



NIELSEN JOURNAL
of MEASUREMENT

OCTOBER 2016-VOL 1 ISSUE 2

USING SINGLE-SOURCE DATA TO MEASURE ADVERTISING EFFECTIVENESS

By Paulina Berkovich, Sr. Analyst

Leslie Wood, Chief Research Officer, Nielsen Catalina Solutions



EDITOR-IN-CHIEF
SAUL ROSENBERG

MANAGING EDITOR
JEROME SAMSON

REVIEW BOARD

PAUL DONATO
EVP, Chief Research Officer
Watch R&D

MAINAK MAZUMDAR
EVP, Chief Research Officer
Watch Data Science

FRANK PIOTROWSKI
EVP, Chief Research Officer
Buy Data Science

ARUN RAMASWAMY
Chief Engineer

ERIC SOLOMON
SVP, Product Leadership

The world of measurement is changing.

Thanks to recent advances in data collection, transfer, storage and analysis, there's never been more data available to research organizations. But 'Big Data' does not guarantee good data, and robust research methodologies are more important than ever.

Measurement Science is at the heart of what we do. Behind every piece of data at Nielsen, behind every insight, there's a world of scientific methods and techniques in constant development. And we're constantly cooperating on ground-breaking initiatives with other scientists and thought-leaders in the industry. All of this work happens under the hood, but it's not any less important. In fact, it's absolutely fundamental in ensuring that the data our clients receive from us is of the utmost quality.

These developments are very exciting to us, and we created the Nielsen Journal of Measurement to share them with you.

WELCOME TO THE NIELSEN JOURNAL OF MEASUREMENT

SAUL ROSENBERG

The Nielsen Journal of Measurement will explore the following topic areas in 2016:



BIG DATA - Articles in this topic area will explore ways in which Big Data may be used to improve research methods and further our understanding of consumer behavior.



SURVEYS - Surveys are everywhere these days, but unfortunately science is often an afterthought. Articles in this area highlight how survey research continues to evolve to answer today's demands.



NEUROSCIENCE - We now have reliable tools to monitor a consumer's neurological and emotional response to a marketing stimulus. Articles in this area keep you abreast of new developments in this rapidly evolving field.



ANALYTICS - Analytics are part of every business decision today, and data science is a rich field of exploration and development. Articles in this area showcase new data analysis techniques for measurement.



PANELS - Panels are the backbone of syndicated measurement solutions around the world today. Articles in this area pertain to all aspects of panel design, management and performance monitoring.

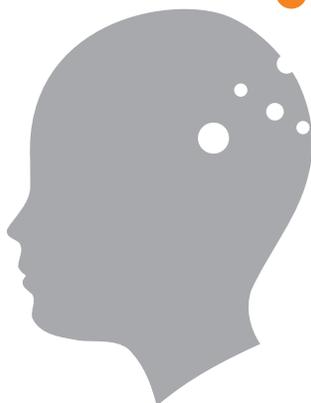
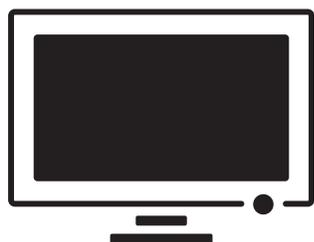


TECHNOLOGY - New technology is created every day, and some of it is so groundbreaking that it can fundamentally transform our behavior. Articles in this area explore the measurement implications of those new technologies.

USING SINGLE-SOURCE DATA TO MEASURE ADVERTISING EFFECTIVENESS

BY PAULINA BERKOVICH *Sr. Analyst*

LESLIE WOOD *Chief Research Officer, Nielsen Catalina Solutions*



INTRODUCTION

How do we know if advertising works? Is there a way to directly measure the in-store sales results of an advertising campaign? Scholars and marketers have grappled with measuring advertising effectiveness for decades. The long-standing joke is that, as John Wanamaker is said to have remarked more than a century ago, “half the money we spend on advertising is wasted—we just don’t know which half.” Thankfully, we’re long past that stage in our understanding of how advertising works, but there’s no question that measuring advertising effectiveness and optimizing performance remain a challenge for advertisers and media companies alike.

