

CROSS-MEDIA MEASUREMENT

**The new medium necessitates
a new approach to an old problem**

Rex Briggs

INTRODUCTION

The Dove Nutrium Bar Cross-media case study represents a breakthrough in marketing mix measurement. It answers the question of where Online fits into the marketing mix. In answering this question, it applies a combination of best practice research techniques, which this paper will review. The study, led by Rex Briggs in partnership with Unilever, MSN, the ARF, the IAB and supported by Dynamic Logic research technology, provides a real-world, side-by-side, dollar-for-dollar comparison of Television, Magazine, Radio and Online effectiveness. The methodology has now been replicated and refined for six additional marketers. The findings shed new light onto the complementary nature of Online and Offline advertising and reinforce the idea of “surround sound marketing.” However, the purpose of this paper is not to review the findings. Rather, the purpose of this paper is to examine the use of best industry practice to develop this break-through research approach. This paper begins with the history and challenges of cross-media measurement and

then dissects the path-breaking methodology and analysis used to overcome these challenges.

MARKETING MIX TAKES THE STAGE

Billions of dollars are spent each year promoting brands through Television, Magazine, Radio and Online advertising. What if balancing the marketing mix could make this investment five percent more efficient in achieving the marketer's branding goals?

The hope of improving results within the same advertising budget is the core argument that has fueled the growth in marketing mix modeling. Think of a marketer with annual sales of \$400 million dollars: a five percent increase in sales is \$20 million dollars without spending more on marketing.

Marketing research, such as matched market tests, regression models, and telephone tracking studies, have made some contribution to helping marketers understand the proper balance in the marketing mix, but all the old approaches have fallen short of properly measuring Online advertising's contribution.

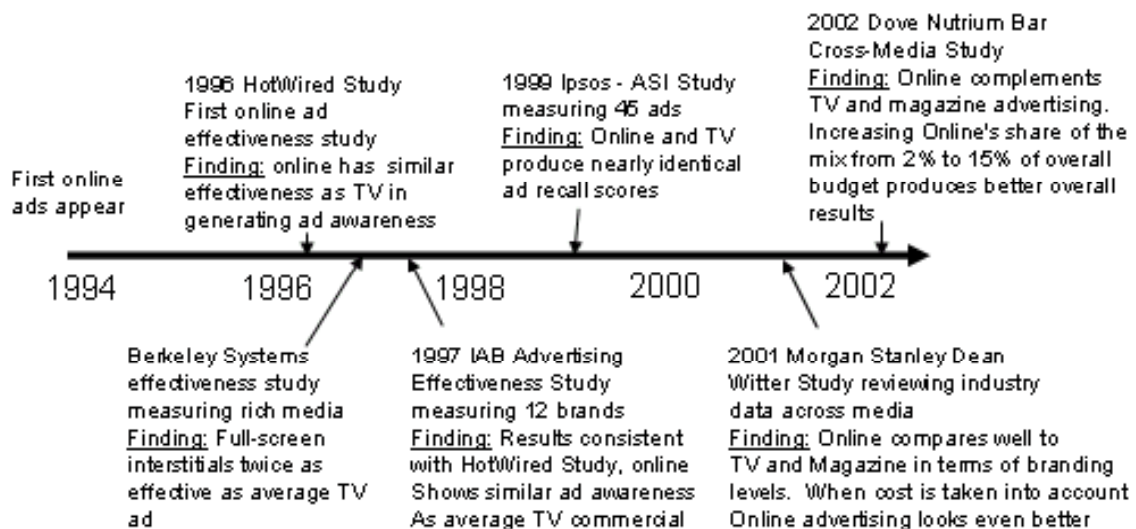
Why do these older methods fail to measure Online's contribution? What has been done to date to measure Online's role in the marketing mix? What new research methodologies can we devise to answer marketers' nagging question, of where Online fits in the mix?

This paper will first address the research efforts to measure Online's contribution, and the objections to the previous research. We will then turn to a discussion of why the old research approaches (designed to measure traditional media) do a poor job measuring the relatively new medium of Online. The balance of the paper addresses the solution to the measurement challenge and details the methodology and analysis by using Dove Nutrium Bar as a case study.

A BRIEF HISTORY OF MEASURING ONLINE VS. OFFLINE (1996 TO 2001)

The first web advertisements began appearing in late 1994 on websites such as HotWired. Within less than two years, marketing researchers responded to marketers' need for quantitative understanding of the value of these new ad units with the first advertising effectiveness study. Over the next five years several studies would tackle the thorny question of how Online advertising compares to Offline advertising. Figure 1 provides a timeline of these key studies.

Figure 1
TIMELINE



I had the privilege of developing and leading the first research effort to measure Online advertising effectiveness back in 1996. The main goal of the 1996 HotWired Advertising Effectiveness Study was to measure the effect of Online ads in the real-world using a classic exposed/control experimental design.

The way the study worked is that we randomly sampled those visiting the Hotwired homepage. Before the homepage loaded, a survey invitation took over the page. After completing a very brief survey, which collected an email address and basic demographics, users were “returned to their regularly scheduled programming”. All respondents were returned to the same page, however, half were blindly given the “exposed” ad and half the “control” ad. (By blindly we mean that consumer’s experience was exactly the same in both control and exposed groups and neither was aware that the ad they were seeing was part of research.)

We followed up a day later with an email invitation to answer “a few more questions”. The surveys were exactly the same for “control” and “exposed” respondents and measured the brand relationship metrics (awareness, brand image, purchase intent) for the “exposed” brand. The “control” group provided the baseline branding level and the “exposed” group showed if there was a statistically significant lift in the key metrics.¹⁾

By virtue of the gold standard experimental design, we could isolate the effects of Offline from Online marketing, thus precisely quantifying the independent value of Online advertising.

While this gave us a best practice for measuring the independent contribution of an Online advertisement to the overall branding levels, we wanted a point of comparison to traditional advertising. In sum, this research had taught us that Online advertising could increase branding metrics; the question that remained was, “how do these results compare to Television or Magazine advertising performance?”

From the beginning, we stretched to compare the results of Online to Magazine and Television advertisements. Thanks to Nigel Hollis’ innovative thinking, we compared the three Online ads we measured in the HotWired study to Millward Brown’s database of real-world Magazine and Television ads.²⁾ Specifically, MBI had a measurement of TV and Magazines’ effect on ad recall based upon the first exposure. Our measurement of ad recall for the Online ads paralleled the methodology and Nigel felt we could therefore compare the data. The finding was that these three Online ads compared quite favorably to the averages of Offline ads in terms of ad recall.

This research process and Online/Offline comparison was repeated in 1997 with four full-screen “interstitial” advertisements³⁾ and then again for twelve Online banners measured in the 1997 IAB Advertising Effectiveness Study.⁴⁾ Comparison of the Online ads to the average for Television and Magazine ads again showed favorable results for Online advertising.

While this approach provided an innovative comparison of Online and Offline advertising, it had its share of problems. In particular, two problems concerned me.

First, we were comparing Online and Offline against “advertising recall,” the only comparable variable Millward Brown had available across media. Most marketers seemed less interested in advertising recall and would have rather seen data on sales, brand image or brand awareness.

Second, and the more serious problem, we were comparing specific ads to averages for Television and Magazine advertising. While comparing norms is a common practice, critics are correct when they complain that such a comparison of norms is not precise. Each brand has unique characteristics in the marketplace and therefore a proper side-by-side study of the same brand advertising across media is a better answer to the relative effectiveness of the Online vs. Offline question. In other words, the marketers were asking for a side-by-side comparison of the same brand across Online and Offline advertising to answer the question of Online/Offline advertising effectiveness.

Side-by-side Comparisons: Ipsos-ASI Responds

While I was hard at work developing real-world measurement for Online advertising, the research firm Ipsos-ASI was hard at work re-tooling their

“laboratory setting” Television pre-test to work for measuring Online advertisements.

What the laboratory setting of Television pre-test research lacks in real-world measurement, it adds in greater control of the consumer experience. By directing consumers to do something specific, like watch a TV program, the research may bias the attentiveness of the consumer (so-called memory priming effects), but the benefit is easy side-by-side measurement of one advertisement vs. another.

Typically, the side-by-side comparison has been used to determine whether one Television advertisement would work better than another for the same brand. To get a mental picture of how these studies work, imagine that you are strolling through a shopping mall and a perky interviewer asks you if you’d like to answer some marketing research questions in exchange for \$10 for your time. You agree.

You and a few dozen people are put in one of two rooms. In the front of the room is a large Television. You watch a thirty-minute Television program and then are given a survey about the Television commercials. You are asked questions such as “do you recall seeing advertising for Brand X.” At the same time you were watching your program, the same program was playing in the next room, however a different commercial for Brand X was playing. By comparing the two groups, the marketing research firm can judge which advertisement is stronger.

Now, imagine the same two rooms of the research facility. This time, instead of a Television in both rooms, the second room contains computers. Half the respondents are randomly directed to the room with the TV, and the other half are directed to the room with the Internet.

You are part of the second group and find yourself in the room with the computer. You surf a website, and are then asked the question about whether you recall an ad for Brand X.

The results from those in the room with the Internet are compared against the respondents in the room that saw the Television advertisement. The researcher judges the relative effectiveness of Television vs. Online advertising.

Such a study comparing Online to Television was conducted on 45 different advertisements for major brands such as Gap, Ralston-Purina and Kraft. In a presentation to the ARF, the report’s author concluded: “If we put ten people in a room, show them a TV program, turn it off and ask them how many remember the ad for Brand X ... four will. If we put ten people in a room, show them some content on a website, turn it off and ask them how many remember the ad for Brand X ... four will.”⁵⁾

In other words, the study found that Online ad recall is about the same as Television ad recall. This 1999 research by Ipsos-ASI answered client's request for a side-by-side comparison of Online and Offline advertising for the same brand. The research was ground-breaking in its own right.

However, those wishing to dismiss the conclusions found it easy to point to the fact that the research was conducted in an artificial research environment. The critics reasoned that perhaps the results might be different if the Television and Online advertisements were measured in the real world as opposed to forcing pre-screened people to watch the ads in a research room.

What these critics wanted is a "real-world" side-by-side comparison of Online and Offline advertising and not a forced side-by-side exposure pre-test of the ads. And, furthermore, they wanted to know the dollar-for-dollar comparison of the two media – not just a comparison of ad recall percentages. In other words, they wanted to know if I invest in Online, what is the ROI, and how it compares to the ROI from TV or Magazine advertising.

Enter Morgan Stanley Dean Witter

In February 2001, the investment firm Morgan Stanley Dean Witter completed an exhaustive study of the financial value of Online and Offline advertising and direct response vehicles. The basis of the study was the collection of advertising effectiveness metrics, cost data, response rates and so forth. According to the report: "We have conducted an in-depth analysis of data gathered from many studies by AdRelevance, Avenue A, The Cable Advertising Bureau, C.E. Hooper, Inc., The Direct Marketing Association, Dynamic Logic, Media Dynamics, NCH, The Magazine Advertising Bureau, and Nielsen Media Research to form our conclusions."

The report was certainly a milestone in that it drew from so many sources of real-world data in an attempt to provide a "real-world" answer to the question of how Online stacks up against Offline advertising. Furthermore it broke new ground by doing what financial analysts do so well and that is making a dollar-for-dollar comparison.

The conclusion is best summed up by quoting from the report directly: "Our primary conclusion is that Internet advertising banners are a cost-effective branding tool ... Banners exceed or are as good as Magazines, newspapers and Television in generating brand recall and brand interest ... When branding is considered in terms of cost effectiveness, banners look even better."⁶⁾

This research was certainly intriguing. But by comparing averages from a variety of sources, it was difficult to say with certainty the relative effectiveness of a specific Online advertisement vs. a specific Offline advertisement.

Despite the converging validity of the 19 real-world ads measured in the Hot Wired, Berkeley Systems and IAB studies, the 45 measured by Ipsos-ASI, and the exhaustive Morgan Stanley Dean Witter analysis, critics would claim that the “Holy Grail” for placing Online in proper perspective within the marketing mix would be a real-world, side-by-side measurement of a campaign that was running in Online and Offline vehicles at the same time.

