.

- There were inconsistencies across and within provider, across all the schedules and metrics being analyzed.
- We executed an analysis that grouped providers based on the underlying data sets—ACR only, set-top box only, combinations of set-top box and ACR or set-top box and panel—but that did not explain the inconsistences.
- Due to this analysis, we conclude that methodology, rather than underlying technology, drives results. In methodology, we include the provider's governing reporting rules such as how "exposure" is qualified (number of seconds a set is tuned to an ad to qualify as an exposure) and if and how data sets are weighted to reflect the total US.

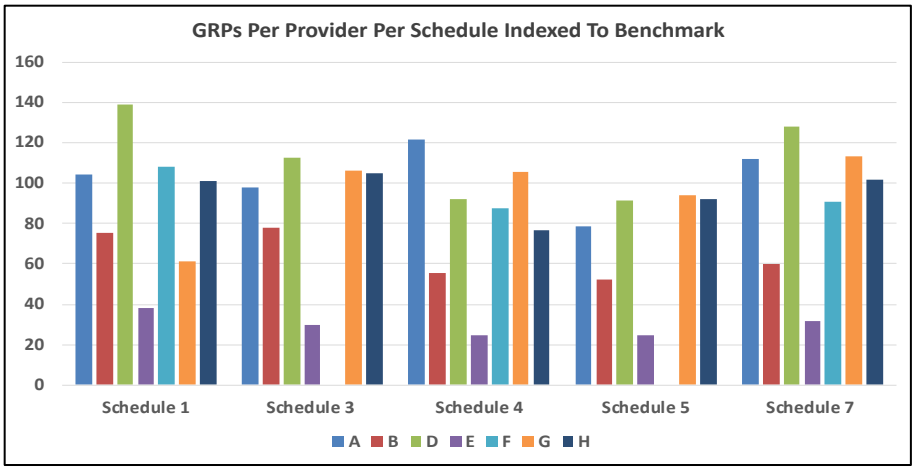
Consistency of Average Schedule GRPs Across Providers

The analysis of average schedule GRPs compared to the Nielsen benchmark shows fairly promising results, with four of seven providers being within 10 index points from the Nielsen benchmark. However, we did identify extremes, with one provider indexing at 32 while another indexes at 122.



Consistency of Individual Schedule GRPs Across Providers

The variance across providers and within provider is clear when we evaluate the index of each individual schedule relative to the Nielsen benchmark. Each provider has a very unique pattern:



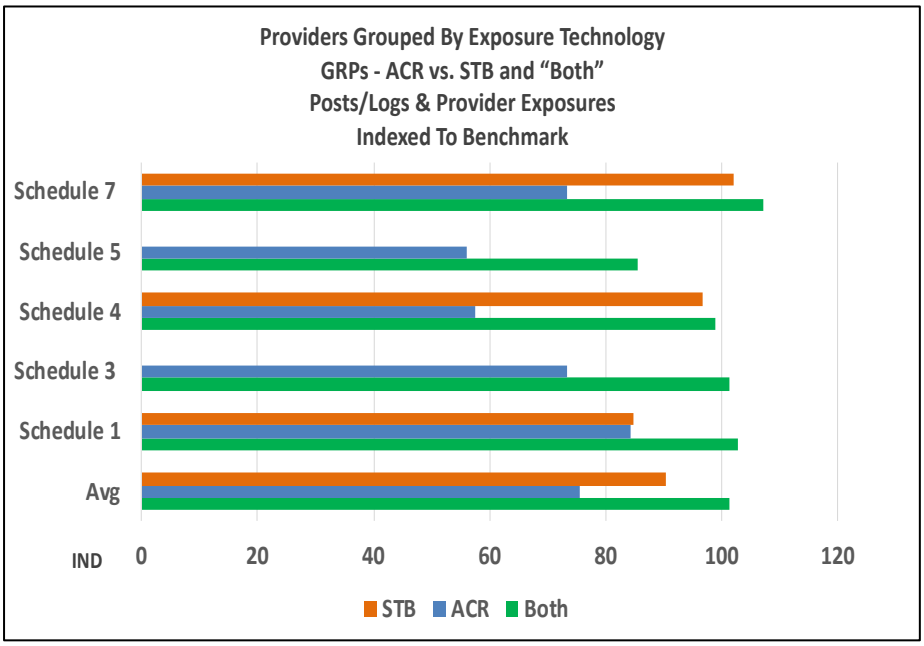
- Provider A: Close to the benchmark for schedules 1, 3 and 7 (close to the 100 IND mark), above for schedule 4, below for schedule 5
- Provider B: Low for all schedules, one schedule with a 50 index
- Provider D: Above the benchmark for schedules 1 and 7, close for schedules 2, 4, and 5
- Provider E: Low for all schedules, schedules 4 and 5 indexing around 25
- Provider F: Close to benchmark on schedules 1 and 7, below for schedule 4
- Provider G: Low for schedule 1, close to benchmark on schedules 3, 4, 5 and 7
- Provider H: Low for schedule 4, close to benchmark for schedules 1, 2, 5 and 7

Do Underlying Sources of Viewing Data Explain Differences in Schedule GRP Levels?