In recent years, massive shopper databases have allowed the industry to develop exciting new research methods. In particular, researchers at Nielsen Catalina Solutions (NCS) have introduced a breakthrough methodology based on single-source datasets to measure the direct effect that an ad campaign has on sales. These ‘sales effect studies’ offer a powerful solution to measure changes in purchase behavior among households that saw an ad campaign, and compare it to the behavior of similar households that were not exposed to it. This paper dives into the details of how those studies are put together.

THE BENEFITS OF SALES EFFECT STUDIES

Whereas traditional ad effectiveness studies relied on separate panels and datasets to collect ad exposure data and sales data, sales effect studies are based on a single-source dataset—that is, ad exposure and sales data are captured for the same households. Before single-source, researchers could only attempt to draw conclusions through correlations between disparate data sources, and changes in spending for a product could only be evaluated at an aggregate market level. Such an analysis was often complicated by external factors, and it was difficult to identify what particular creative execution or placement was driving sales—if at all. The advent of single-source data has made it possible to control for external factors, and attribute changes in purchase behavior much more accurately to advertising seen by the household.

These new studies provide value to a broad spectrum of companies. For advertisers, they are a critical piece of a holistic approach to planning, activating and measuring campaigns. With the insights gained from evaluating a campaign, they can allocate budget and impressions to the most effective media and improve returns on future advertising spend. For media companies, on the other hand, sales effect studies demonstrate accountability to advertisers and help make a case for increased spend in the future.

Advertising campaigns are run with a number of objectives in mind. At the most basic level, they aim to increase sales of a brand—both in the short term and the long term. However, more detailed objectives vary widely. Ad campaigns may seek to convince buyers to try out a new product, or bring lapsed buyers back to an established brand. They could be trying to shift share from competing brands, or increase spend among existing brand buyers. Single-source data makes it possible to quantify changes in purchase behavior in enough detail not only to calculate incremental spend, but to answer many other questions as well: Did a large number of new households try the product? Which competitors was share taken from? Which customer segments are most responsive to advertising, and are my existing customers buying my product more regularly? All of these insights can be used to uncover the brand's real strengths and plan more effective campaigns in the future.

NCS uses sales effect studies to evaluate campaigns on a number of channels, including digital, mobile, TV, radio,

magazines and direct marketing using a client's proprietary data. Our single-source dataset contains over two years of purchase data, spanning more than 44,000 brands across hundreds of categories. The purchase data of the single-source dataset comes from the store loyalty cards of 90 million unduplicated households, a subset of Catalina's data warehouse. Catalina's retailers include major grocery stores and drugstore chains, and the data is commingled for a holistic view of spending for each household. Product categories represented in the data include food, beverage, baby products, pet food, and general merchandise, just to name a few. To date, studies have been delivered for thousands of campaigns, resulting in a series of advertising benchmarks that allow advertisers to compare campaign performance to industry standards.

Let's examine how those studies are conducted.

A SALES EFFECT STUDY TYPICALLY INVOLVES THE FOLLOWING STEPS:

- TAGGING THE CAMPAIGN AND SELECTING HOUSEHOLDS
- APPLYING ALL-OUTLET ADJUSTMENTS TO THE PURCHASE DATA
- MATCHING TEST AND CONTROL HOUSEHOLDS
- MEASURING HOUSEHOLD-LEVEL INCREMENTAL SALES
- IDENTIFYING THE DRIVERS OF INCREMENTAL SALES
- CALCULATING REACH AND RETURN

TAGGING THE CAMPAIGN AND SELECTING HOUSEHOLDS

When a sales effect measurement is ordered, tags are embedded in all creative units that are part of the campaign, in order to identify not only which households are exposed, but also which specific version of the creative execution

they were exposed to. This is considered the “media exposure” element of the single-source dataset. The media types supported for tagging include addressable TV, digital, and mobile. Radio campaigns are tracked with codes, and magazine campaign exposures can be determined based on the households that subscribe to the corresponding publisher as well as secondary (or pass-along) readership. Linear TV data is collected from Nielsen panels as well as set-top boxes.

