



Better Representing Magazine Effects in Marketing Mix Modeling

Wayne Eadie (MPA), Mark Reggimenti (Brand Science), Judy Vogel (PHD), Dr. James Collins (MRI)

Winner, Chairman's Award at the 2009 Worldwide Readership Symposium

Introduction

We are in unprecedented times in our industry. Consumer control over content creation, distribution and consumption is expanding at a rapid rate. New media choices emerge on an almost daily basis. Advances in technology have helped move the marketer/consumer relationship to very much more of a two-way interaction. All of which has made the job of connecting with and motivating consumers more challenging. Yet at the same time, we media practitioners are required to demonstrate greater proof of performance, answering questions ranging from “Did it work?” and “What worked?” to “How can we make it work harder?”. Coupled with the growth in the influence of client side finance, procurement and international budget setting/control disciplines, the need to provide robust evidence substantiating marketing investment plans and executions is significant.

Part of the solution to addressing these critical business questions has been marketing mix modeling (MMM). A member of a large family of Econometric modeling techniques, MMM is generally described as the use of multivariate and time-series statistical analyses to estimate the impact of various marketing elements on sales or other marketing objectives. The models are often used to forecast the impact of future marketing activities and often guide fundamental media investment decisions. Hence, MMM has found wide acceptance as a trustworthy marketing tool among the major consumer marketing companies. As with all analysis and modeling, critical to an accurate MMM is the data used in its development and deployment.

Hypothesis

Hence, the principal hypothesis of this paper is the obvious, but often compromised proposition that better media audience data leads to better MMM results, better predictive models and hence more efficacious marketing and media strategies. Magazines are one medium where the complexity of audience measurement has limited its impact and statistical relevance in MMM practice. In comparison, digital and television audience measurement of advertising is relatively precise and immediate as the exposure is served and then disappears, rarely (not withstanding delayed and repeat television viewing) to be seen again. The ratings and audience of these advertising exposures are measured discretely for each separate advertisement or at least advertising minute.

Magazines are a more persistent medium, exposing people to their advertisements as people read the publications and this readership occurs over the course of several weeks or even months. Magazine audience has also been traditionally measured on a bi-annual basis where all issues during that period possess the same audience (i.e. average-issue) rating. This could be as few as three issues for bi-monthlies to as many as 26 issues for weeklies. We would never apply the same rating to a half year of Yahoo! Front pages or Primetime television shows, so why is this acceptable for magazines?

The goal of any genuine “media neutral” analysis should be to place all media on as level a playing field as possible by employing the best, most representative data available when assessing marketing outcomes and ROI.

Approach

In practice MMMs are built to assess the relationship between brands’ marketing efforts and a variety of possible marketing objectives. As marketing objectives YouGov’s Brand Index health/sentiment metrics have proven to be both meaningful and versatile in a variety of Brand Science’s projects across a panoply of business sectors. In particular, Brand Index metrics have proven valuable as both key performance indicators (KPI) in upper funnel analytics and as business drivers in Sales/ROI modeling. Thus, while any particular MMM exercise may employ its own ROI measure(s), Brand Index metrics nicely represent ROI metrics for typical MMM work.



Thus, the goal of this paper is to look beyond the merits of particular statistical techniques and client-unique ROI measures, by focusing on the more fundamental matter of what is the most robust data to employ for print in MMM's. Even before an argument can be made for the use of one or another modeling technique or ROI outcome, there needs to be some consensus among analysts and users of MMM results as to what is the best, most representative (print) media data.

The dramatic enhancements in the measurement of print advertising audiences have been largely neglected in MMM practice – much to the detriment of MMM practice and print medium and ultimately advertising campaign planning and spending efficacy. This neglect has a variety of possible sources including:

- 1) Focusing unbalanced attention on the particular statistical techniques employed in individual MMM executions
- 2) Inadequate access to the media planning/processing tools and data sets necessary to retrieve and process these more precise data
- 3) Commitment to productizing particular analysis techniques to efficiently generate MMM results
- 4) Lack of comprehensive knowledge of available, relevant (print) media and advertising data

Thus, the primary foci of the work at hand are 1) the review and analysis of the variety of print metrics available as data for MMM's and 2) the identification of the most suitable and robust of these various metrics for effective MMM's. The review and analysis are based on six brands from across a relatively extensive array of consumer categories:

- Household products
- Batteries
- Hotels
- DTC Pharmaceuticals
- Food
- Financial Services

Each brand's analysis has been performed on three Brand Index metrics derived from YouGov:

- Positive Buzz - Respondents are prompted for anything they've heard in the media (news or advertising) or in conversations among friends and family. This is to elicit whether people have noticed good or bad news, advertising or PR campaigns, product launches, or whether there is any "word on the street" emerging.
- Positive Impression - Respondents are asked for which brands in a sector they have a "generally positive" or "generally negative" attitude.
- Willingness to Recommend - Respondents are asked which brands they would recommend to a friend? Which would they suggest best be avoided? Changes in this metric (potentially) suggest future change in market share.