In the author’s opinion, it would be beneficial for research to go beyond “Online vs. Offline” and begin to look at Online alongside Offline advertising – to see how the two might complement one another and provide a more balanced and effective marketing mix.

FROM ONLINE VS. OFFLINE TO ONLINE’S ROLE IN THE MARKETING MIX

The previous research compared Online vs. Television and other media. In the 1999 ESOMAR paper that Horst Stipp and I presented at NetEffects⁷⁾ we argued that we should strive to measure Online advertising in context with Offline advertising. We suggested that there were hints of Online and Offline synergy in the early research studies. The challenge was how to measure Online’s contribution to the marketing mix.

This is the challenge that a collection of dedicated marketers and researchers would rise to address with the Dove Nutrium Bar Case Study. But, before we discuss how the Dove Nutrium Bar case study provided a “real-world”, side-by-side, dollar for dollar comparison of Online and Offline advertising, let us address the broader challenge of measuring Online’s role in the marketing mix. Specifically, let us address why older marketing mix measurement approaches generally fail to measure Online’s contribution (and why we needed to develop a better measurement approach).

The Failure of Traditional Marketing Mix Measurement and the Need to Create a New Measurement System

With the growth of Online advertising, some marketers have turned to the old tried and true methodologies to shed some light on the value of their Online marketing investment relative to their Offline spending.

The problem is that these “tried and true” methodologies fail to measure Online’s contribution in all but the most specialized of cases. In fact, if marketers are looking to have a bright spotlight focused on the value of the Online spend, the best that these methodologies can yield is a dim, flickering candle.

At the heart of the problem is the fact that Online is simply a small part of most marketers' current advertising plans. To properly measure Online's contribution, the researcher must know who has been exposed and who has not. Consider, for a moment, the traditional approaches and the unique challenges of measuring Online. Regression models, matched markets, telephone tracking – each suffer from key limitations when it comes to measuring Online advertising's contribution.

At the heart of the measurement problem is the need to identify who is exposed to what media at a respondent level. Most of the old marketing mix methodologies approached the issue of identifying who is exposed to what media in a very simplistic fashion. We needed a more sophisticated approach. There are solutions to this conundrum. For each of the traditional research approaches, we will highlight how to update the methodology to measure Online's influence.

Modeling of Advertising and Sales Data

The idea here is to take the marketing inputs such as the level of spend and regress them against the outcome variables such as sales. Whether one uses multiple regression, or more complex neural, Liserl, Bayesian, or another modeling technique that looks at the total marketing inputs vs. outputs, each run into a common challenge. The challenge is that without knowing who specifically is exposed to Online advertising and who is not, the tiny amount being spent to buy Online ads simply washes out of these models as a rounding error. I call this the “small investment problem”.

This is not to say that modeling does not hold promise, however, since Online represents only about 2% of most advertising budgets, it is too small to be meaningful when modeled against the variance inherent in weekly (or daily) sales figures from store or panel level data.

Regression modeling (and its analytic cousins) can be made to work if one of the two conditions is met: 1) the marketer makes Online 20% of their budget so that the model can better detect the influence; or 2) the marketer determines which respondent is exposed to which media.

Determining who is and who is not exposed to Online ads is probably more palpable to an advertiser that may be reticent to plunk down 20% of their budget into Online for the purposes of conducting a proper marketing mix model analysis. Determining who is exposed and the influence on sales will probably entail some type of empanelment approach. Such an approach can link sales to respondents for whom Online advertising exposure is known. For example, based on some of my consulting advice, IRI and DoubleClick linked IRI panelists with DoubleClick's ad server, thus enabling exposed/not exposed

experimental design for Online advertising. This can be combined with Offline advertising data to perform a proper marketing mix analysis on the influence of sales among those not exposed to Online ads, thus providing a full picture of Online and Offline advertising's influence on sales.

Matched Markets

Matched market measurement has been used for many years to evaluate TV, Radio and other marketing programs. Rather than spend nationally, marketing programs can be cost-effectively tested with a small investment. The way this works is rather than spending 20% of the marketing budget nationally, certain markets are chosen so that media can be "heavied up", thus allowing the marketer to measure the impact on sales compared to the control matched market.

This approach has worked for Offline media, but this approach fails to work for Online advertising. It does not fail for the typical problem of actually finding well-matched markets (which can be a problem with matched market studies). It fails because Online's regional targeting capability works well for avoiding advertising waste, but is generally limited when it comes to helping the marketer "heavy up" a specific market.

Yes, there is geographic targeting for Online ads, but it works differently than geographic targeting for traditional media. Currently, Online publishers who sell geographic targeting do so to avoid waste. In other words, if I buy Online ads only in the New York City ZIP codes, the Online ad server sends ads only to those web browsers that can be confirmed to some degree of certainty to be in New York City. If the ad server cannot tell where the user is based (which is a big chunk of the time), it simply skips over the user and does not serve them the regional ad – even if the person does in fact reside in NYC.

Why might the ad server skip over someone? Let us say the user dials into America Online. Unless the user has provided declared data on where the user is based, the ad server reads the user location as Dulles, Virginia, which is where AOL is based. The ad server has been programmed to skip over AOL users that lack declared data since it cannot determine where the user is really based. The Online local ad serving is a godsend for those marketers wishing to avoid waste. The consequence of this technology solution for avoiding waste is that a marketer generally can't "heavy up" Online to the degree necessary in the market for the purpose of research.

The geographic ad serving technology is great if I want to avoid serving ads to people outside of New York City. Even though many people within New York City are likely to be skipped over, the marketer gets those that the ad server can determine with certainty are based in New York City. But if I want to get

the highest possible Online reach within New York City, then “skipping over” those that cannot be identified as living in the city makes it very difficult, perhaps impossible, to heavy up a market without contaminating the control market.

Therefore, given the current state of geographic targeting Online, matched markets are not an option for cross-media measurement. As with regression modeling, we can overcome this challenge by devising a strategy to measure at the respondent level who is exposed to Online advertising and who is not exposed. If Online ads can be served to panelists based on location, then matched markets can be used to detect different media combinations. For one of Marketing Evolution’s clients, Radio ads were in ten cities and Online was national but included a control group. We could therefore measure the combined effect of Online and Radio and compare it to the effect of Radio only or Online only, but we were only able to execute this analysis because we could detect on a respondent level who was and who was not exposed to Online ads.

Telephone Tracking

Telephone surveys have been used for decades to examine points in time prior to a Television ad airing and the points in time after. The innovation nearly thirty years ago was to shift from waves of surveys to continuous surveys. Moving averages of branding metrics were trended against advertising levels. Periods of hiatus offer the opportunity to measure the pre-campaign vs. the post campaign effects. Can telephone tracking studies be used to measure Online contribution? The answer is generally “no”. Telephone tracking does not accurately measure Online advertising.

The telephone tracking approach runs headfirst into the “small investment” problem. If only one in ten consumers are exposed to Online ads, how can telephone tracking detect it? Even if exposure to Online ads boosted sales by 10 points, the tracking measurement would only measure a one point gain (10 points x 10% reach = 1 point). And, given the typical sample size of 50 to 100 respondents per week, this increase of one point would fail to pass statistical significance testing.

In other words, unless the reach of Online is in the 45% or greater range,⁸⁾ the top down telephone tracking approach will mask the effect of Online advertising exposure even if Online’s effect is substantial.

This is not just a problem with Online advertising. Magazine, Radio, and Newspaper promotions are also penalized by the telephone tracking bias toward “big reach” TV advertising campaigns.

The solution to Online's small investment problem is to determine who is and who is not exposed. We can determine who is and who is not exposed by moving the survey Online and "tagging" all Online ads.

Tagging ads is a procedure that results in a cookie being set each time an advertisement is delivered. This cookie can be automatically read when a consumer is randomly surveyed. For each respondent, we can simply observe whether or not he or she has been exposed to the advertisement. This now allows us to measure the affect of Online exposure, even if the Online campaign is relatively small.

Continuous Sampling with Cookie Ad Tracking

Online surveys to measure real-world effects of Online advertising have been in practice for the past six years. The approach has been executed on hundreds of campaigns by a variety of research firms and has received wide industry support. Expanding this proven research approach to measure cross-media campaigns holds tremendous promise.

There are two other key benefits of conducting the survey Online, and one potential drawback. First, the benefits: not only can we measure Online's contribution even if the spending level is fairly small, another wonderful benefit of moving the survey Online is the ability to measure Magazine's impact more precisely.

Magazine campaigns typically have lower reach than Television and are often more difficult to properly measure. By conducting the survey Online, we can show respondents covers of Magazines and ask "which of the following Magazines have read or looked through". This recall-based measure links back to the vehicle reach and is consistent with the way that Magazine advertising is bought and sold.

Another benefit is the ability to achieve larger sample sizes than could be achieved with telephone, face-to-face or mail surveys. For the Dove Nutrium Bar study we surveyed 12,990 consumers over six weeks. The cost of conducting such a survey via telephone would have been many times more expensive.

The downside of conducting the survey Online is that we are systematically excluding those that do not use the Internet at all. How big of an issue is this? We can calculate the degree to which this sampling approach might bias the results through sensitivity analysis, or we can add in a telephone fill-in sample to double check trends among those without Online access. Since the potential for sample frame bias is an issue, let us address the question directly.

Addressing the Issue of Sample Frame Bias

If the purpose of the study is to quantify the value of Online advertising among those with Online access, then sampling those with Online access is ideal.

However, if the goal is to understand the effect of both Online and Offline advertising among the total target population, then we must examine the effect of only surveying those with Online access. To calculate the degree to which our findings are sensitive to the sampling frame bias, we begin by measuring the percentage of the target audience we are excluding by using an Online sample from.

This is an important place to start because the smaller the excluded percentage, the smaller the potential bias (the potential for difference in the excluded group to change the overall results is proportional to the excluded groups population percent).

According to the UCLA random digit dial telephone survey of the U.S. population conducted during the same time that we were conducting our survey for Dove Nutrium Bar, 82% of women ages 25 - 49 used the Internet either at home, work or some other location (excluding email). Nielsen Net Ratings estimates the percentage using the Internet at closer to 60% (though unlike the UCLA study where the enumeration methodology is publicly documented, it is not entirely clear exactly how Nielsen's enumeration studies count usage and locations of access). For arguments' sake, let us suggest that the percentage of the target population that uses the Internet is at the mid-point between the two estimates: 70%. This would imply that we are systematically including 70% of all consumers and excluding 30%.

Since we are including 70%, we can feel confident that our data is projectable to most of the target audience. What we need to consider is the potential impact of the 30% we excluded.

Consider the impact on purchase intent. The baseline measure pre-campaign is 8.7%. We examine those people that did not see any Online ads. In other words, these people only have an opportunity to see Television and Magazine advertising. This is the same type of advertising exposure that the 30% of non-Online users would have. If we look at those in our survey that saw only Television and Magazine advertising, we find purchase intent at 11.5%. That is a three point gain over the six weeks of tracking. That is a very respectable increase from Television and Magazine advertising that would lead most brand managers to congratulate their advertising agency.

We could assume that this same increase observed among the 70% of the target audience that accessed the Internet within the past month is projectable

to the 30% of the audience that also saw TV and Magazine advertising, but does not have access to the Internet.

If we make the assumption that the two samples have similar reaction to TV and Magazine advertising, then our results are generalizable to the entire target audience.

But what if those that do not access the Internet are more likely to want to buy a product advertised on TV and Magazine advertising than those that do not access the Internet? If this is the case, then we need to examine the impact on our overall results, and recommendations to determine the practical implications. We can examine the potential impacts with parallel telephone tracking of those without Online access.

Two of the study participants opted to run a parallel telephone tracking study. While these studies add significant cost, they are a definitive way to address the question of sample frame bias.

In performing the analysis of the Offline sample vs. the Online sample, we screened those in the telephone study to only include those without Online access. We also removed those in the net-user sample that had seen Online ads (to remove the incremental branding effect) and weighted both populations back to the U.S. Census population targets for age and sex. This provided an apples-to-apples comparison of the Internet users and non-net users.