Online publisher partners, such as Yahoo or AOL, use first-party cookies to link exposures to a unique household. This is called a “direct match,” and tagging is not required in those situations because the publisher already has all the necessary linking information. With third-party cookies, on the other hand, a third-party company is required to link exposure to households and overlay demographic information. This is a technique that’s typically used for programmatic ads, and referred to as an “indirect match” or “cookie match.” In either case, no personally identifiable information is attached to any household. To protect the privacy of all households represented in the data, each household is simply represented by an anonymized ID number¹.

To be included in a sales effect study, households must meet a pre-period “purchase static” to ensure that they are actively

using the loyalty cards that form the basis of the purchase data used in the study (or, for Homescan households, that they are actively scanning their purchases). An additional static is applied for the post-period. The second static maintains, as closely as possible, the same ratio as the pre-period static. For example, if a study has a pre-period of 4 quarters and a post-period of 2 quarters, and the pre-period static requires that a household purchase a particular product in two of those four quarters in order to qualify (e.g., a ratio of 1/2), then that household would need to register a purchase in at least one of the two post-period quarters (e.g., the same ratio of 1/2) in order to be counted.

The purpose of the statics is to retain the greatest possible percentage of dollars spent while excluding households where large amounts of data are missing. They ensure that households with no spend on a particular product or category truly didn’t purchase it (in statistics, those are referred to as “true negatives”). Statics are necessary to balance the need to protect data quality with the need to retain enough households to offer a robust basis for analysis. They’re critically important and an ongoing topic of investigation for the research community. We’re currently using different statics for different product departments (or product categories), as shown in the following table:

STATIC	DEPARTMENTS INCLUDED	MINIMUM PURCHASE THRESHOLD	MINIMUM PURCHASE FREQUENCY
FOOD	BAKERY, DAIRY, DELI, FROZEN FOODS, GROCERY, MEAT, PRODUCE	\$30	3/4 QUARTERS
HEALTH CARE	HEALTH CARE	\$5	2/4 QUARTERS
PERSONAL & BEAUTY	BEAUTY CARE, PERSONAL CARE	\$5	2/4 QUARTERS
NON-FOOD	GENERAL MERCHANDISE, HOUSEHOLD CARE, PET CARE	\$10	3/4 QUARTERS

¹NCS is committed to protecting the privacy of all households represented in its data. A dedicated privacy council is constantly reviewing and revising NCS protocols to ensure that they meet industry standards. As the possibilities of single-source data are developed further, maintaining security will remain a central concern

Sample sizes can vary widely, depending on the length of the campaign and the size of the product category. For a Homescan study, a typical sample size is a few thousand households. Digital studies typically produce the largest test groups. In one extreme case, a multi-month campaign for the orange juice category had a sample size of approximately two million households. The only sample size constraint that must be met is a minimum buyer count for the item or items being studied. Without a robust buyer count, the difference in purchasing between test and control groups would not be statistically significant. For frequent shopper studies, the buyer count is generally larger, but here too it depends on the product category. The instant breakfast study had a bit more than 4,000 buyers. In contrast, the household item study had over 100,000 buyers.

How do we identify the products that are purchased by the households in the study? All items and categories in sales effect studies are defined using classifications already existing in NCS data, or using custom UPC lists provided by clients. Precise product definitions allow a client to assess a campaign's impact on the purchase of a particular subset of the brand, and category definitions make it possible to calculate how much share the brand has captured from competitors within a particular grouping of products. UPC lists that are both comprehensive and up-to-date are critical to get accurate results.