These three Brand Index measures are employed in this work as they 1) have been found to well represent current and future market outcomes in MMM's and 2) are among the measures about which many advertisers most care.

Each brand's analysis for each Brand Index metric considers the impact of magazines using five different audience and surrogate audience metrics:

- Nielsen Ad*views rate card spend applied to the issue date*
- MRI Average Issue Audience GRPs applied to the issue date
- MRI Issue Specific Audience GRPs applied to the issue date
- MRI Average Issue Audience GRPs audience accumulated
- MRI Issue Specific Audience GRPs audience accumulated

GRP data for the six consumer brands and their associated target demographics are outlined below:



Consumer Category	Target Demographic
Financial Services	Males 35-64
FMCG (3)	Women 25-54
Travel	Adults 25-54
DTC Pharmaceuticals	Adults 35+

Magazines included in the campaigns but unmeasured by Mediamark Research & Intelligence (MRI) were prototyped following the guidelines established by the Advertising Research Foundation (ARF) (See Appendix B for guideline details).

*ad spend from Nielsen, versus actual, was used as is typical of marketing mix models where competitive spending (where no actual data is available) is factored in.

Data Treatment

In addition to treating unmeasured magazines through prototyping, a few additional, relatively conventional data treatments were applied.

The Brand Index KPIs are analyzed on a four week moving average the better to control for random sampling error common in health and consumer brand sentiment tracking data.

Moving Average

$$\widehat{BIX}_{it} = \frac{1}{\sum n} \sum_{\tau=0}^3 (BIX_{i,t-\tau} * n) \quad \forall t \in T$$

where

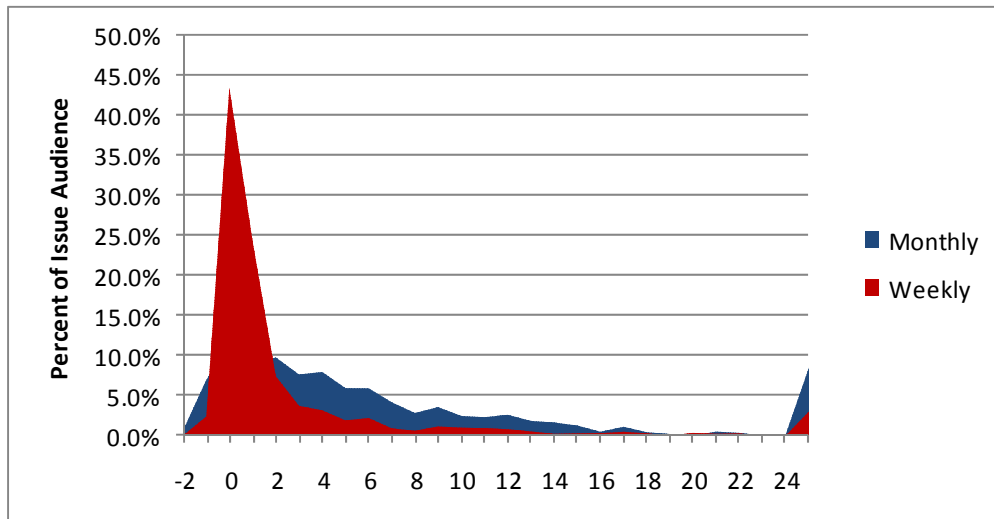
- \widehat{BIX}_{it} : The smoothed i^{th} BrandIndex Key Performance Indicator at time t .
- $BIX_{i,t-\tau}$: The raw i^{th} BrandIndex Key Performance Indicator at time $(t - \tau)$, i.e., period $t, t-1, t-2$, and $t-3$.
- n : Weekly sample size
- t : Current week
- T : Full modeling period

MRI Issue Specific Ratings

For the most precise and discrete rating of each magazine's audience MRI's Issue Specific audiences have been applied to each advertising insertion. Title issue specific ratings are available for 131 of the 197 (66.5%) titles utilized in this study and these titles account for 15,066 of the 16,349 (92.2%) GRP's. For those titles without issue specific ratings, we applied the average issue ratings.