In an attempt to evaluate whether the source of each provider's viewing data is a driver of the differences we saw with schedule GRP levels, we aggregated providers based on their underlying data as follows:

- Providers who utilize Smart TV ACR data only
- Providers who utilize set-top box data only
- Providers who utilize both

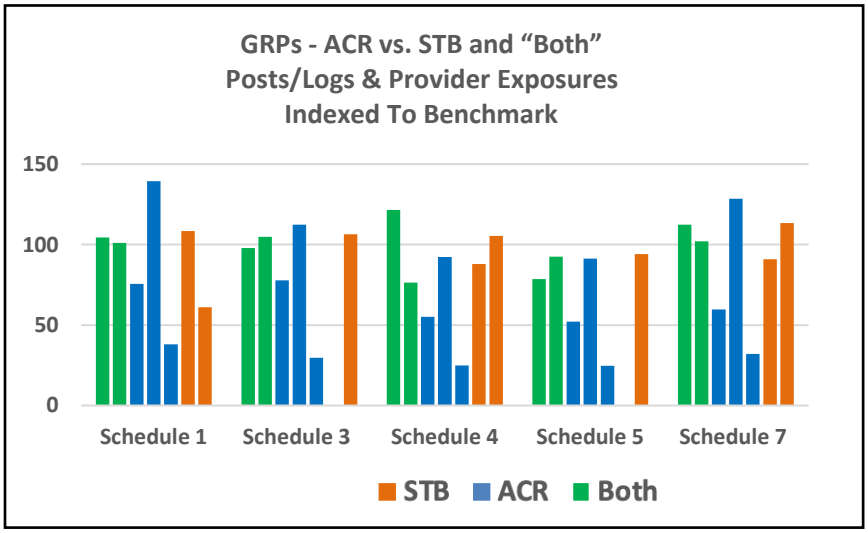
Across the average of all schedules, this analysis shows that providers who use both are most comparable to the Nielsen benchmark, providers who use set-top box data only are lower than Nielsen but within 10 percentage points, and providers who use ACR only are more than 20 points lower than Nielsen.



But again the data highlighted inconsistencies by schedule. Schedule 1 indexed lowest for set-top box providers only but performed best for ACR providers only; schedule 5 indexed lowest across all schedules for providers who utilize both.

Further, the side-by-side comparison of providers based on their underlying data sources shows the level of inconsistency that exists within providers that utilize the same basic technology.

Providers using both STB and ACR are fairly close for all schedules except schedule 4, where there is more than a 40 index point difference between the two providers.



The general relationships between providers using ACR only is consistent from schedule to schedule, but for schedule 1 and schedule 7 there is nearly a 100-point difference between two providers. Providers using STB only are relatively close for three of the schedules, but there is nearly a 50 index point difference for schedule 1.

These analyses suggest that:

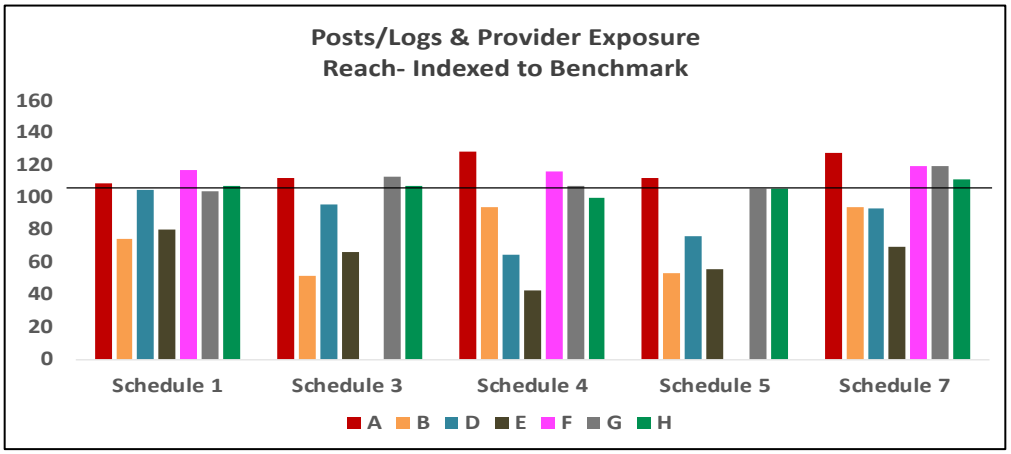
- Overall, the total GRPs reported for schedules is fairly close to the Nielsen benchmark for 4 providers, with 2 providers reporting materially lower GRPs and 1 provider reporting materially higher GRPs.

- The relationship of total GRPs reported for a schedule differ greatly provider to provider, with no consistent pattern.