APPLYING ALL-OUTLET ADJUSTMENTS TO THE PURCHASE DATA

Catalina's retail partners include major drugstore chains and grocery stores. These retailers add up to a very large frequent shopper dataset (FSD), but it's not an exhaustive list of all outlets where a consumer might purchase goods and services. Thankfully, both the Nielsen Homescan panel and Nielsen's Retail Measurement Services (RMS)² can help. Intentionally selected to be nationally representative in terms of key demographics, the Homescan panel provides a comprehensive look into household purchasing. It can be used as the sole source of purchase data for certain sales effect studies, but we're also using it to calibrate purchase data from the FSD and estimate where else consumers might have purchased the products in the study. This process, called all-outlet adjustment (AOA), leverages the overlap between

Homescan and Catalina's FSD. It works by estimating the proportion of Homescan buyers who shop at FSD retailers, calculating how much they spend inside and outside FSD retailers, and scaling dollar differences accordingly. We use the RMS data in a similar way to calibrate sales projections. Via the AOA process, we can account for sales beyond FSD outlets, and correctly identify households that did or did not make any purchase anywhere within the product category during the time period covered by the study.

A recent study for an instant breakfast brand was run using FSD data and showed that households in the test group spent an average of \$2.04 on the product in the post-period. However, with the all-outlet adjustment, that average dropped to \$0.40. This reflects the fact that the penetration of the instant breakfast category is limited. Few households purchase products like powdered mixes or bottled ready-to-drink shakes. Additionally, households that purchase those products do so primarily at grocery stores, which are well-represented in frequent shopper data. As a result, we need limited adjustment to account for purchases of the brand at retailers not captured by FSD loyalty cards.

In contrast, another recent study looked at a common household product. From the raw FSD data, the test group spent an average of \$3.23 on the brand being studied. But after the all-outlet adjustment, the average surged to \$8.20. This increase reflects both the fact that the category is purchased by a large proportion of households (meaning that no downward adjustment is needed), and the fact that the product is commonly purchased at retailers other than the ones in the frequent shopper data. In both cases, using the adjusted value to calculate the incremental spending driven by the campaign results in a more accurate assessment of its true impact.

The all-outlet adjustment has also a substantial effect on product penetration. For the household product, frequent shopper data alone indicated that it had a penetration of approximately 23% among the test group. However, after the all-outlet adjustment, the penetration jumped to nearly 43%. This adjustment captures the large number of consumers who purchase the product exclusively at non-FSD retailers. On the other hand, the breakfast brand had a penetration of about 17% within households included in the study (which were required to purchase in the instant breakfast category in order to qualify), but an all-outlet penetration of just over 2%. This decrease reflects the low penetration of the category as a whole.

²The Nielsen RMS data captures sales and price data from virtually every major retail chain at the point of sale and at the UPC level

MATCHING TEST AND CONTROL HOUSEHOLDS

Once the list of exposed households that will form the test group is created, each of the households on that list is matched with a household in the NCS data pool that was not exposed to the ad campaign (a “non-exposed household”) and most closely resembles it. The purpose of matching is to be able to attribute any increase in spending to advertising rather than to an existing preference for the product being studied. To that end, the variables used to determine test-control household pairs are customized for each project to pertain to the particular product category.