MRI Audience Accumulation

For the correct timing of each magazine’s audience delivery MRI audience accumulation curves have been applied to each insertion. Title specific curves are available for 148 of the 197 (75%) titles utilized in this study and these titles account for 15,628 of the 16,349 (95.6%) GRPs. For those titles without audience accumulation curves, we applied the average curve for that publication’s publication frequency.



Model Criteria

The basis on which the validity of the magazine audience data will be judged is the Studentized t -stat of that magazine variable in each model. A greater t -stat signifies a tighter relationship between that variable and the KPI.

Studentized t -stat

$$t_i = \frac{\hat{\beta}_i}{\frac{se}{\sqrt{n}}}$$

where

t_i	:	Studentized t statistics for the i^{th} independent variable.
$\hat{\beta}_i$:	The i^{th} estimated coefficient from a regression
se	:	Standard error
n	:	Number of observations



Findings

The results of the study have proven to be quite successful in demonstrating the value of using better, more granular data as inputs for magazines in MMM's. In general, moving from spend to ratings to audience accumulated ratings, has produced higher *t*-stat values, as the tables below illustrate.

Positive Buzz

Category	Spend	Average Issue GRP	Issue Specific GRP	Average Issue GRP Adcume	Issue Specific GRP Adcume
Household Products	1.55	1.82	2.23	2.45	3.02
Batteries	2.53	1.59	3.06	4.05	4.12
Hotels	0.55	0.41	0.42	0.50	0.52
DTC Pharmaceuticals	2.40	2.61	2.58	2.89	2.75
Food	1.02	1.04	1.09	1.59	1.64
Financial Services	2.60	2.69	2.74	3.23	3.28

Positive Impression

Category	Spend	Average Issue GRP	Issue Specific GRP	Average Issue GRP Adcume	Issue Specific GRP Adcume
Household Products	0.43	0.67	0.78	0.61	0.86
Batteries	2.27	2.29	2.28	2.35	2.32
Hotels	1.34	1.59	1.41	0.82	0.71
DTC Pharmaceuticals	1.98	2.01	2.03	2.61	2.65
Food	2.29	2.54	2.55	4.89	4.89
Financial Services	1.83	1.87	2.05	2.18	2.43

Willingness to Recommend

Category	Spend	Average Issue GRP	Issue Specific GRP	Average Issue GRP Adcume	Issue Specific GRP Adcume



Household Products	0.73	0.75	0.78	1.07	1.10
Batteries	1.77	1.92	1.90	1.79	1.76
Hotels	0.29	0.41	0.38	0.49	0.50
DTC Pharmaceuticals	2.10	2.07	2.01	3.14	3.13
Food	1.75	1.87	1.96	1.97	2.00
Financial Services	2.23	2.41	2.45	2.89	2.95

For virtually all of the categories examined and effectiveness measures employed, we found consistent improvement when moving from ad spend through the increasingly precise GRP's. (In the case of Hotels wherein this was not the case, it may well be that the economic conditions prevailing during the campaign period and their impact on both business and personal travel may be responsible for these somewhat anomalous results.)

Beyond this general conclusion two more discrete findings emerge:

- 1) Generally the most significant improvement in model performance is made with the incorporation of Adcume.
- 2) Issue Specific metrics do not seemingly demonstrate significant effect.

Both of these findings are to be expected.

The relatively modest incremental performance of the issue specific metrics beyond the average print vehicle audience ones, is largely due to the averaging effects among vehicles in these campaigns. While individual print vehicle issues may generate varying levels of GRP delivery, over the course of the campaign they will tend to average out to something approximating the vehicle's average-issue audience. Moreover, while at any single point in time throughout the campaign different issues of different vehicles may be generating differing GRPs, at any single point in time averaging among these vehicles' issues will also arise. Thus, the minimal differences in MMM performance between the average issue and issue specific metrics.

With that said, while the value of the issue-specific metrics is modest for MMM purposes, it is still quite valuable for ROI generally. Print, as currently with television and the internet, inevitably will be held to account for audience delivery/performance and if a particular magazine fails to deliver consistently among a number of selected issues in the campaign its contribution to the campaign is diminished irrespective of how the campaign in aggregate performs.