Consistency of Individual Schedule Reach Across Providers

As expected, there is slightly less variance with schedule Reach than with GRPs. Having an accurate measure of campaign Reach is crucial to accurate ROI or ROAS measurement because most providers will calculate ROI/ROAS by multiplying the incremental impact among households/people exposed by the actual number of households/people exposed.

The variance across providers and within provider is clear when we evaluate the Reach index of each individual schedule relative to the benchmark.



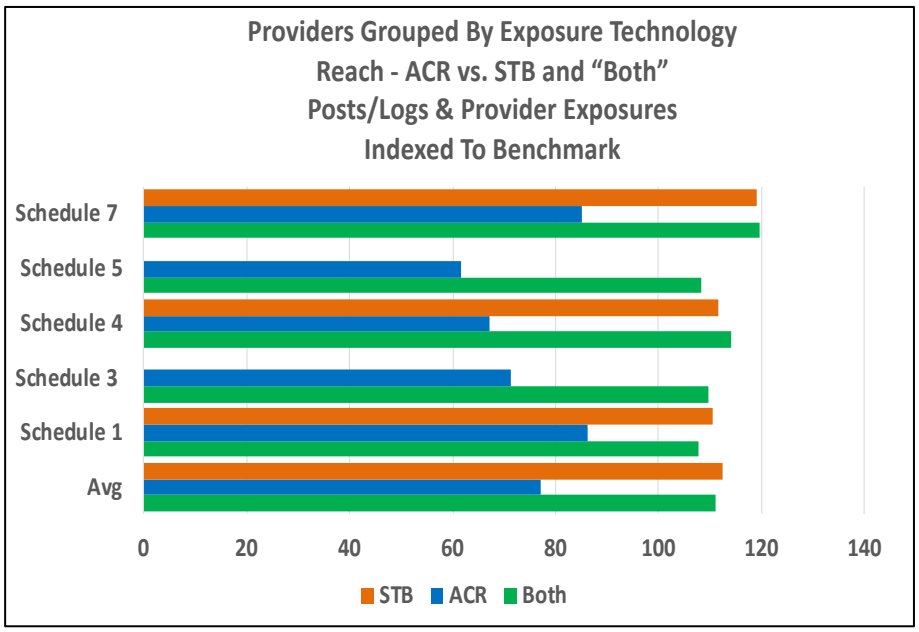
Again, we see that each provider has a very unique pattern:

- Provider A: Close to benchmark for schedules 1, 3 and 5; above for schedule 4
- Provider B: Below the benchmark for schedules 1, 3 and 5; close for schedules 4 and 7
- Provider D: Close to benchmark for schedules 1, 3 and 7; below for schedules 4 and 5
- Provider E: Below the benchmark for all schedules, schedule 4 indexed at 65
- Provider F: Close to the benchmark for all schedules
- Provider G: Close to the benchmark for all schedules
- Provider H: Close to the benchmark for all schedules

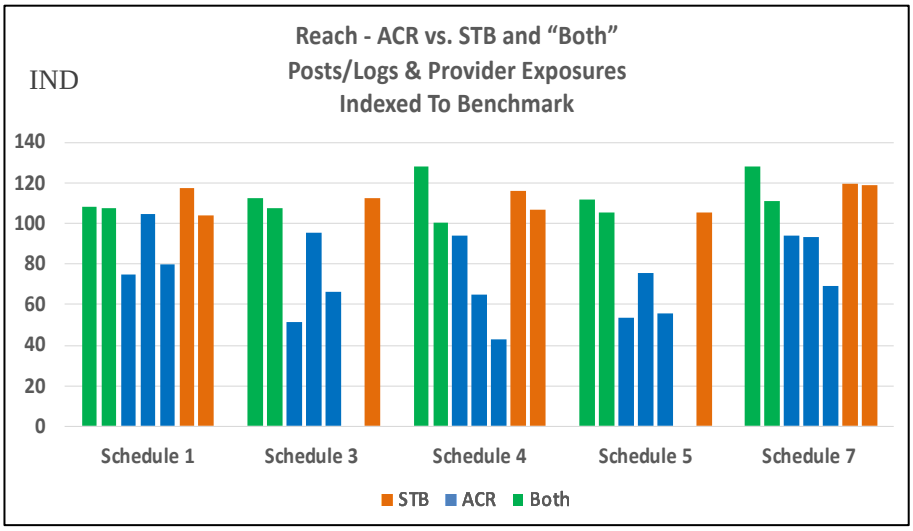
Do Underlying Sources of Viewing Data Explain Differences in Schedule Reach Levels?

Across the average of all schedules, this analysis showed that providers who use both and providers who use set-top box only are slightly above the Nielsen benchmark, and providers who use ACR only are more than 20 points lower than Nielsen.

But again the data highlighted inconsistencies by schedule. Schedule 7 indexed at about 120 for ACR/STB providers and STB only providers; schedule 5 indexed at 62 and schedule 4 indexed at 67 for ACR-only providers.