MATCHED HOUSEHOLD PAIR

Matched on similar

- 1 Brand and Category Penetration
- 2 Total Pre Period Brand Dollar/Volume/Units
- 3 Total Pre Period Category Dollar/Volume/Units
- 4 Trips by Retailer
- 5 Household Demographics
- 6 Geographic Location
- 7 Lapsed Time since brand purchase



We use hundreds of matching variables. Those include: past spending on the advertised item being studied, competitor brands, and category as a whole; purchase amount at different retailers; recency of purchase; dollar purchases with a discount; and demographic information. Match variables include not only aggregated variables that capture total spend for the pre-period and post-period, but also monthly spend variables to ensure that households are at a similar stage in their buying cycle. For example, if a household consistently purchased a product for the first six months of the pre-period but did not make any purchases in the final six months, it would not necessarily match well with a household that did make recent purchases, even if the total dollar amount they spent over the course of the year was the same. Controlling for the stage in the purchase cycle helps separate out the

impact of advertising from purchasing by a household that had simply run out of the product and would have purchased it at that point in time anyway.

Purchase-based variables are weighed more heavily than demographic variables. Specifically, the variables that receive the greatest weight in the match relate to buying patterns during the year prior to the ad campaign. These include dollars spent on the item, the number of units purchased, the number of purchase occasions, and average dollars spent per trip. These variables help establish that the test-control pairs resemble one another in their buying patterns of the brand being studied. A household that bought a large quantity of a product on one occasion might have the same overall spending as another that bought smaller quantities on multiple separate shopping trips, but matching on occasions and dollars per trip helps ensure that the paired households exhibit a similar consumption behavior. That way, an extra purchase occasion in response to the ad campaign can be separated out from a regular frequent purchase.

By comparison, demographic variables such as household income, presence of children, and the age of the head of household are weighed very lightly because these characteristics are not highly predictive of future purchase behavior. To the extent that they are—for example, parents of young children are much more likely to buy baby products—this information is already largely contained in the purchase data. However, information about the demographics of households that saw a campaign is provided in the form of indices that show whether a certain group is overrepresented among those exposed.

Control households may be paired with more than one test household because linear television campaigns have such high reach that there are not enough control households for a one-to-one match. However, research has shown that as long as resampling is limited to prevent any single control household from having too much influence, it does not adversely affect the quality of the results.

To guarantee that the incremental sales impact seen in the post-period is truly attributable to advertising exposure, the total difference in certain key variables between the test and control groups is less than 1% in the year prior to the start of the campaign. Key variables include total purchase, price, and purchase on deal. That 1% difference does not necessarily apply to individual household pairs. For example,

in the instant breakfast brand study, test households spent an average of \$18.99 within the category during the pre-period of the study. Control households spent an average of \$19.03. But among the individual test-control pairs, there were cases where the test household spent \$100 more than the control household, and also cases where it spent \$100 less. These pairs made it into the study because they resembled each other closely on other variables.

MEASURING HOUSEHOLD-LEVEL INCREMENTAL SALES

Once test-control pairs are created, the final step is to compare the aggregated purchases of the two groups in the post-period, thus quantifying average incremental sales per household (see our sidebar: metrics and calculations). Decomposition of the incremental sales can also be calculated by aggregating the differences for sub-groups in order to evaluate if a particular segment of consumers responded particularly well. This information helps the advertiser decide if there is a better way to allocate impressions for future campaigns.

In the household product study, the test group spent an average of \$8.20 on the brand throughout the four-month period being studied, after all-outlet adjustment. The control group spent an average of \$7.99. The results suggest that being exposed to the campaign led to an incremental spend of \$0.21 per household, a 3% increase. Statistical significance was also calculated based on the distribution of the purchase data and the sample size. A significance level of 99% confirms that the results were conclusive and not a result of sampling variation. The robust sample size of many studies makes a significance level over 90% common even for studies where purchases increase by only one or two percent.

For this particular campaign, the client ran two different versions of the creative, and the exposure data was used to split the test group into those who saw one version or the other and an overlap group that saw both. Purchase data for each of the groups showed that one of the creative executions drove the entirety of the incremental revenue. While the results of the decompositions are not always statistically significant, they can highlight the greater success of one portion of the campaign compared to others. They

can be computed for a variety of custom segments based on demographics, ad exposure, or purchase history.