That adcume generally demonstrates the greatest improvement in MMM performance is again expected. The practice of assigning print's audience (or expenditure) in aggregate to discrete single points in time clearly and fundamentally misconstrues the way in which print delivers impressions over time and thus is misaligned in relationship to the MMM's outcome ROI measure. By way of reference, even weekly magazines deliver only about 50% of their audience in the week of publication, with monthlies exhibiting more extended delivery dynamics. Hence, the inclusion of adcume is all but mandatory in responsible MMM practice involving magazines.

Conclusions and Future Work

While expected, these findings are significant for MMM practice. As the MMM outcome is 1) diachronic and b) sensitive to changes in advertising penetration and extent, the use of print metrics reflecting the actual delivery of advertising exposure over time is paramount. Magazine metrics falling short of this threshold cannot be expected to perform adequately in diachronic MMM modeling, quite simply because they misrepresent in a fundamental manner how print actually delivers audience/impressions.

With these findings in hand several other matters await consideration:



- 1) Repeat exposure to the same vehicle issue
- 2) Reach / Average Frequency

As was noted in the introduction, print is not an ephemeral medium; readers may return again and again to a single issue of a magazine. Some metric, average page exposure, for example, might well show significant effect in MMM as repeat exposures to an advertisement in the same issue of the same magazine almost inevitably will some deliver incremental benefit. Similarly, reach and/or some measure related to average frequency may prove significant as the number of people being reached and their frequency of exposure at points in the campaign could reasonably be posited to have some impact on the extent of the response.



APPENDIX A



Nielsen MonitorPlus

Data collection practices for the various media explored in this study are detailed in the following:

Network Television

Costs: Program/daypart costs are provided by individual Nielsen-monitored networks where available. Where broadcast data are unavailable, expenditure information is supplemented by other industry sources and may include data derived from proprietary models. Expenditures are weighted based on their relationship to a 30 second spot.

GRPs: Nielsen Television Index (NTI) provides ratings data from the national People Meter panel on a weekly basis. Program averages are reported on an individual program basis. The program rating is applied to each commercial within the program regardless of duration, pod or position. The minimum reportable rating (before rounding) is .05. Available data streams include Live, Live+SD, Live+7 and C3.

Syndicated Television

Costs: Expenditure data reported for commercial activity are derived from data supplied by SQAD and from proprietary models. Expenditures are weighted based on their relationship to a 30 second spot.

GRPs: Nielsen Syndication Service (NSS) provides program ratings data estimated from the national People Meter panel on weekly basis. Program averages are reported on an individual program basis and are not adjusted according to duration. The minimum reportable rating (before rounding) is .05. Available data streams include Live, Live+SD, Live+7 and C3.

Spanish Language Network Television

Costs: Expenditure data is derived from program costs provided by individual monitored networks. If a network does not provide broadcast data, expenditure information is supplemented by other industry sources. Expenditures are weighted based on their relationship to a 30 second spot.

GRPs: Nielsen Hispanic Television Index (NHTI) provides program ratings data estimated from the national Hispanic panel on a weekly basis. Program averages are reported on an individual program basis. The minimum reportable rating (before rounding) is .05. Data streams Live, Live+SD, Live+7 and C3.

Spanish Language Cable Television

Costs: Expenditure data is derived from program costs provided by monitored networks. Where broadcast data are unavailable, expenditure information is supplemented by other industry sources and may include data derived from proprietary models.

GRPs: Nielsen Hispanic Homevideo index provides program ratings data on a weekly basis. The minimum reportable rating (before rounding) is .05. Data streams include Live, Live+SD, Live+7 and C3.

Spot Television

Costs: Expenditure data is derived from household CPP information supplied by SQAD. CPP information is based on actual spot television buys placed by advertising agencies and media buying services

GRPs: Commercial occurrences are linked to the Nielsen Station Index (NSI) Viewers in Profile (VIP) ratings by report book month, station, day of week and quarter hour. The minimum reportable rating (before rounding) is .05. Data streams include Live, Live+SD, Live+7 and C3.

Cable Television

Costs: Expenditure data is derived from SQAD's NetCosts product. NetCosts' cost per point data is based on actual national television buys as reported by participating advertising agencies and media buying companies.

GRPs: Nielsen Homevideo Index (NHI) provides both program and half hour ratings data estimated from the national People Meter panel on a weekly basis. In order of preference, Nielsen Monitor-Plus utilizes program ratings based on: Monday-Sunday averages across all dayparts, Monday-Friday averages across all dayparts, individual telecast and half-hour averages. Half-hour



ratings are used for non-program oriented networks. Average audience estimates are based on the total national universe without regard to cable coverage. The program rating that is applied to each commercial is not adjusted according to duration. The minimum reportable rating (before rounding) is .05. Live, Live+SD, Live+7 and C3 data streams are available.