We compared the two data sets and found that there were no statistically significant differences between net users and non-net for unaided awareness (first mention and all mentions), nor were there any differences on aided brand awareness, however, we did find a difference on aided product awareness. This difference is caused by the fact that an Online survey question showed a product shot while the telephone study simply provided a verbal cue. This led to much higher recognition among those seeing the product shot vs. those simply provided with a verbal cue over the phone, without the visual aid of a product shot.

Since the rest of the survey is based on whether a consumer is aware of Colgate toothpaste, the two datasets are not easily compared. This challenge notwithstanding, we attempted to compare brand attitudes and purchase intent among those aware of Colgate Total within the two datasets. We found the following:

- Net users and non-net users have nearly identical purchase intent levels, but net users have more positive brand perceptions overall than non-net users
 - Non-net users show a slight down-trend in the last week of TV advertising, while net-users show a continued gradual increase in positive
-

brand perceptions in the last week of TV advertising. On this metric, the net sample shows a more intuitive trend against television advertising weight and timing

In summary, the similarities between the datasets leads us to conclude that using a net-user sample balanced to U.S. census targets would produce very similar data to that of the telephone methodology, with the exception being in the area of product recognition, where visual stimuli used in the Online survey produce a different measurement than a less cognitively rich verbal prompt read over the telephone.

The Kleenex brand also performed a telephone comparison study. Kleenex's goal was a quick check of validity of the Online sample and was therefore not as rigorous as Colgate's parallel study. Kleenex has an ongoing telephone tracking study which trends key metrics such as brand awareness. The trends observed in this telephone tracker paralleled the trends observed in the Online based study.

These points of converging validity give us confidence in the use of an Online based sample, weighted to census targets, to produce nationally projectable data for targets aged 13 - 64 years. For certain targets, such as the elderly (65+), we are more cautious since we know that the population is less likely to have Internet access compared to their younger cohorts. But for mainstream marketers, the Online sample approach not only allows accurate measurement of Online advertising (which is not possible without conducting the study Online), it also allows better measurement of Offline media such as magazines, and provides larger samples more cost-efficiently.

There is an alternative method to accessing the potential for sample frame bias. It does not require the expensive parallel telephone tracking study. We detail this alternative approach in the technical appendix at the end of this paper.

**INTRODUCING DOVE NUTRIUM BAR:
THE FIRST REAL-WORLD BRANDING CAMPAIGN MEASURED
ACROSS TELEVISION, MAGAZINE AND ONLINE
ADVERTISING VEHICLES**

In the fall of 2001 we completed a six-month effort to develop the first real-world, side-by-side measurement of a branding campaign running Online and Offline advertisements. The research represents the collaboration of the marketer (Unilever), their online and offline advertising agencies, a key publishing partner (MSN), industry associations including the ARF and the IAB, and researchers. The measurement approach has been described as "best

practice” through and through and received a nomination from ESOMAR for Best International Research of the year. Before we delve into the methodology, let us recap the objective.

The “Big Question” we all worked together to answer is, “Where does Online fit in to the marketing mix?” Online has grown faster both in terms of the potential reach and time spent with the medium than any other advertising vehicle, but does that mean a marketer should use Online advertising to help achieve goals? Consumers in the United States give Online between 10% - 15% of their total time with media,⁹⁾ yet advertising spending as a percentage of the total budget advertising is at a mere 2% on average.¹⁰⁾ Is this level appropriate given the advertising effectiveness of Online relative to Offline? This research was designed to answer that very question.

As we addressed this question, we were mindful of the previous research and attempted to answer critics by:

- conducting the research in the real world (as opposed to conducting a forced exposure study);
- using one brand advertising across media (as opposed to comparing norms across media);
- using a side-by-side dollar-for-dollar comparison (as opposed to comparing branding metrics in a vacuum without consideration of ROI);
- measuring a range of brands spanning multiple categories (as opposed to conducting a single study in one non-representative category – Dove is the first of a half dozen brands participating).

In response to the measurement objectives, we refined and adapted proven measurement techniques, and, where necessary, developed some new approaches.

What are the key findings of this breakthrough research? What are the implications? Online is a potent complement to Television and Online advertising. Adding Online allows marketers to increase the productivity of the marketing investment.

We found that overall branding results could be increased by 14% for purchase intent and 8% overall for Dove Nutrium Bar by increasing Online’s reach and frequency to more optimal levels. This 8% gain in branding effectiveness is achievable without increasing budget. Instead, if marketers reduce Television and Magazine frequency slightly (leaving reach alone) enough dollars can be freed to boost both Online reach and frequency levels.

Achieving expanded reach for Online and optimal frequency would require increasing Online spending to 15% of current budget. We have found

justification for similar increases in budget for Colgate, McDonald's, and Kleenex, and continue to expand our analysis to examine other brands. The implication of the research is that marketers should make it a priority to measure whether they too can achieve better overall results by expanding Online's role in the marketing mix.

The objective of this paper is not a discussion of the findings and implications of this research. Instead, the goal of this paper is to continue the discussion of the research approach for measuring the branding phenomenon of an Online and Offline advertising campaign. The goal is to communicate the best practice approaches used to develop greater support of this methodology.

What gives these results legitimacy? Why should you believe these findings? There are at least two reasons for taking heed of these findings. First, the study represents an impressive use of best industry practice to properly quantify who was exposed to advertising across each medium. Second, the collective effort of a diverse and brilliant team from Unilever, their advertising agencies, the ARF, Dynamic Logic and Marketing Evolution ensured a media neutral measurement and analysis that yields replicable results grounded in a precise measurement of a variety of advertising effects.

The heart of the measurement examines cause and effect. We measure whether exposure to advertising "causes" increases in branding metrics. So, how do we measure exposure? What do we define as "branding" effects? First, let us talk briefly about how we define branding, and then we will turn to the critical question of how we measure exposure. We will then discuss some of the key analysis, such as the impact of increasing reach and frequency and the calculation of dollar-for-dollar impact across media.

Measuring the Branding Metrics

We begin with the brand manager's marketing goals. We translate the goals into survey questions and use the consumer's response to the survey to quantify each branding metric.

We can, at the client's request, connect the exposure data to actual purchases. Dove was satisfied to measure core branding measures specifically. We measured three standard core branding metrics:

1. Brand awareness (unaided and aided);
2. Positive brand image;
3. Intent to take action (such as purchase the product).

And while we did not measure sales directly, it is certainly possible to extend the methodology to measure sales through empanelment of respondents. In this

case, the client opted to use their purchase probability measure as a proxy for sales. Since the survey question of intent to purchase is by itself media neutral, the measurement should provide a relative measure of the influence on sales by each media independently and in synergy.

By continuously surveying new groups of consumers over time (while the advertising weight levels wax and wane based on media schedules) we can trend the consumers' branding levels over time. We then link-up consumers' exposure to advertising to determine the effect. (For a fuller discussion of the branding metrics, please see the following Discussion of Key Branding Metrics.)

Discussion of Key Branding Metrics

Brand awareness: Brand awareness is measured in two different metrics. We analyzed unaided brand awareness and aided brand awareness. To measure unaided awareness we ask consumers to type in the soap brand that first comes to mind. If the goal of the marketer is to become more "Top-of-mind," and salient, then this unaided awareness question is a good measure of that branding goal. The other way we look at brand awareness is to measure "aided brand awareness." This means we provide the consumer a list of several brand names such as Dove Nutrium Bar and ask consumers which of the following are you aware of. If the branding goal is to generate higher overall brand recognition, than this is a good measure of that goal.

Positive Brand Image: Often times, the marketers' goals go beyond brand awareness. Often the goal is to create a positive brand association. If we are talking about the beauty bar soap, we want people to be informed that our product "contains Vitamin E", for example. And we want them to agree that the product will "nourish their skin." We also want the consumer to internalize the message and agree that the brand is "for people like you." These brand image statements allow us to hone in on whether or not advertising is creating the right kind of awareness and association. For the Dove Nutrium beauty bar, we measured five key statements.

Intent to take action: Our final branding metric is intent to take action. The goal of all marketing is to generate sales at some point. We do not expect that the Television ad, the Magazine ad or the Online ad will result in people jumping out of their chair and running to the grocery store to buy the product the instant that these ads are shown. Nor do we expect them to clickthrough, or dial an 800 number and order immediately. That is an unrealistic branding action in the context of this product. We do expect purchase action at some point in the future will result from these ads. For Dove Nutrium Bar, Unilever opted to use the standard purchase probability question "How likely are you to purchase this brand?" The choices included "definitely," "probably," "might or might not," "probably not," "definitely not."

Advertising Measurement in the Real World

The approach of tracking response in the real world is considered a gold standard in advertising measurement. It avoids directing consumers to watch a certain program, or visit a website and therefore avoids the memory priming problem inherent in forced exposure studies.

Considering our branding measures of awareness, brand image, and intent to purchase, you will probably agree that we have a fairly comprehensive measurement of advertising success. The key question is how we unobtrusively observe consumers in the real world to measure whether or not these advertisements moved the needle on key branding metrics.

Measuring Exposure – An Impression by any Other Name

Measuring exposure properly is central to the success of cross-media research. Since our goal is to make dollar-for-dollar comparisons of the effectiveness of marketing elements, it is critical to measure advertising exposure using the same approaches that the advertiser uses to pay for each advertising vehicle. Why is this so critical?

For better or worse, the definitions of what constitutes an ad impression paid for by the marketer varies from one media to the next. If we fail to account for these differences, we cannot link back to the financial data on media cost. This may be an intuitive point, but as it is such a key point I will elaborate.

Consider the potential difference between an Online, a TV and a Magazine impression. Each are used as a media currency and advertisers talk about the CPM (cost per thousand) as the denomination of their currency. For better or worse, 1,000 Television impressions means something slightly different from 1,000 Online impressions, or 1,000 Magazine impressions.

Let's start with Television. Purchasing 1,000 Television impressions for \$20, for example, does not necessarily mean that 1,000 people will sit through your thirty-second TV commercial with rapt attention. In fact, in the United States the behavioral ratings that define the audience size are not based on measuring consumers watching a specific ad. Instead, it is based on the average viewing audience within the quarter-hour that the ad appeared.

Is it possible that some of these 1,000 people that the advertiser paid for did not actually see their ad? Yes. If I am a viewer during that quarter-hour and I leave the room a minute before the ad starts and return a minute after the ad finished playing, I would still be counted as part of the 1,000 users.

Is this a problem for us and our attempt to measure the value of Television exposure? No, it is not a problem at all, as long as we link back to the cost paid vs. the impact when comparing the media.

In other words, if we are asking, “What impact does \$20 spent on advertising in a given media produce?” then we have the right measurements in place to answer this question. For instance, if \$20 spent on Television results in 60 people becoming aware of the brand (the increase over the pre-campaign level), then it is really beside the point whether 1,000 people really saw the ad and listened to every moment of the TV spot or whether the number is slightly different. The costs are the same and the results are the same regardless of the “true” number of actual attentive viewers.

Certainly the savvy researcher will be interested in the qualitative experience of the user included in the 1,000 media defined package. We might be interested to know how many of these 1,000 impressions count people in the room. How many of these people watch the entire commercial? How many of them press the mute button, and how many listen to the sound? How many are watching only the TV commercial and not multi-tasking (eating is a popular activity while watching TV)? There are plenty of qualitative questions that will influence the overall effectiveness of Television (or any medium for that matter). But if we want to do a proper dollar-for-dollar comparison, these issues are not relevant to our particular analysis.

The key issue that is relevant is what the advertiser paid, and what the advertisers got in return for their investment. The bottom line we focused on measuring was exposure to the media based on the same framework that the advertising agency uses to purchase the media – and that gives us the ability to analyze the data properly.

I have used Television as an example for the qualitative viewing vs. media buying definitions, but this same principle applies to Magazine and Online advertising. What we as an industry have defined as “an impression” (the number that advertising agencies use to buy a media with a commonly defined approximation of the number of eyeballs they are reaching) is used to define an impression in this study – and that consistency is valuable for proper measurement.

For Magazine advertising the definition of an impression is tied to those reading or looking through the Magazine vehicle. Does the Magazine consumer actually look at your ad? Do they read through it? Those are research questions that are addressed by qualitative companies, but they are non-germane to the return-on-investment analysis. We can simply focus on the industry-accepted definition of what constitutes an impression that the advertising agency pays for and the resulting branding increases. As with Television, we use cost vs. impact to ensure an accurate measurement of the contribution of Magazines.