A commonly used decomposition is one where we split the exposed group into subgroups based both on their past category purchases and on their loyalty to the brand being studied. For example, households that were already heavy buyers of the category but were only moderately loyal to the brand might have responded well to the campaign and become more loyal to the brand. This type of insight can help brand managers identify more specialized consumer segments and other diagnostics to improve future campaigns.

For parent brands with multiple sub-brands, the incremental revenue can be split up by sub-brand to determine which ones saw the greatest gain. In the household brand study, one particular sub-brand of the parent brand accounted for less than 20% of its sales in the pre-period, but represented over 40% of incremental sales in the post-period. Campaigns do not impact all products within a brand family equally, making the sub-brand contribution a valuable insight when planning what products to emphasize in future creative executions.

Finally, using the purchase data, the brand being studied can be compared to competitors in terms of category share requirements. This makes it possible for the advertiser to see which brands, if any, it's taking share from. The competitor products can be defined based on a custom UPC list or based on product designations within NCS data, allowing for flexibility when determining how purchasing of a specific competitive set has changed.

A distinction must be made between the results seen in a sales effect study and overall market performance. The incremental sales are reflective of sales among the test households only, so spending among exposed households should not be expected to align with spending among all households. Single-source data makes it possible to separate out exposed households, distinguishing their purchase behavior from the purchase data of all buyers during that time span. This type of analysis isolates the impact of the ad campaign among those exposed to the advertising, and shouldn't be confused with measures of sales in the total market.

IDENTIFYING THE DRIVERS OF INCREMENTAL SALES

The behavioral changes that drive incremental revenue are not always the same. For instance, an advertiser may want to know whether buyers are new to the brand or whether they had purchased it in the past already. It's possible to answer that question using purchase data from the pre-period. Households that didn't purchase in the pre-period but did in the post-period are classified as trial buyers, while households that purchased in the pre-period and again in the post-period are classified as repeat buyers. These totals are compared for the test and control groups. In the household product study, the brand was well-established and already had a high penetration. Over 40% of households in both the test and control groups were repeat buyers. But the percentage was higher within the test group, suggesting that the campaign successfully led more buyers of the brand to stay loyal to it.

Differences among households that made a purchase are also analyzed by calculating the dollar buying rate. This number represents the average household spend during the post-period, excluding those that didn't buy the brand being studied. In the case of the household brand, the average dollar buying rate was \$0.14 higher for the test group than for the control group. This suggests that the campaign not only drove more households to make a purchase, but also drove those households to spend more. However, increased penetration was a larger contributor to its success.

The dollar buying rate can be decomposed even further, into the number of purchase occasions and the average dollar amount spent per shopping trip. The average number of purchases made by the test and control groups for the household brand was almost identical. This suggests that the campaign did not drive purchases on a greater number of separate occasions. Incremental sales came instead from consumers choosing to buy more of the brand at a time.

The purchase drivers were significantly different for the breakfast brand. The campaign drove an overall lift of 28%, a number reflective of a smaller brand with room to grow its sales. Of the total lift, 10% came from increased penetration and 10% came from a higher number of purchase occasions. This campaign ran for less than a month, but it successfully persuaded consumers to choose the brand more frequently, at the expense of its competitors.

CALCULATING REACH AND RETURN

The approach to estimating campaign reach varies depending on the data available from each media platform and the particular project's specifications. In general, the percentage of households from the NCS dataset that have been exposed is multiplied by a universe number to project to the general population. For example, in digital campaigns using third-party cookie matching, the number of unique households reached is divided by the total number of active online households in the NCS dataset to obtain a reach percentage, which is multiplied by the total number of online households—estimated at 96.5 million as of 2016—to get a final count. The reach percentage for linear TV studies comes from Nielsen Ad Intel and Nielsen's national TV panel. For some media types, such as magazines, the publisher provides an estimated reach.