The side-by-side comparisons of providers based on their underlying data shows the level of inconsistency that exists within provider types. Providers using STB and ACR are fairly close for all schedules except schedule 4, where there is nearly a 30 index point difference between the two providers. The general relationships between providers using ACR only is less consistent across schedules than it was for GRPs.



Schedules 1, 3 and 5 have fairly similar relationships. The first ACR-only provider has the highest reach for schedule 4 and schedule 7. Providers using STB only are relatively close for three of the schedules.

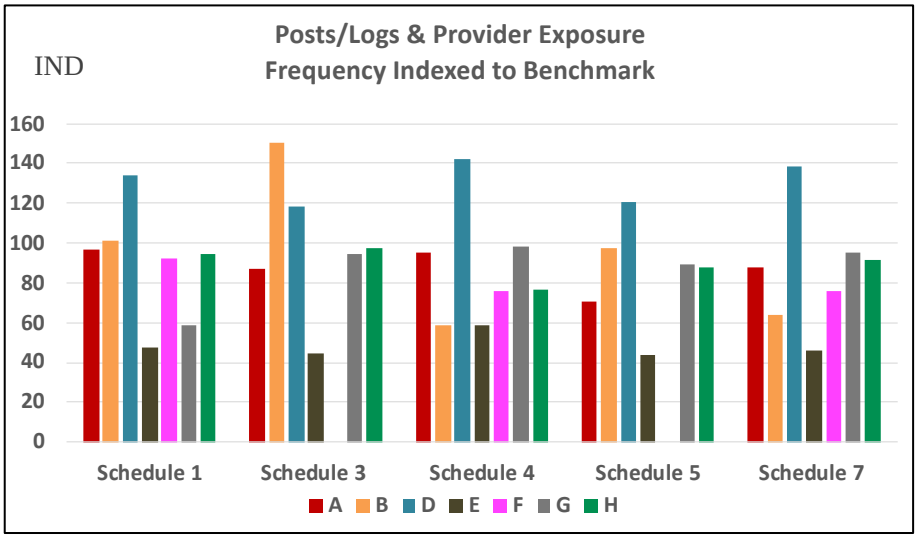
These analyses suggest that:

- As expected, the differences across providers for Total Schedule Reach is less than the differences we saw with GRPs. The differences range from one provider schedule that was 30% above the Nielsen benchmark to one provider schedule that was nearly 60% below Nielsen.
- The relationship of Total Schedule Reach differs greatly provider to provider, with no consistent pattern.
- While the source of the underlying data set does provide some clarity, wherein providers who integrate STB and ACR and providers using STB only are fairly close, providers using ACR only are furthest from the Nielsen benchmark.
- Within groups of providers using the same technology, there are material differences from schedule to schedule.

Analysis of Individual Schedule Frequency, Indexed to Nielsen Benchmark

Accurate measurement of frequency is also very important for accurate ROI measurement. Most ROI/ROAS studies have shown that there exists a natural frequency/impact curve, where impact grows with incremental frequency but there is a point where incremental frequency drives less and less incremental impact. A data set that overreports Frequency may understate incremental impact per household or persons exposed; a data set that underreports Frequency may overstate incremental impact per household or persons exposed.

The variance, across providers and within provider, is clear when we evaluate the index of each individual schedule relative to the benchmark. Each provider has a very unique pattern, as follows:

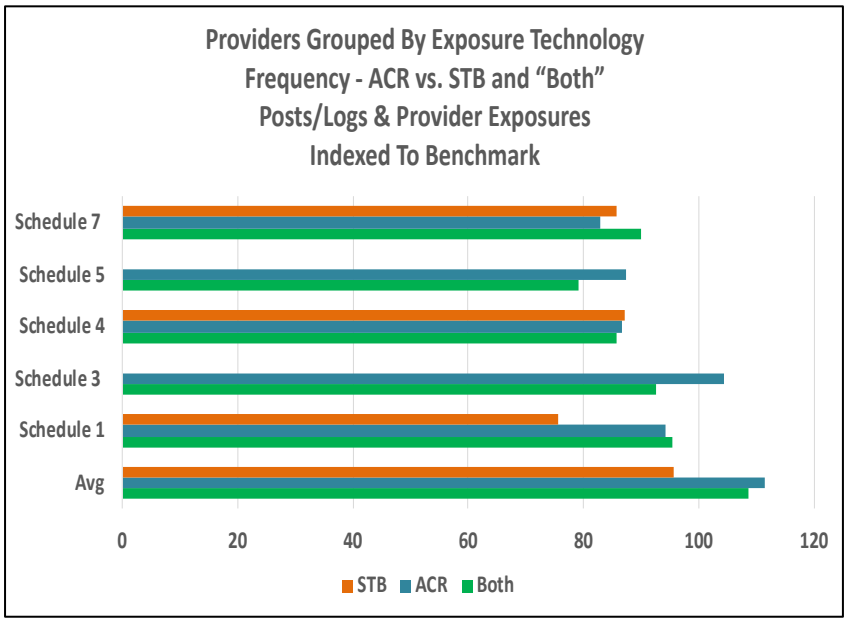


- Provider A: Close to the benchmark for schedules 1 and 4, below for schedules 3, 5, and 7
- Provider B: Close to the benchmark for schedules 1 and 5, above for schedule 3, below the benchmark for schedules 4 and 7
- Provider D: Above the norm for all schedules, index of 150 for schedule 3
- Provider E: Below the norm for all schedules, index of 44 for schedule 3 and 45 for schedule 5
- Provider F: Close to the benchmark for schedules 1, 3 and 5, below the norm for schedules 4 and 7
- Provider G: Within tolerance for schedules 3, 4, and 7, below the norm for schedules 1 and 5
- Provider H: Within tolerance for schedules 1, 3, and 5, below the norm for schedules 4 and 7

Do Underlying Sources of Viewing Data Explain Differences in Schedule Frequency Levels?

In an attempt to evaluate whether the source of each provider's viewing data is a driver of the differences we saw with schedule GRP levels, we looked at schedule Frequency levels and aggregated providers based on their underlying data as follows:

- Providers who utilize Smart TV ACR data only
- Providers who utilize set-top box data only
- Providers who utilize both



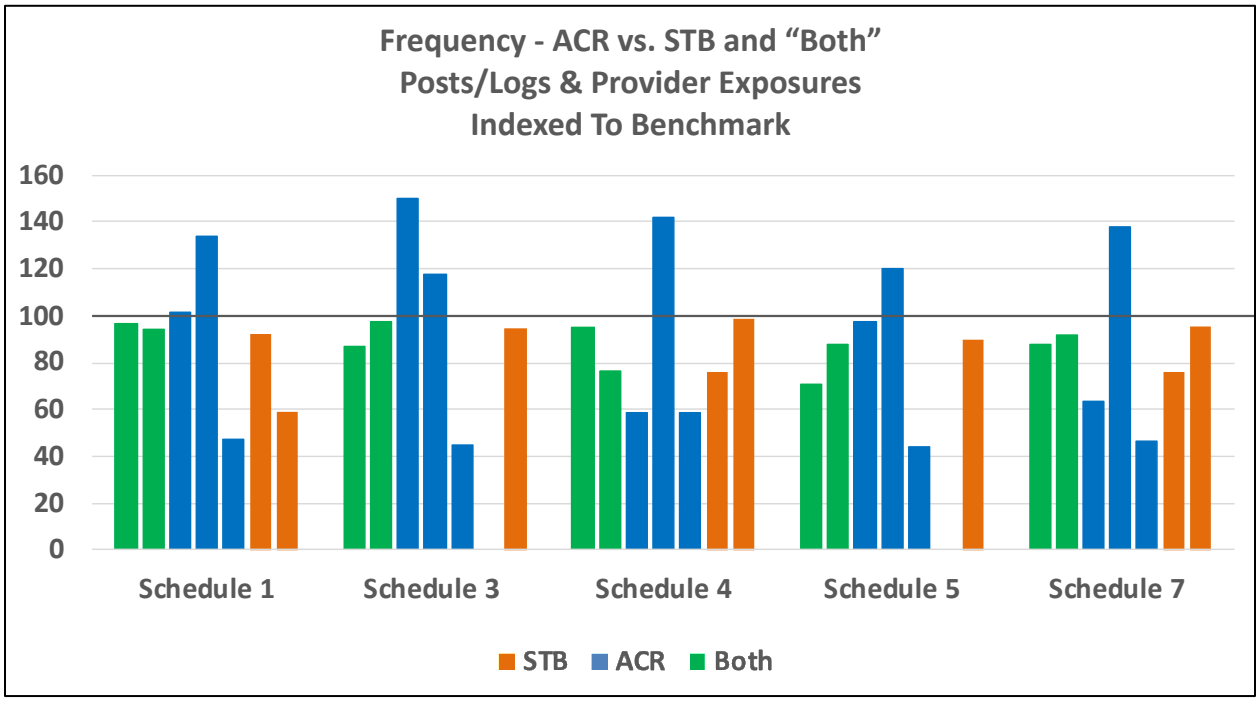
Across the average of all schedules, this analysis showed that providers who use both are most comparable to the Nielsen benchmark, closely followed by providers who use ACR only. The providers who use set-top box data are the lowest of the three provider types.

Across all providers, there is a general tendency to underreport frequency.

But again the data highlighted inconsistencies by schedule. Schedules 4, 5 and 7 reported Frequency levels lower than the Nielsen benchmarks for all provider types. Schedules 1 and 3 reported Frequency levels in line with the Nielsen benchmarks for providers using ACR only and ACR and set-top box, but schedule 1 reported Frequency below the Nielsen benchmark for STB-only providers.

Again, the side-by-side comparisons of providers based on their underlying data shows the level of inconsistency that exists within provider types and the other media metrics. Providers using STB and ACR are fairly close for all schedules except schedule 4 and schedule 5, where there is over a 20 index point difference between the two providers.