Just as with Television and Magazines, we must look at what constitutes an Online advertising impression. Online impressions are currently measured through the process of an ad server (a computer) receiving and responding to a request to deliver a specific ad impression. This measure is not perfect. For a variety of reasons, not every Online advertising image fully loads. (The failure of an ad to fully load can occur if a user presses the stop button, or clicks on another link while the ad is loading, thus never receiving the full ad). And, just as with TV and Magazines advertising, a consumer may not give their attention to the advertisement. We perform the same procedure with Online that we performed with Television and Magazines: we apply the amount paid for 1,000 impressions and measure the impact to determine the value (ROI).

With the definitions of impressions and the rationale of using the media currency impression definition, let us specifically talk about the ads we measured. Let us look at target audience as well as the reach, frequency and timing of the ads.

DOVE NUTRIUM'S AD BUY

Let us discuss where you, as a consumer, might have seen the ads. The first question is whether you are part of the target audience. Defined in media buying terms, the Dove Nutrium bar's audience is women aged 25 - 49 years.

For those readers that are a female in the United States aged 25 - 49 years, over the six weeks in October and November 2001 (when we were measuring the campaign) based on the information provided by the advertising agency, about 85% of you would have seen the Television advertisements. On average, you would have seen six ads. About half of you would have seen the Magazine ads, and on average, you would have seen the ad three times. And, about one in ten of you would have seen the Online ads, and on average you would have seen the ad two times.

Now, this does not necessarily mean that you will remember seeing any of the ads, but these are the exposure reach and frequency numbers that the advertising agency paid for.

We can measure the effect of the campaign by comparing the branding levels after the campaign (post) to the branding levels before the campaign (pre). In other words, if we surveyed someone before the campaign started, we can get a measure of where the brand stands, pre-campaign. If we conduct surveys every day among a new set of consumers, then as Television, Print and Online starts, we can see if the advertising lifts the branding level by trending the combined branding effect of the advertising over time. Specifically, we can see if the campaign increased key branding metrics over the pre-campaign levels.

This type of measurement is typically referred to as continuous tracking. It was first introduced to advertising measurement in 1973 and has gained wide acceptance since then. In fact, of the top 100 advertisers, over one-third use continuous tracking to gauge their advertising effectiveness. It is a proven approach geared almost exclusively to measuring Television campaigns.

The innovation of this research study was our ability to look not only at the combined effects of the ad campaign, but also at the separate effects of Online, Print and Television.

Typically, continuous tracking studies survey only a few hundred people and it is not possible to look at subgroups such as those exposed to Magazine only. We surveyed many times the typical tracking research number. We surveyed 12,990 for Dove over six weeks. This gives us more than enough people to analyze the effects of each media separately and in combination with one another.

How did we pull apart the independent effects of Television, Online and Magazine advertising? We did not want to bias consumers answers to our branding question by prompting consideration of media habits, so here is what we did: We randomly surveyed consumers and asked our branding questions first. This gives us a “clean” read on how branding levels trend over time.

Later in the survey, we asked the media consumption questions to allow us to determine advertising exposure levels using the same “currency” definition used by the media. To be specific, we used the industry sanctioned impression definition to measure consumer’s opportunity to see advertising within each media.

When we designed the sampling structure we ensured that we had a large random sample of consumers. We augmented this sample with a random over-sample to create large “control” and “exposed” groups to better measure the precise effects of Online advertising relative to TV and Magazine. Let us go through each media and discuss our measurement of the opportunity to see an advertisement.

Television

For Television, we knew the exact start date of the advertising, as well as the reach and frequency. We applied this information to those surveyed while Television advertising was running. If we surveyed 100 respondents on day four of the study, we know the opportunity to see (OTS) TV ads from this campaign is zero – TV ads did not start appearing yet. If we surveyed 100 respondents on day 42 of the study, we know that TV’s campaign reach is 85%, which means 85 of the 100 people had an OTS for the TV campaign.

Our approach gives us a good measure of the opportunity to see Television ads across the population surveyed after TV starts. We can further define an individual's opportunity to see the TV campaign by arraying respondents based on their self-reported number of hours of TV watched yesterday, past week, times of day watched over the past week, specific programs and channels watched each day over the past week, etc. These media habits were measured at the end of the survey and allow us to calculate a probability that any given respondent was reached by TV (based on daily cumulative reach) and the frequency of exposure (based on cumulative quintile frequency). Someone who watches lots of TV during the same dayparts Dove was advertising is calculated as more likely to have been reached than someone who watches very little TV, or who never watches during the dayparts Dove was advertising on.

The accuracy of the assignment of individual TV reach and frequency can be cross-checked by comparing the population reach adjusted average to the respondent assigned exposure probability. To be specific, the baseline for purchase intent is 8.7%. If that level is increased 1.3% points to 10% among those not exposed to Magazine or Online advertising when TV has a reach of 80%, then we calculate that the true effect of TV is 1.6% (or an increase of 10.3).

We arrive at this conclusion by observing that only 80% of those surveyed were reached by TV. Therefore if we divide the branding increase by 80%, we arrive at the branding increase among those reached by TV.

Since the calculated 1.6% increase is a population-based figure that required no assumption or modeling of exposure, we are extremely confident in this number. We can now use this number to compare our individual respondent level derived "reached by TV" definition.

The respondent level calculation for TV reach does require the assumption that those that watch more TV during the dayparts that Dove was advertising on are more likely to be reached by the TV ads. By cross-checking the respondent data to the population data, we ensure our assumption for assigning individual TV reach and frequency is sound and produces accurate data.

Magazines

We have similar reach and frequency data for Magazines. To figure out which consumers were reached with Magazines, we showed consumers covers of 27 Magazines. We asked respondents to indicate which they had "read or looked through." (We also gave respondents an opportunity to fill in additional magazines they read or looked through in open ended-fields so that no respondent would feel compelled to "choose the closest magazine" to signal

that the respondent reads.) The approach of asking consumers to indicate the magazines they have read or looked through is fairly consistent with how the ratings for reach that the agency uses are derived. (In fact, using the actual magazine covers exceeds current best practice in Magazine measurement.)

Because Magazines accumulate over a longer period of time, we went back about two months and included titles that were still accumulating readers. We also updated our Magazines list the day new issues hit the newsstands. Some of these 27 Magazine issues contained ads for Dove Nutrium Bar, and some did not.

In fact, over the course of the study, about one-third of the Magazines we included in our list contained Dove Nutrium Bar ads. Therefore, we can code respondents that indicated that they had “read or looked through” any of the Magazines that contained Dove Nutrium ads as having an opportunity to see Print advertising for the brand.

As a point of comparison, we examined the agency forecasted reach with the survey reported reach. This self-reported data is very close to the ratings estimated data the advertising agency uses to buy and pay for Magazine advertising. The agencies reach number based on ratings is 53% while our self-reported data (which covers a two-week shorter time period) indicates a reach of 48%.

Online

For Online, we use the industry definition of ad request impressions. Thanks to Dynamic Logic’s Ad Scout system, every single advertisement that Dove served Online included a transparent pixel. Serving this transparent pixel allows us to record the exact time and location that every ad is served. When we survey consumers, we can simply examine this behavioral data and use it in our analysis.

We can see, for instance, that one consumer did not see any advertisements for Dove Nutrium Bar, while another consumer had an opportunity to see seven ads over the past three weeks.

Our approach to collecting advertising exposure data for each media represents a harmonization with the industry accepted currency. We came to this measurement approach after consultation with the Advertising Research Foundation (ARF), the client, the advertising agency and other industry experts. Magazine ratings tell us that the average frequency over the Dove Nutrium Bar campaign is about three. But if a researcher chose to use a different qualitative definition of what constitutes an impression – such as a decision rule requiring the consumer to have looked at the headline and some of the text body – then maybe the count would really be 2.5 or 3.5. Regardless

of qualitative definitions, we link back to what the agency pays for the impressions that will generate the three, and therefore we want to align our analysis with the “currency” of the media buyer – and that’s precisely what we did. The measurement is therefore neutral. And because we tie back to the currency, the approach does not favor one media over another.

Analyzing Cause and Effect

Let us turn to how we apply this exposure information and branding metrics to our analysis. If we look over time at the Dove Nutrium Bar campaign, you can see that starting on the left of the figure 1, on October 11th, we began surveying consumers. Other than a few old Magazines that were still gathering their last bit of reader accumulation, the campaign was dark. There were no Television spots and no Online ads. By removing the few people who indicated that they read or looked through one of the old Magazines that we coded as having contained Dove Nutrium Bar ads, we are left with a pure read on the “Pre-campaign” branding levels. (See the first image in figure 1).

A baseline level¹¹⁾ from which we can measure the branding gain is important. For purchase intent, for example, the level is 8.7% (definitely will/probably will buy).

As we continued to interview consumers in the next week, new Magazine ads began to appear within the new issues of titles such as *Glamour* and *InStyle*. Online ads also began to appear. We separated consumers that saw Online ads, and did not read or look through Magazines that we coded as containing Dove Nutrium Bar ads. This allowed us to analyze the branding effects of those who were exposed to Online ads only. (See the second image in figure 1.)

Figure 1
DISSECTING ADVERTISING EFFECTS

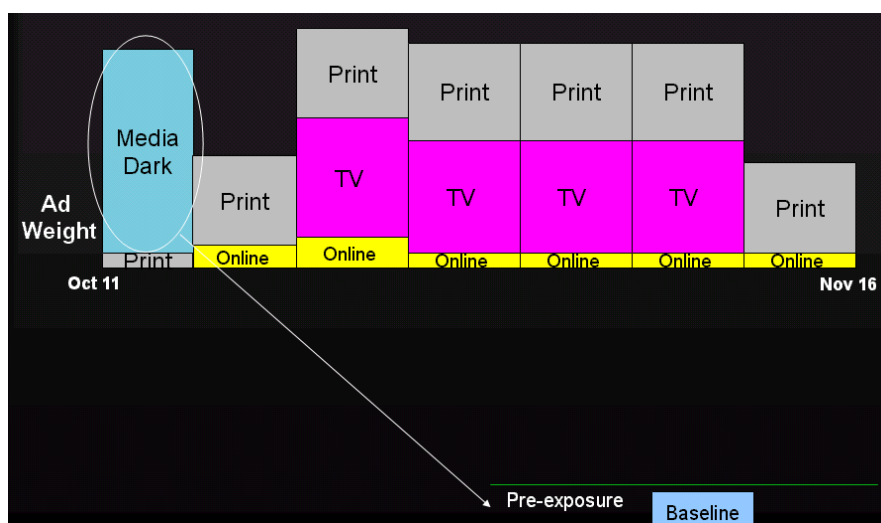
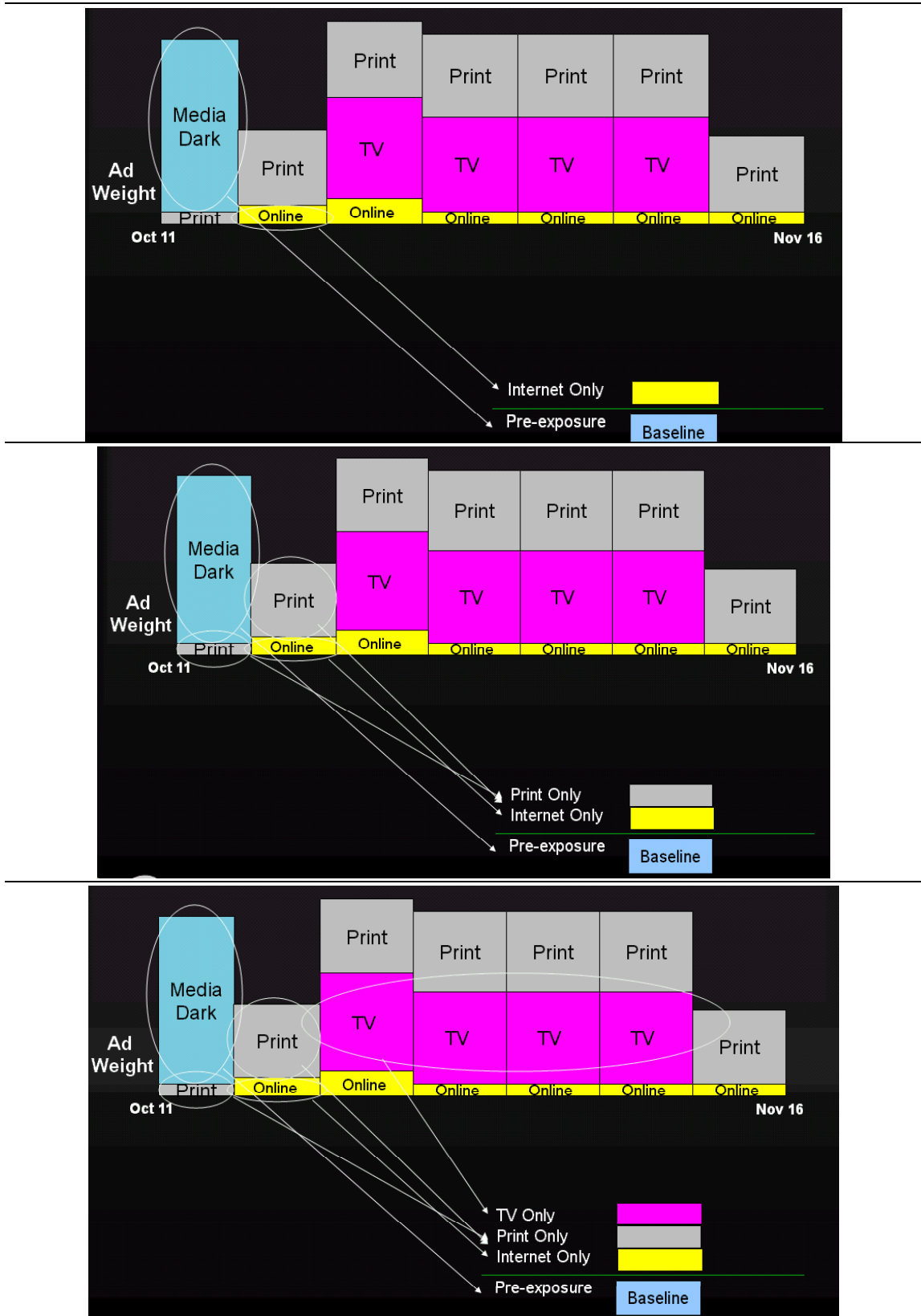