For cross-media studies, reach can be estimated for the entire audience that saw the campaign on a certain platform. Alternatively, it can be estimated for the audience that only saw it on that platform, with an overlap group capturing households that saw it in more than one place. Estimation of the overlap is made possible by exposure data from a large number of partners, and research is ongoing to improve estimates not only between partners on the same platform (such as multiple digital publishers) but also between platforms, such as linear TV and digital.

The total amount of incremental sales attributable to the campaign is calculated by multiplying the per-household incremental sales by the reach number. This calculation is based on the assumption that incremental sales are consistent between households in the sample and exposed households not represented by NCS purchase data. The samples on which the studies are based are robust enough that this assumption can be justified.

With total incremental sales calculated, other key indicators such as return on advertising spend (ROAS) can be calculated and compared to past campaigns and to benchmarks for the specific product category. With the insights provided in these studies, advertisers can determine effective ways to improve future campaigns. Single-source data takes the guesswork out of determining which aspects of a campaign are driving results.

CONCLUSION

Sales effect studies are a powerful tool available to researchers to measure the direct effect that an ad campaign has on in-store sales. One avenue for future research is to enhance the list of variables used in the matching process. While the purchase and demographic variables currently in use are robust and relevant to product buying patterns, there might be others that could help improve the attribution of incremental sales to the ad campaign. The relative weights of the variables could also be refined, along with the way the values are standardized.

This research is opening new doors to understand advertising effectiveness. For instance, with buyer basket analysis, we're starting to understand how ad exposure might affect cross-buying behavior—that is, consumers' propensity to purchase other products alongside the item being studied, during the same shopping trip. We're also starting to understand the many ways in which exposure on multiple platforms might combine to affect purchasing. The opportunities provided by single-source data are virtually limitless and have only begun to be explored. [1](#)

METRICS AND CALCULATIONS

Average Incremental Sales (can be in dollars, volume, units):
 $(\text{Test Group Spend} - \text{Control Group Spend}) / \text{Total Analysis Households}$

Total Incremental Sales:
 Average Incremental Sales X Total Households Reached

Penetration (percentage of households that made a purchase):
 $\text{Number of Buyers} / \text{Total Analysis Households}$

Buy Rate (average household spend on the brand during the post-period):
 $\text{Average Incremental Sales} / \text{Penetration}$

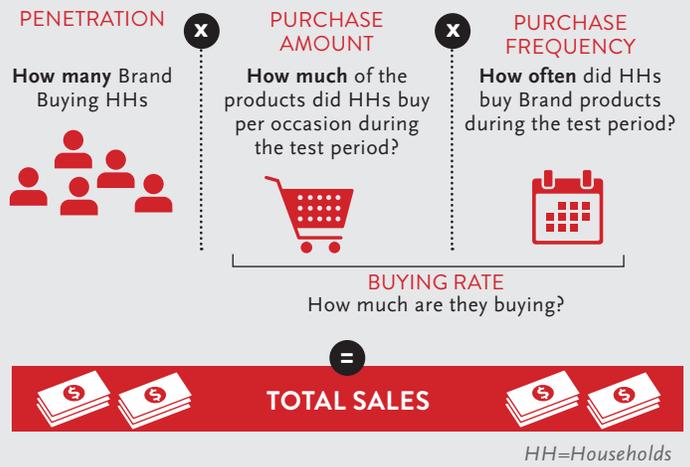
Purchase Frequency (average number of purchases made during the post-period by households that bought the brand being studied – this excludes non-buyers)

Purchase Amount (average amount spent on each purchase occasion):
 $\text{Buy rate} / \text{Purchase Frequency}$

ROAS (return on ad spend):
 $\text{Total Incremental Sales} / \text{Spending on Campaign}$

DPM (for digital studies, return per 1,000 impressions):
 $\text{Total Incremental Sales} / \text{Impressions} \times 1000$
Corresponding metric for TV studies is DPP, or return per GRP

SALES IMPACT CALCULATION



nielsen
.....