The general relationships between providers using ACR varies dramatically between schedules. For schedule 3, there is nearly a 100 index point difference between the ACR-only provider with the highest Frequency compared to the ACR-only provider with the lowest Frequency. Providers using STB only showed differences as large as 33 index points for schedule 1, and the relationship between the two providers flips between schedules 1 and 4.



What Do We Take Away From This?

- As expected, the differences across providers for Total Schedule Frequency is greater than Reach, but less than the differences for GRPs. The differences range from one provider schedule that was 50% above the Nielsen benchmark to one provider schedule that was nearly 40% below Nielsen.
- The relationship of Total Schedule Frequency differs greatly provider to provider, with no consistent pattern.
- While the source of the underlying data does provide some clarity—providers who integrate STB and ACR and providers using STB-only are fairly close—providers using ACR only are slightly lower.
- Within groups of providers using the same technology, there are material differences schedule to schedule.

Summarizing GRPs, Reach, Frequency Delivery

In this section of the report, we highlighted the lack of consistency in exposure measurement across providers and across the various schedules. The table below summarizes the situation. Each provider has a unique signature, with different relationships compared to the Nielsen benchmark.

Posts/Logs & Provider Exposure Average Schedule GRPs, Reach, Frequency- Indexed to Benchmark							
	A	B	D	E	F	G	H
GRPs	**101**	**72**	**122**	**32**	**91**	**90**	**101**
Reach	**115**	**70**	**93**	**68**	**115**	**110**	**99**
Freq.	**105**	**121**	**157**	**57**	**93**	**98**	**122**

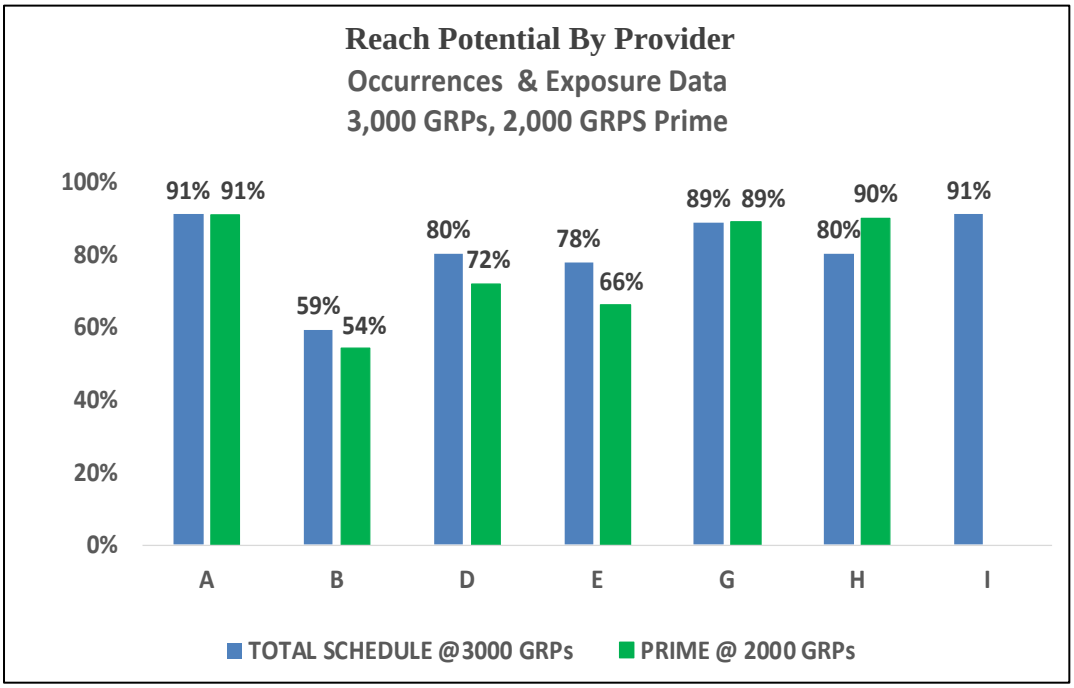
Analysis of Reach Potential by Provider

The analysis of exposure has clearly shown that provider measures of GRPs is a major driver of Reach differences. We also thought it important to evaluate—assuming equal GRPs across providers—whether each provider would report the same level of estimated reach.

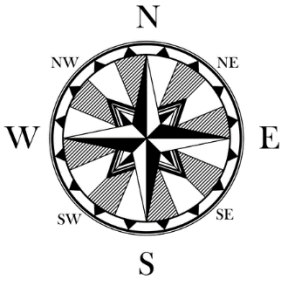
Using all of the schedules that were rated, a logarithmic function was created to develop an estimate of Reach at different GRP levels for different dayparts.

Based on industry experience, our expectation is that a schedule of 3,000 GRPs across broadcast and cable and all dayparts would yield a reach of around 90% of the population.

The analysis shows that Providers A, G and I reach that 9% benchmark, Providers D, E and F are within 11% of the 90% reach estimate, but Provider B is far from that benchmark.