FIGURE 1, CONTINUED



We could apply the same principle to Print ads. We can look at people who read or looked through Magazines that we coded as containing Dove Nutrium Bar ads and, subtract those we observed as having an opportunity to see Online ads. This tells us how Magazine advertising performed on its own. (See the third image in figure 1.)

As we continued to survey consumers in the following week, Television ads hit the airwaves. To analyze how Television performed on its own, we can subtract those that read or looked through Magazines that we coded as containing Dove Nutrium Bar ads, and we can subtract those that we observed as having an opportunity to see Online advertisements. This leaves us with people that only had an opportunity to see Television advertising. We can then access how TV performed on its own in terms of moving the branding needle. (See the fourth image in figure 1.)

Each of these analyses is very powerful. They tell us how each media performs on its own. Furthermore, we can combine the cost data for each media to get a measure of relative efficiency in achieving branding goals.

Understanding how each media performs independently is insightful, but we were very interested in understanding how media complement one another and work in synergy.

Figure 2 walks through how we measure advertising synergy. In the second week of data collection, we examined those who were exposed to both Online and Print ads. Since TV advertising had not yet begun, we can look at the interaction of Magazine and Online without the influence of TV advertising. This analysis allowed us to better understand the synergy between the two media. (See figure 2.)

For example, the Online banners and the Magazine ads both contain the very clear pink and white Dove embossed beauty bar. However, while the Print ad has a clear “Nutrition Bar” headline along with “Dove Nutrium Bar,” the Online ads only have a small reference to “Nutrium Bar” in the very last frame.

Not surprisingly, we found that the Online ads increased awareness for Dove, but had little effect, on its own, in improving awareness specifically for Nutrium Bar.

By examining the advertising synergy, we found that the combination of Online and Magazine ads were far more successful in driving awareness of Dove Nutrium Bar than the Magazine ad alone. This suggests that the consumer was keying in on the pink and white bar. Seeing both the Magazine and Online ads really reinforced brand awareness substantially.

Synergy is an important part of our analysis. We looked at each combination of “Online and Print”, “Online and TV”, “Print and TV”, and all three together. (See figure 2.)

This analysis was very insightful. Not only did it underscore the generally held belief that all three media perform better than any one media on its own, it provided insight into how a properly balanced media mix can improve overall branding results.

Figure 2
DISSECTING ADVERTISING EFFECTS

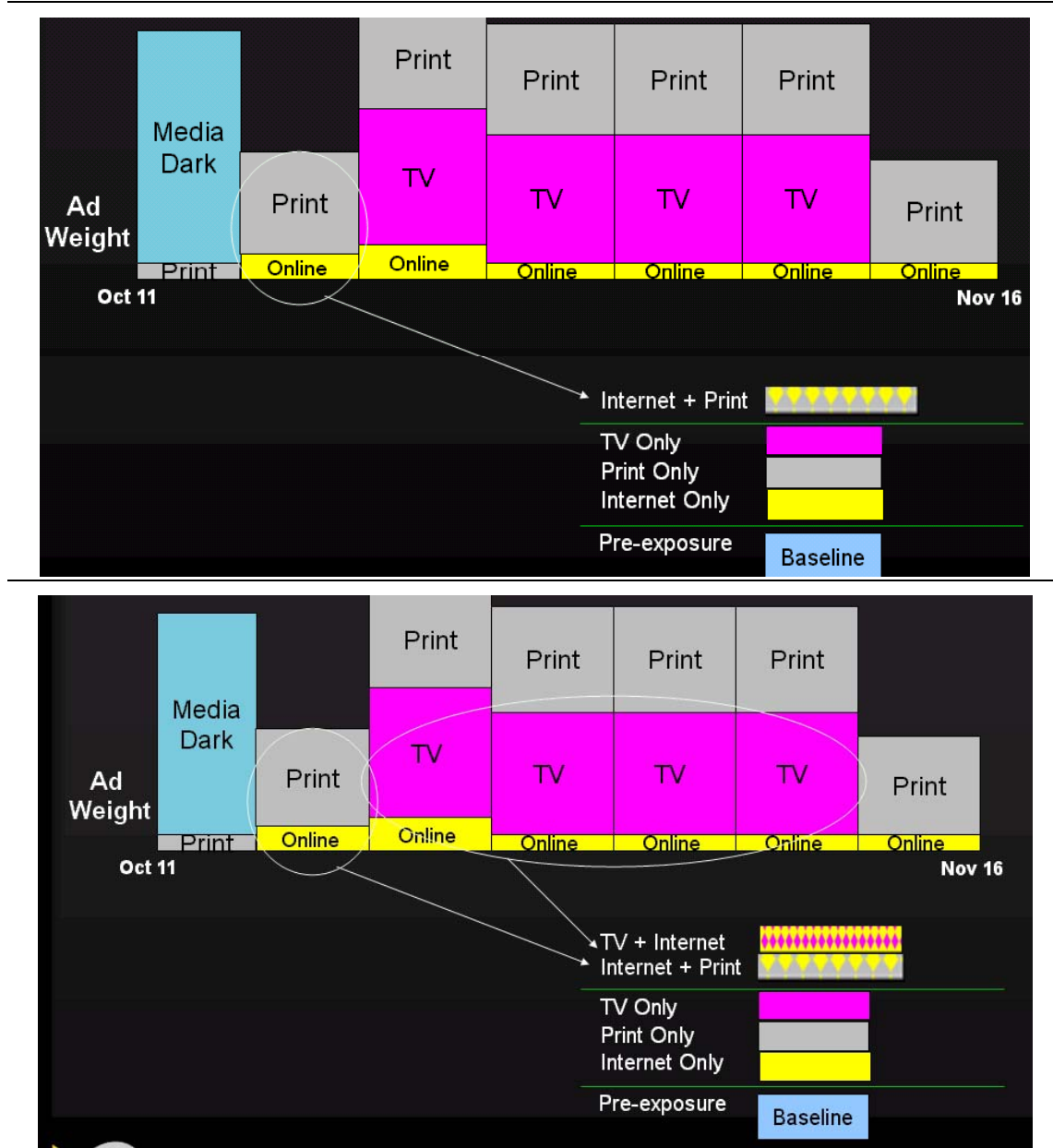
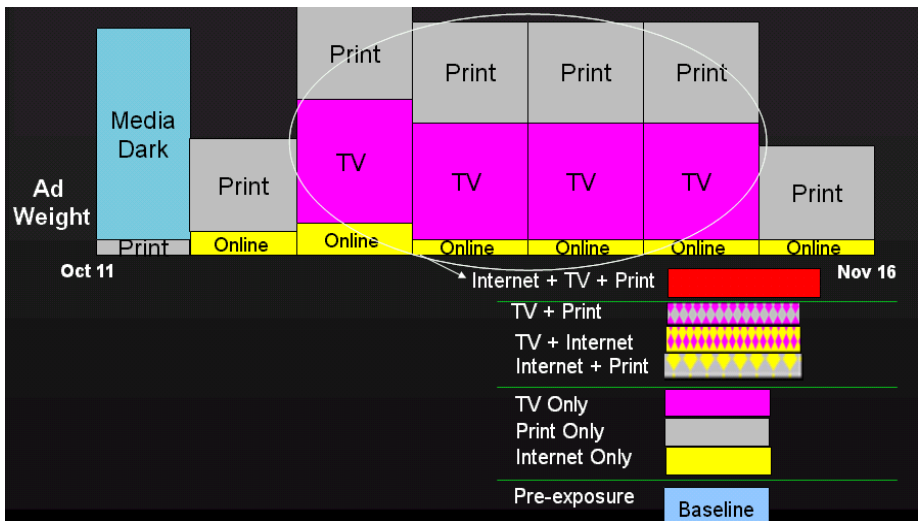
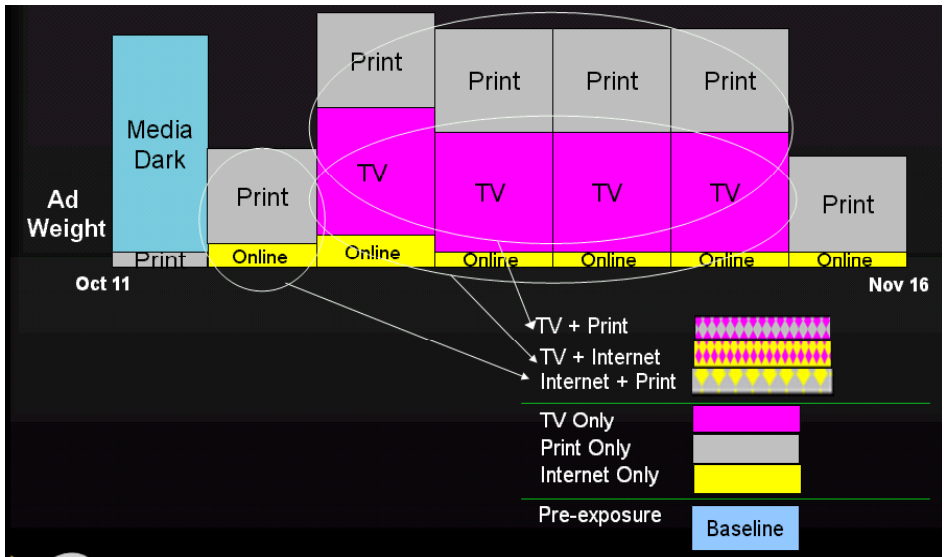


FIGURE 2, CONTINUED



Cost Efficiency Analysis

We have observed synergy between Offline and Online advertising. How do we know that it is “synergy” as opposed to simply the effect of added advertising weight? There are two ways to make this determination: One is cost efficiency analysis and the other is weight adjusted branding forecasts.

Cost efficiency analysis takes the cost of the advertising and divides it by the absolute branding increase from pre-campaign levels. This produces the cost of affecting each person on average.

For example, if Dove spent \$1,000,000 on the TV ad campaigns and the TV advertising increased purchase preferred by 1.5% among the target of women 25 to 49,¹²⁾ we could calculate the average cost per person affected.

We first project the total number of people affected by multiplying the total US population of women 25 - 49 years by the branding increase ($1.5\% * 52,674,822^{13}) = 790,122$). We then divide the total spend by the total number of people affected to arrive at the average cost for TV to affect each person, which is \$1.22 ($\$1,000,000 / 790,122 = \1.22).

We can calculate the incremental branding effect of adding Online, and calculate the incremental cost. By performing the same cost per branding effect calculation, we can determine if adding Online is more cost efficient or less cost efficient than using TV alone. If there is no synergy, then adding incremental media will add incremental branding increases at an increased cost (because there are diminishing returns). If there is synergy, the incremental cost per person affected will decline. In Dove's case, we found that adding Online was 32% more cost-efficient compared to TV and Magazine advertising alone – which is a powerful argument for using Online advertising.

The second approach to confirming synergy is weight adjusted branding forecasts. Since this analysis relies on measuring frequency, let us return to weight adjusted forecasts after we have thoroughly discussed the role of frequency.

Role of Frequency

We have talked about our objective, our definition and measurement of “branding,” our tracking of Online and Offline advertising exposure, and our analysis of the relative effects of Television, Print and Online. Now let us turn to our last area of analysis, the affects of frequency and diminishing returns.

The previous analysis, which examined the relative effectiveness and synergy of Offline and Online advertising, led us to conclude that Online is often more cost efficient in terms of improving overall branding results. The implication to marketers is this: increase Online's reach so that you can get these improved branding results among more people.

There is another dimension to optimizing branding results, and that is frequency. We wanted to understand the value of increasing (or decreasing) Television, Print and Online advertising frequency in terms of the core branding results. Here is how we approached this measurement.

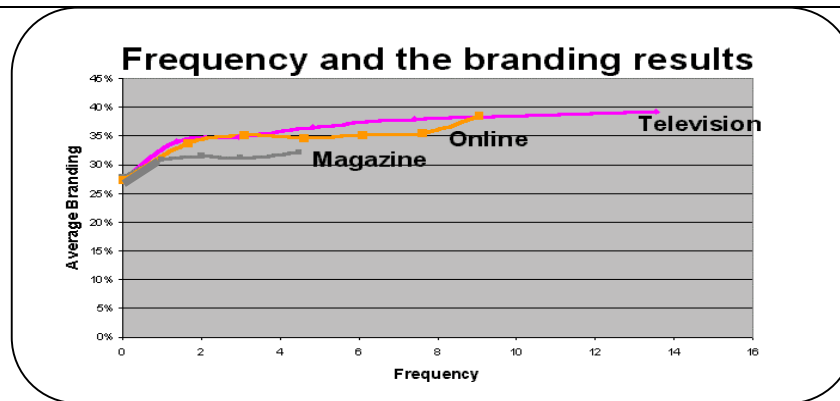
For Online, we looked at the number of opportunities to see Online advertisements and then broke people into one or more impressions, two or

more impressions, three or more, and so forth. We then looked at each branding metric against these frequency levels and plotted the curve.

What you can see, in this example from Dove Nutrium Bar, is how increasing Online frequency from the current 1.7 level boosts overall branding. However, increasing TV from the current level of 6.0 (over five weeks) produces very little incremental effect (see figure 3). The point is this: the methodology allows us to look at the value of increasing (or decreasing) frequency for Television, Print and Online.

Figure 3
FREQUENCY AND THE BRANDING RESULTS

**Average
Branding
Increase**



Number of OTS ad exposures

How do we determine frequency? For Online, we can use the observed count of ads served. This matches back to the currency with which Online is bought and sold.

For Magazines, we can use the self-reported data of the number of Magazines read, and match back to the ratings data. This match back allows us to ensure that we are aligning with the Magazine impression level corresponding to the frequency that the advertising agency purchased and paid for.

For Television, we use a quintile analysis of respondents. By quintile analysis, we simply mean we divide the population that has an opportunity to see Television advertisements into five groups, from heaviest exposure to lightest exposure for each day of the study. We do this based on cumulative daily quintile figures provided by the agency. We base this quintile analysis on the amount, daypart, and TV programming each respondent reports watching. We then match back to the ratings quintiles for the actual average frequency levels to our self-reported quintile. We cross-check the data from respondent level

with population figures to ensure the respondent assigned frequency is consistent with the population figures.

Frequency analysis is not only useful in terms of gauging the incremental branding from additional advertising impressions, we also plug-in the cost information.

For example, the first image in figure 4 shows that there is clearly value in increasing Online frequency, for example, from six impressions over six weeks to twelve impressions over six weeks in terms of improving brand image. But increasing frequency by this amount implies additional cost. The next part of our frequency analysis looks at the cost per branding increase at different frequency levels.

Cost per Branding Increase

Here is how the analysis works: Take the positive brand image frequency trend we looked at just a moment ago as an example. For Dove Nutrium Bar, the pre-campaign average positive brand image level is 30.7%.

At 6.1 impressions over six weeks, brand image jumps to 42.1%. That is an eleven point increase. In other words, if we talked to 1,000 people, we would find that 307 would agree to such statements as “Dove Nutrium Bar nourishes your skin” before we started this new ad campaign. After Online ads run at a 6.1 frequency level brand image increases. Now 421 people would agree with the same brand image statements. That means Online advertising made 114 (421-307=114) more people agree with the desired perception of the brand (see the second image in figure 4).

Figure 4
CALCULATING COST EFFICIENCY

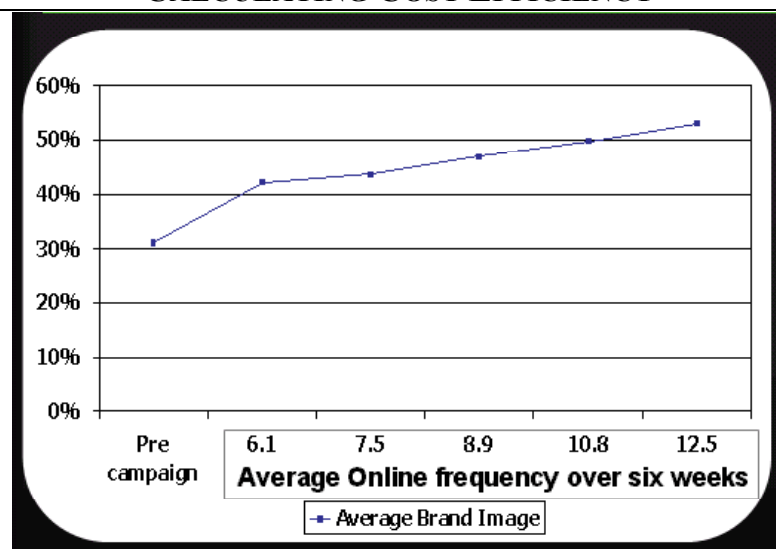
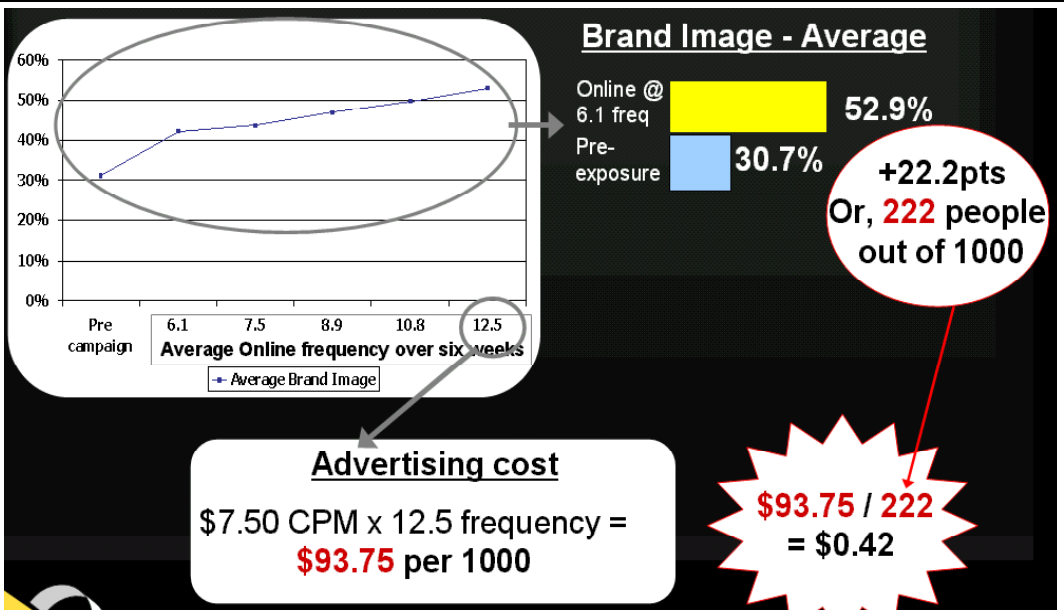
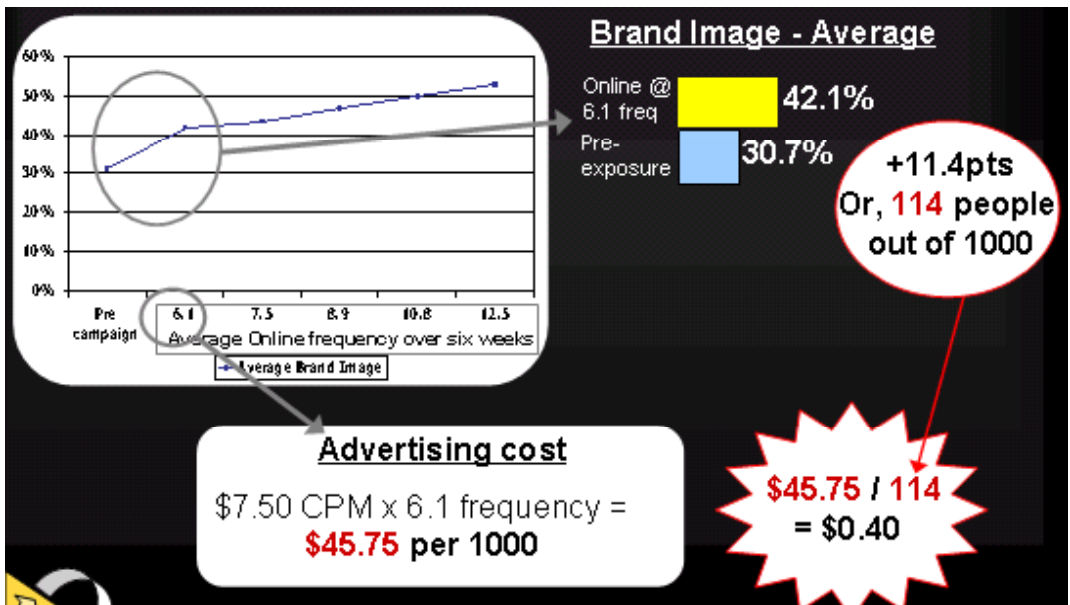
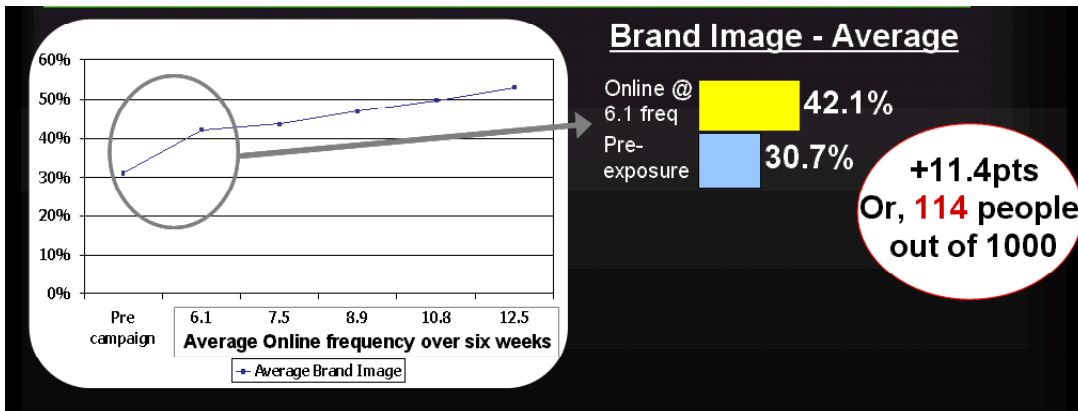


Figure 4, continued



How much did it cost to make these 114 people agree with the statements? Based on the CPM, it cost about \$45.75 per thousand at the frequency of 6.1 over six weeks. But let us translate this into a cost figure that means more to a marketer. How much does it cost, on average, to make someone agree with the key brand image statements? Taking the cost and dividing it by the effect (among 1,000 people), we find that it costs about forty cents to make each person agree with the brand image statement (see the third image in figure 4).