It may be a combination of issues with the underlying data sets (e.g., lack of primary TV set coverage), weighting methodology, and lack of data cleansing rules such as unification that could cause a provider to understate reach maximums.



Final Summary

- **Occurrences and exposure data are highly inconsistent across providers.** The accuracy of spot detection and all of the different exposure data elements—GRPs, Reach, Frequency—differ from provider to provider. This is not a good thing. Identifying and the counting of exposures, or impressions, should be standard starting points. How providers then connect the dots from exposures to outcomes should be their points of difference. If they are all identifying different occurrence levels and evaluating different exposures, then they may as well be evaluating different campaigns.

- **Lift outcomes differ significantly.** Because of the differences in both occurrence and exposure data, measurement of lift provider to provider yielded different results, both in terms of magnitude and direction of lift.

- **Provider exposure data impacts lift results more than occurrence data.** The study found that while there are differences between providers for both occurrence data and exposure data, the differences in exposure data are a much larger contributor to differences in lift measurement.

- **Methodology, rather than underlying technology, drives results.** For both occurrence data and exposure data, the underlying data elements—monitoring of network signals to create an accurate ad occurrence file, ACR and/or set-top box data that measures what households are viewing—are similar. The methodology of converting that data into final ad occurrence files and exposure data, including weighting, editing and other data processing rules, is believed to be the cause of the differences between providers.

For More Information

Please reach out to Janus Strategy & Insights or Sequent Partners.

Howard Shimmel

Janus Strategy & Insights

Howard@JanusStrategyandInsights.com

Jim Spaeth

Sequent Partners

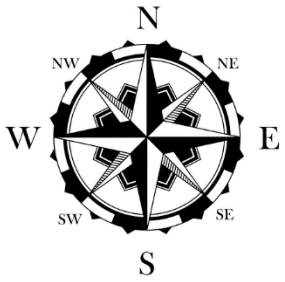
jim@sequentpartners.com

Alice K. Sylvester

Sequent Partners

alice@sequentpartners.com

Provider Profiles



From the beginning of the project, we agreed not to release the findings of this study with provider company names identified. Though that has frustrated some people, CIMM and the providers agreed that releasing unblinded findings violates the spirit of learning in a collaborative, low-risk environment we had promised. CIMM is very grateful for the openness and support of television attribution providers. The project was labor intensive and involved a large number of complicated custom data pulls. Their cooperation was inspiring and their desire to learn about their own data and emerging best practices is laudable.

CIMM offered the providers this opportunity to share, in their own words, their approaches to occurrences, exposure data and attribution in general.

605 offers an independent, deterministic TV viewership data source to measure TV audiences among US households with one or more televisions. Viewing activity is collected from more than 21 million homes with STB-connected or smart TVs. Viewership data from these households is transformed, weighted, and overlaid with same-home demographics and consumer information, yielding more than 10 million households, to support next generation TV audience analytics.

The processing workflow starts with the collection of second-by-second live and DVR/VOD time-shifted viewing data, station schedules, and advertising as-run logs and continues through an extensive process of data extraction, transformation, and loading into 605's secure data lake. Within the data lake, 605 generates household weights projecting to the universe of US TV households, then overlays station, programming, and advertisement data layers.

To qualify for reporting, a household's demographic characteristics are obtained from Experian through a matching process that is blind to 605 to protect personally identifiable information. Additionally, a home must have watched one or more seconds of television during the past 90 days to qualify.

Household weights balance, adjust, and project household television viewing to correct for biases in the raw data sample and ensure that published metrics mirror the overall population.

605 uses raking weighting (also known as iterative proportional fitting or rim weighting) to calculate a weight for every qualified home. The raking method of generating weights iteratively adjusts weights until the marginal distributions of the weighted measured population match the total target population.

Alphonso is a TV data and measurement company, with an audience footprint that provides brands and agencies with near real-time TV ad campaign measurement, closed-loop attribution for TV ads, and TV audience extension across digital devices. Alphonso TV Data Cloud services are used by hundreds of brands and agencies in the US.

Alphonso's audience footprint in the US includes about 15 million opted-in households that report viewership data to Alphonso via smart TVs and connected devices of major brands such as Sharp, Toshiba, Hisense, LG, TiVo, Seiki and Skyworth. Alphonso's balanced panel, which is representative of the geographic and demographic distribution of the US households, uses this deterministic TV viewership data to help advertisers and agencies pinpoint and measure their TV audiences and understand the journey from ad exposures to business results.

Alphonso uses its patented video AI technology to automatically detect ads running on linear TV in both national and local DMAs. This allows Alphonso to monitor all ads running on TV without requiring any creatives from brands or agencies and enables Alphonso to report on all advertising activity in near real-time and at scale.

Its SaaS offering, Alphonso Insights, delivers actionable TV measurement and closed-loop attribution with offline data in real time, to help brands understand the true impact of TV advertising. Alphonso offers various types of attribution such as tune-in attribution, website or store visit attribution and purchase data attribution for both linear TV and OTT campaigns.