Now, let us look at a higher frequency level of 12.5 impressions over six weeks. We see that brand image increases from 42% at 6.1 impressions (over six weeks) to 53%.

That is a big increase. Clearly higher Online frequency produces better brand image results. Out of 1,000 people, Online advertising would now make 222 people agree with the branding statements. But increasing frequency costs more money. In fact, the cost per thousand at this effective frequency level is \$94.75. Is this extra frequency worth it?

Let us look at the cost per branding effect. What we find is that the cost, on average, for each new person agreeing with the branding statement is now forty-two cents (see the fourth image in figure 4). We found an increase over the forty cents at a 6.1 Online frequency.

We can plot the cost per branding effect at each frequency and determine the optimal frequency level. We call this chart the “cost per effect curve”. The point of this analysis is to get very precise about the cost and benefit of increasing frequency Online and Offline. Armed with the cost per effect curve, the marketer can now make a better decision about the appropriate frequency levels.

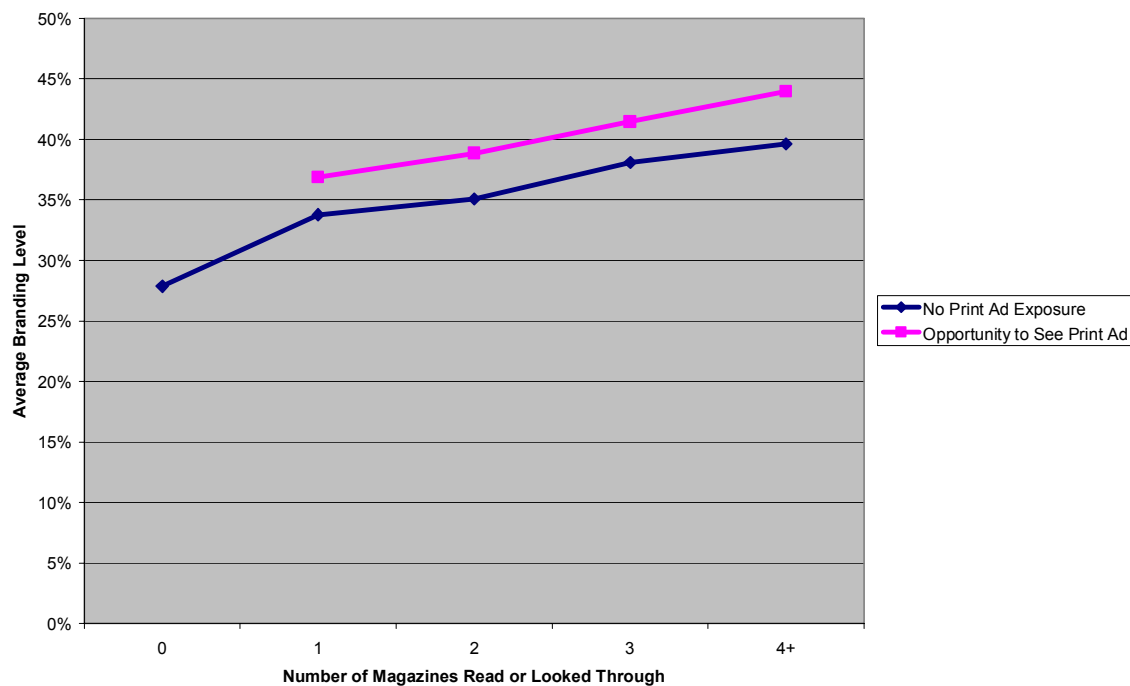
Depending on the most important branding objective to the marketer, different frequency levels may be selected. In this example we focused on brand image. Had we focused on brand awareness, we would have found that the diminishing returns set in earlier and a lower frequency over six weeks may have been more prudent. The point is not to suggest an optimal frequency for all brands to achieve all metrics. Rather, the point is that this type of research allows marketers to examine the role of frequency and diminishing returns in relationship to differing branding objectives.

We have talked about Online to illustrate this analysis, but we also performed the exact same analysis for Television and Print. We found that Television produces very little incremental benefit from increasing frequency beyond the current levels of 6.0 over four weeks. Moreover, because the cost is greater, there are steep diminishing returns.

The story for Print is similar. In some cases, increasing Print produces bigger effects, but the increase in costs tend to outpace the branding improvements. Also, in Print's case, those who indicated that they read or looked at more women's oriented Magazines tended to have a more favorable brand rating for Dove Nutrium Bar.

Figure 5 demonstrates that Magazine advertising certainly works – as those who had an opportunity to see the Magazine ads have more favorable branding levels than those that did not have an opportunity to see the Magazine ads. However, the figure also indicates that increasing Magazine advertising frequency does not produce that much of an incremental gain above those reading a similar number of Magazines (but not exposed to the Dove Nutrium Bar ads).

Figure 5
BRANDING IMPACT OF MAGAZINES (DOVE NUTRIUM BAR)



In other words, there is a “heavier reader bias” that should be subtracted to get a more accurate view of the role of frequency in this campaign. Heavier media user bias can be examined for each media.

To reiterate an earlier point: Our goal here is not to suggest that there is an optimal frequency for all brands revealed in this study. Rather our point is that a marketer can examine results at different frequencies for each media against

specific branding goals and use this learning to improve the overall performance of the marketing mix.

Weight Adjusted Branding Forecast

Now let us return to the question of synergy vs. added advertising weight. Are incremental branding gains caused by adding an additional advertising medium, or are they caused simply by additional advertising weight? Weight adjusted branding forecasts is a direct way of answering this question. While determining the value of adding advertising from another medium, such as Online, based on whether the cost per branding effect decreases is one sound and logical approach to addressing this question, weight adjusted branding forecasts is another answer to the question. The weight adjusted forecast approach does not require any assumption about diminishing return. The two approaches taken together provide converging validity.

In the case of the cost per branding effect analysis, we assume that putting more money (or weight) into TV will produce branding increases at the same cost per branding effect level (or at a less efficient level because of diminishing return). Therefore, if adding Online is more cost efficient, then TV, we conclude that the right decision is to increase Online advertising.

If we need to trim back TV slightly to free up money for Online, we examine the cost per branding frequency curve for TV to calculate the point at which TV's cost per branding effect is equal to Online's cost per branding effect. Provided that we are not cutting TV beyond this "cost-balance" point, we have produced a more optimal mix.

Weight adjusted forecast takes a different approach. We calculate the curve linear equation to describe the exact branding results by frequency. The equation is custom derived as a near perfect fit ($R^2 = .99\%$) of the observed data. The equation includes TV alone and TV and Online combined. This equation allows the analyst to input specific weight levels and receive as output the exact observed branding levels.

This means that we can determine if adding Online produces better overall results. Specifically, weight level is controlled by trimming down TV when Online is added. The overall branding results of TV alone and TV plus Online therefore represents a comparison of equivalent weight levels.

The conclusion from the weight adjusted forecast is the same as the cost per branding increase. Adding Online to TV produces better overall results.

SUMMARY

The Cross-media research study is a landmark analysis that gives us a chance to peer into the actual branding affects of Television, Print and Online across several brands and product categories. Marketers that invest the time and effort in this type of measurement can finally get a handle on where Online fits into the marketing mix. The indication from the studies conducted since the ground breaking Dove study demonstrates a substantial opportunity for improved marketing efficiency. We hope that this discussion of the methodology helps clarify the mechanics of our measurement and encourages others to replicate the cross-media measurement methodology.

ACKNOWLEDGEMENT

This research would not have happened without the intellectual contribution of Charles Newman of Unilever research; Jim Spaeth, President of the ARF; Piya Panyavetchawat, Mike Sigenthaler and Heide O'Connell of MSN, whose unconditional support allowed us to independently create a media neutral methodology; Greg Stuart of the IAB, whose experience as an ex-advertising executive inspired him to replicate the research so marketers could know rather than guess at their advertising ROI; and Dynamic Logic, whose flexibility in modifying their methods to accommodate our unique measurement needs was essential to our success.

FOOTNOTES/REFERENCES

1. Briggs. (1996). Hotwired Advertising Effectiveness study.
 2. Briggs and Hollis. (1997). Is There Response Before Clickthrough? *Journal of Advertising Research*.
 3. Briggs. (1997). Berkeley Systems You Don't Know Jack, the Netshow study.
 4. Briggs and McDonald. (1997). IAB Advertising Effectiveness Study
 5. Foley, Marianne. (1999). Ipsos-ASI report. ARF Conference,
 6. Russell. (2001). *Does Online Work? Yes, But...* Morgan Stanley Dean Witter Equity Research. February.
 7. Briggs, Stipp. (1999). How Internet Advertising Works. *Proceedings of the ESOMAR Conference on Net Effects*.
 8. The 45% is based on 75 interviews per week over three months and uses 10% lift from Online as the minimum level statistically significant at the 90% confidence level (using a one tailed test). A marketer percent reach by calculating the Online reach 45%, the measurable lift from Online (let's assume 10%) the starting branding level (let's assume 50%) and the total sample size over a period of time (let's assume 900 - 75 over 12 weeks). The calculation is as follows 50% baseline * 10% Online
-

lift * 45% reach = 2.25%. Statistical significance is (calculated as $1.28 * \text{Square root of } PQ/n$ – that is: $1.28 * \text{Square root of } 50\% * 1-50\% / 900 = 2.1\%$). A marketer can use this same equation to solve for other scenarios to ensure adequate sample size and reach levels.

9. Knowledge networks, Scarborough and UCLA
 10. IAB, PWC report
 11. The “dark” period is not essential to the measurement, but to the extent a brand goes off air it provides a unique opportunity to see how introducing (or reintroducing) ads in different media combinations affects consumer attitudes.
 12. Cost and branding levels are not the actual Dove figures, these figures are used to illustrate the calculation only.
 13. Census QT-P1 report 2002
-

TECHNICAL APPENDIX 1

MATHEMATICAL TESTING OF SAMPLE FRAME BIAS

Mathematical testing of sample frame bias without conducting an expensive parallel telephone tracking study

We can examine the potential impact of sampling only those with Online access by mathematically testing different assumptions of branding response for the 30% without Online access. For example, for this analysis, we can assume that those that do not access the Internet are 50% more likely to want to buy a product that they see advertised on TV and in Magazines than those that use the Internet (and also watch TV and read Magazines, but are not exposed to any Online ads for the brand). What impact would such a difference have on our overall results?

The answer is very little. Since the population that we have left out is only 30% of the overall results, Television and Magazine advertising could increase purchase intent four times higher for those that do not use the Internet compared to those that do use the Internet and it still would not change our overall finding that Online adds value to the overall marketing mix. In other words, the potential for sample frame bias is rather insignificant and not worthy of concern. Let us take a closer look at the findings and the mathematic potential for bias to better understand why we can dismiss sample frame bias as a concern.

The two key findings of the Dove Nutrium Bar study are that the addition of Online advertising increases purchase intent from 11.5 without Online to 14.2 among those reached with Online ads at an average frequency of 3.1. We know these findings are true for the 70% that have Online access since they are the sample frame from which we are projecting results.