In terms of validation, we have worked with a variety of different partners and clients to establish best practices. For instance, we work with numerous TV stations directly and verify our data with the station's data to make sure there is a high degree of accuracy. We are the attribution provider for all of CBS O&O stations, Tegna, Sinclair, Hertz and many others. We work with companies such as Experian and Nielsen to match our panel to their demographic data to look at national representativeness of the data.

We use iSpot as our source for ad occurrence tracking for national campaigns. For local cable executions, we use our own as-run affidavits as the source of truth.

To generate ad exposure, we sync these timestamped ad logs (this particular ad aired at this time on this network in this geography) with household-level STB viewership to create ad exposure.

Television viewing behavior is sourced from TV Essentials, which combines second-by-second set-top-box tuning data from AT&T U-verse, Charter Spectrum (includes legacy Time Warner Cable households), Cox, DirecTV and Dish Network.

Tune data from each source is separately ingested, cleansed, schedulized, quality checked and conformed to a uniform schema by Comscore's Data Operations team. Then the disparate data sources are combined into a single national database from which Comscore builds projections created and maintained by the Statistical Operations team.

Ad schedules are licensed by Comscore from Kantar Media Intelligence and overlaid on the tuning data at a household level by aligning the timestamps of the ad start and end times with the tuning start and end times on each network.

iSpot monitors ads on more than 130 channels across the Top 30 DMAs in the U.S. As new ads run, we immediately ingest them into our system and add their fingerprints to our catalog of ads that now includes more than 1 million different creatives. Our proprietary ad catalog is then used alongside the Vizio ACR technology to detect the individual ads as they appear on any network in any DMA on the screen of the 15 million Smart Vizio TVs opted into the ACR panel. We receive this data, combine it with our Gracenote schedule data, and now have occurrences and their associated panel impressions. Our Vizio panel is well dispersed geographically and has been weighted at the household level to be representative of household viewership and demographics when we scale it up to the full US population. The fully balanced and representative panel powers our reach, frequency, impressions and GRPs.

NCS believes that "smart data" is the combination of Big Data informed by small, representative, currency quality panel data. In practice, this means that the Nielsen NPM currency viewing data is used to inform the on/off and persons-based viewing models for cable set-top box (STB) data using machine learning techniques. Nielsen does this cleaning for NCS. Weights and population projections are based on comparisons between our currency and Big Data depending on the data elements involved in a study.

The size of the panel we used for this analysis was 9,321,712 households. A unification of 75% was applied to this panel. The occurrence data is from Nielsen's Ad Intel service.

For this project, Samba TV created a footprint of households with ACR-enabled Smart TVs that are balanced on demographics by geo-region to project within 99.99% to the US Census to ensure a representative sample. The represented sample accurately reports both traditional heavy linear viewership households and those with newer trends that consume TV content from nontraditional methods such as streaming or with heavier time-shifted consumption. Including households with none or very little viewership to linearly broadcasted content ensures all audiences are represented in the analysis.

The methodology used to assign exposure of a TV impression is based on content viewership to the exact timestamp of the ad exposure logs provided by CIMM. If a household enabled Samba Smart TV was viewing the content/network at the exact time of an ad occurrence, then the household would be labeled as an exposed impression. Using this methodology ensures Samba TV is assigning exposure accurately to households and significantly removing the chance to assign false-positive exposures.

TVSquared's platform can work with viewing data from various sources. For the purposes of this study, TiVo's panel of 2–3M households was used. More commonly, TVSquared uses Inscape's panel with its larger footprint. Data is cleaned and filtered to include only households where both exposures and responses over the whole campaign period can be tracked. Viewing data is calibrated to the national TV viewing population. TVSquared's Inscape data integration allows for the use of household-level demographics as well as DMA to weight the data.

VideoAmp is an interoperable measurement and optimization platform. Advertisers, agencies and media owners leverage their privacy-first suite of data and software solutions to gain a true deduplicated read of performance across linear TV, OTT, digital and walled garden media by connecting the dots between ad exposures, audiences and outcomes.

VideoAmp's proprietary commingled TV viewership data set consists of set-top box data from MVPDs and Smart TV ACR data from TV manufacturers, deduplicated to form the TV viewership data sets. VideoAmp utilizes unique characteristics from each data source to correct the other, creating a more comprehensive, unified data set.

By leveraging household identity matching and consumer profiling partnerships, VideoAmp connects TV viewership to digital activities for advanced targeting and measurement. If consumers of a particular segment are targeted by an ad and convert at a high-rate, then it is important to understand whether this high conversion rate was caused by ad exposure or belonging to that particular segment. Thus, to quantify the effect of TV ad exposure, VideoAmp can find a similar user with the same ad exposure probability who is not exposed to that particular ad. By doing so, VideoAmp's model measures the incrementality of TV ads by controlling for the propensity of exposure, creating a matched control group for effective and reliable measurement.