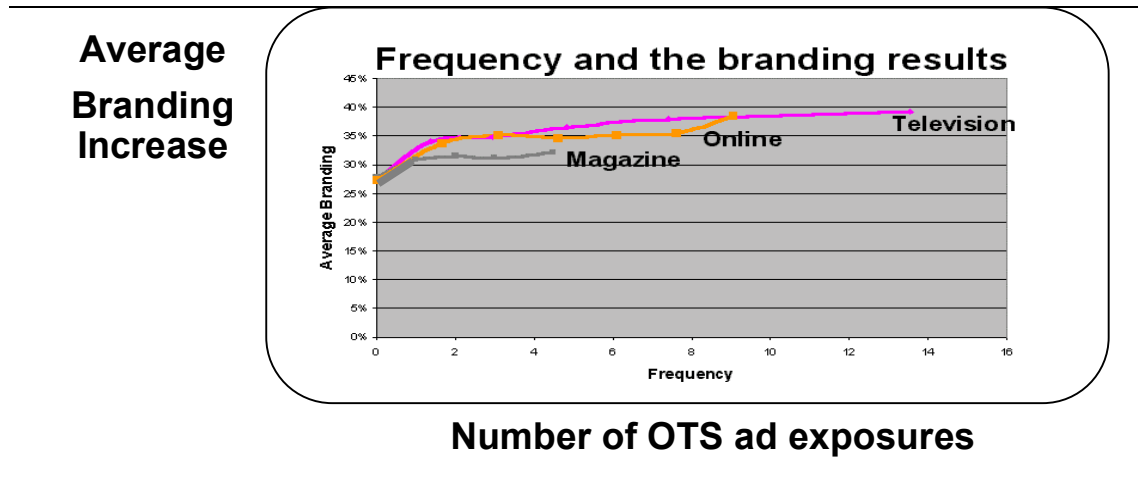
We also know that the 30% that do not have Online access will receive no benefit from the Online advertising (as they have no access to be exposed to Online ads) and therefore we need to focus on the impact of the slight trimming of TV frequency from 6.0 to 5.5 and Magazine from 2.7 to 2.0.

To reiterate, this finding is specific to those with Online access, therefore these findings will not change for 70% of the target population regardless of the differences among those without Online access. (See figure 6.)

The recommendation to redistribute budget does warrant additional analysis to determine whether differences in advertising frequency curves among those without Online access might change the conclusion.

In the Dove Nutrium Bar study, we examined the branding effect by frequency and found that Magazine and TV advertising could be trimmed back slightly allowing Online to be increased within the same budget. We concluded that this would increase purchase intent overall by 14% without Dove spending more on advertising. This conclusion is based on the assumption that the 70% sampled are representative of the total target audience.

Figure 6
BRANDING EFFECT PROJECTED FROM 70%



But what if the 30% that do not have Online access differ dramatically from those with Online access in terms of how advertising frequency in TV and Magazines affects their purchase intent? Rather than ask “what if”, we can test the different scenerios mathematically to determine the possible outcomes.

First, a key point: When we stated that Online advertising could be increased by trimming TV and Magazine slightly, we did mean slightly – TV was trimmed from 6.0 over 4 weeks to 5.5 and Magazine was trimmed from 2.7 to 2.0.

So, the practical question of sample bias is really confined to whether the effects of TV advertising are drastically different between consumers with Online access and those without it over the range of 5.5 and 6.0 frequency for TV and 2.0 and 2.7 frequency for Magazines.

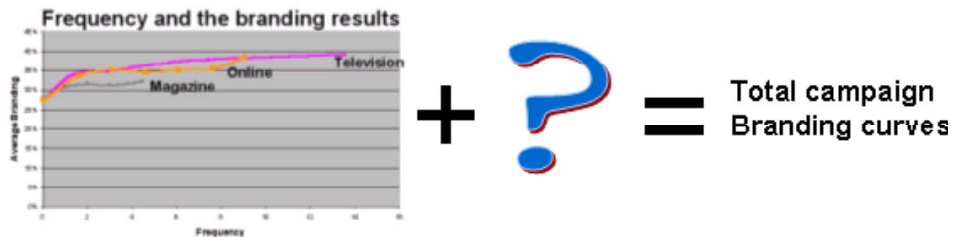
In examining the branding increase from 5.5 to 6.0, we find less than 0.25 gain for TV, indicating that TV weight has hit the point of diminishing returns. Same for Magazines. The gain from 2.0 to 2.7 is zero. We know that the delta between 5.5 and 6.0 for TV and 2.0 and 2.7 for Magazines is true for the 70% with Online access, but what if the 30% without Online access exhibit greater branding gains from 2.0 to 2.7 and 5.5 frequency to 6.0?

How big of a gain would we need to observe among the 30% without Online access to alter our recommendation to increase Online advertisings’ share of the budget? The answer is rather than the observed average gain of 0.10 we would need to see a gain of over nine points, which is 90 times greater than the gain we observed among those with Online access.

From a practical standpoint, this suggests that it is next to impossible that the Offline users are different to the degree that we will change our recommendation.

The mathematical mechanics tells us that a 90 times difference is required to change our conclusion – but is such a difference possible given what marketers know about TV and Magazine frequency curves?

Figure 7

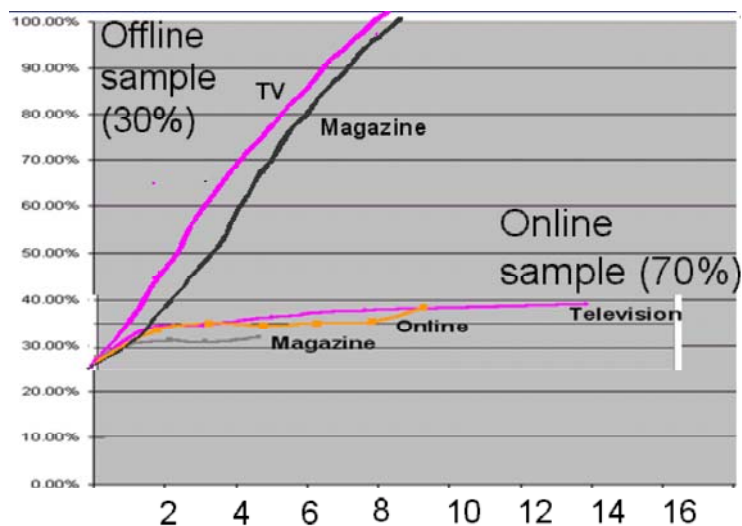


Online sample (70%) + Offline sample (30%) = Total effect

We can return to the frequency curves to answer the plausibility question. We know what the frequency curve looks like for the 70% with Online access.

These curves fit the classic Convex-linear diminishing return curve most media researchers have come to expect. Our question is what this curve looks like for those without Online access. If we examine the frequency for 2.0 to 2.7 for Magazine and 5.5 to 6.0 for TV, and try to fit a 9.0 branding gain over this small range, then either we must assume that diminishing returns have not set in, or the other alternative is that the curves between Offline and Online samples look similar up to 5.5 for TV and 2.0 for Magazine. And then, for some reason there is a major inflection point.

Figure 8

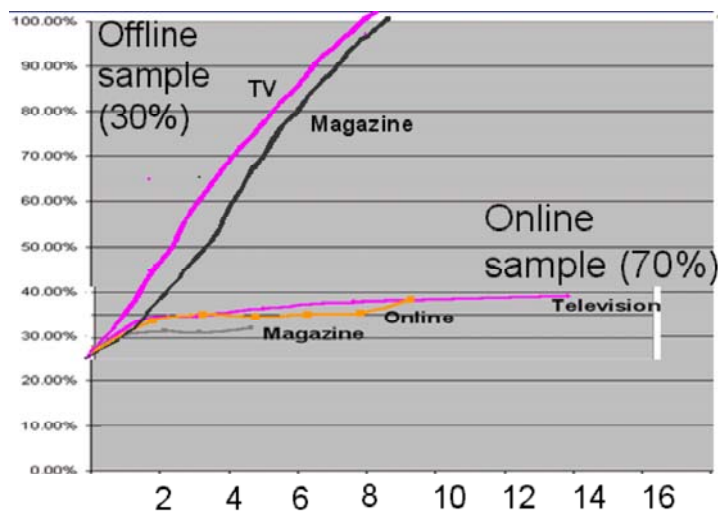


The first hypothesis of a 9.0 gain over 0.5 TV frequency and 0.7 Magazine frequency would suggest that each increase in frequency is providing about 20 points in branding gain for TV and Magazine. If the curves are linear that would mean that by a frequency of 5 for TV, the branding levels would be over 100%!

Since this is impossible, we can reject the idea that among Offline users, branding levels consistently increase at the level required to produce a 9.0 gain over the frequency of 5.5 and 6.0 for TV and 2.0 and 2.7 for Magazine, (the level at which we would reject our conclusion based on the Online sample).

The second mathematic way that Offline users could theoretically be different enough in their response to Television and Magazine advertising is for their general branding levels to be similar to the 70% of the target with Online access, except at the precise frequency for which we wish to trim back TV and Magazine. That is to say that there is a significant inflection point at 5.5 to 6.0 for TV and 2.0 and 2.7 for magazines.

Figure 9



While these frequency curves are mathematically possible, we have never observed such inflection points in the middle of a frequency distribution and find it unlikely that we would find such a pattern only among those without Online access (and no hint of inflection among those with Online access).

We can therefore reject this second hypothesis as well. The most plausible assumption is that Online users and Offline users are not 90 times different over the frequency range of 5.5 to 6.0 for TV and 2.0 to 2.7 for Magazines. We therefore can be confident in our recommendation to redistribute budget by trimming TV and Magazine frequency slightly.

TECHNICAL APPENDIX 1

SAMPLING PROCEDURES AND ALTERNATIVES

Sampling consisted of creating three distinct sub-samples. First is the broad sample drawn from a large selection of websites that reached the key target audience. Second is a sample of those exposed to Online advertisements, and third is a sample drawn from the exact same sites as the sample of exposed respondents, but with controlled exposure so that they never see an Online advertisement.

All samples were recruited continuously over the course of the campaign with the same pop-up invitations executed by Dynamic Logic. Consumers are surveyed only once. Branding levels are tracked by detecting shifts in the population. In this way, the study is a classic continuous sample as opposed to a longitudinal study.

The first sample is the representative sample. It is balanced back to U.S. Census targets. The second and third samples are matched to ensure no demographic variation. The second and third samples are also given a proportion weight (to adjust for the over-sample).

Sampling of those exposed to Online advertisements can occur anywhere from immediately after exposure to up to several weeks after ad exposure. The same is true for exposure to other media as well. We address the concern of surveying after immediate exposure by conducting a decay analysis. This analysis examines the time delta between exposure and the survey and examines how the branding levels decline over time. We perform a similar decay analysis for TV (once TV goes off-air).

There are several alternative approaches to the sampling. First, we could use a panel and direct a continuous stream of consumers to take the survey Online. We could read their cookie and determine if they were in the exposed or control cells. This type of an approach is practical if the Online campaign reach is significant enough to produce a readable base size among those naturally exposed. For example, if the campaign has a 10% reach, it will take 1,000 interviews to produce 100 consumers with exposure.

The panel approach can be modified if the panel is linked with an ad server which can give preferential delivery to those in the panel. In such a situation, natural reach can be “over-delivered” to those within the panel. This can occur passively, thus preserving the integrity of the “real-world” study design.

Another alternative is to deliver advertisements to “exposed” and “control” cells and monitor some post-advertising behavior, such as purchase. Those who have taken the action (such as purchase) can then be randomly/ representatively sampled and directed to an Online survey. The past media behaviors of these purchasers can then be examined to determine the effect of various media combinations on the desired action.

TECHNICAL APPENDIX 3

DATA WEIGHTING PROCEDURES

Data weighting consisted of balancing the sample to know U.S. Census targets on age and sex. Cell weighting was generally used as it is more precise than RIM weighting.

However, in some cases, we used additional purchase pattern and other demographics to weight the data. For example, for Universal Home Video, the client wanted to ensure proper matching of light, moderate and heavier DVD purchasers by survey week (for trending the data) and among exposed and control cells (to ensure comparisons are not biased by any anomalies in cell matching). Further, it was desirable to ensure proper balance of those with kids in the household among the analytic groups. We examined the proportion in the total population and applied this same proportion to each of the cells thus ensuring cell balance on the key profile variables.

Weighting ratios of the largest weight divided by the smallest weight were generally less than 4, meaning that even in the most extreme cases, one respondent's survey input would carry four times the weight of another's.