National Magazine

Costs: Expenditure data is computed through a calculation procedure that utilizes the published advertising rates for each magazine, provided by SRDS. Costs reflect discounts earned by corporations (parent companies) by accumulating the advertising activity of all brands within the corporate structure.

Audience Data: MRI provides magazine audience data. Monitor-Plus occurrence data is reported with the most current MRI adult audience study. Impressions, GRPs, CPPs and CPMs beginning with January 1999 are available.



APPENDIX B

ARF Magazine Prototyping Guidelines

Background

There are a multitude of magazines serving the various niche audiences vital to the promotion of U.S. consumer brands. The nature of these publications' audiences is such that hundreds of them cannot be rigorously measured by syndicated services. As a result media agencies and publishing companies produce hundreds of their own prototypes. After an ARF audit of these practices, the ARF Magazine Prototype Committee was formed to frame general principles and guidelines for best practices for use by agencies and publishers.

Definition

The Advertising Research Foundation (ARF) defines prototyping as "estimating a magazine's audience based on a subscriber study and using model (judgmental and/or computerized approach used individually or together) to select surrogate 'measured' publications."

Prototyping provides a basis for creating an estimate of an "unmeasured" magazine audience for subsequent computer-assisted media analysis within the context of an appropriate syndicated study.

Unmeasured Publication Candidacy

In order to develop a prototype for any publication, that publication must meet the following requirements:

- Publications must have published at least four issues
- A minimum circulation of 100,000 is recommended
- Geographically skewed publication prototypes should be clearly disclosed as such
- The magazine's demographic profile must be consistent with the syndicated service used for the host selection

Circulation Audit

The following circulation data should be included in any prototype:

- An ABC or BPA statement or a letter of confirmation from ABC or BPA of a formal filing for audit as well as a sworn publisher's statement of circulation
- Detailed circulation plans for either rate base increases or reductions, publication frequency changes and any changes in the mode of distribution as "on the record"
- The effect of ABC reporting changes and their potential effect on any prototype estimate should be made explicit
- Complete explanation of the home subscription development plan, newsstand performance or future plans for newsstand sales, public place programs, including all forms of "non-traditional manners" like daycare centers, publications as part of membership or loyalty programs, etc. (controlled vs. paid)

Audience Source

Prototypes are typically developed for use within syndicated audience surveys. Publishers, agencies and the research syndicators, notably the latter, may establish their own guidelines.

Bases for Prototype

Subscriber Study Requirements - Subscriber studies on which prototypes should be ABC audited, or closely follow ABC guidelines.

Newsstand Circulation Proportions - High quality newsstand surveys requiring random probability sampling of newsstand locations are required to provide a reasonable estimate of the profile of the newsstand reader.

Frequency - Prototypes should be considered for re-evaluation based on new subscriber or newsstand studies at least every two and a half years, or sooner depending on the evolution of the editorial and the circulation.

Tie Profile to Syndicated Sources - Each prototype requires a unique analysis of the potential host publications in the syndicated service (based on the same subscriber study).



Frequency of Prototype Execution/Currency Issue

New prototypes should be prepared from scratch whenever a new, preferably audited, subscriber or newsstand study becomes available.

Matching/Host Selection

Analysis of all available *evaluated* data is recommended. For example:

- High-quality subscriber study, preferably with an ABC verification audit as appropriate
- Data from other reputable third parties (MMR, JD Power, IntelliQuest, if a publication is measured in one of those services)
- Review of all data and media kits from the publication and its competitors, as well as discussions with the publisher to determine the competitive (demographic, editorial, or lifestyle) set
- Complete demographic analysis

Any prototyping model should be systematically and scientifically built, and well-tested. The modeler should disclose all underlying assumptions, when the model was developed, the composition basis, the number of variables being used, and a detailed validation, including the date of such validation.

Readers-per-Copy

Sufficient supporting documentation should be provided to enable a user to reproduce the estimate. Key considerations include competitive set, public place copies (numbers and general description by location) and mode of distribution.

For this analysis, when we encountered a situation where a prototypes had not historically been created, some host books were selected based upon the category averages of the unmeasured publication, for both audience composition as well as RPC.



APPENDIX C

BrandIndex

BrandIndex, developed by YouGovPolimetrix, is a daily measure of public perceptions of more than 1,000 U.S. consumer brands across 40 industry sectors. Seven key indicators of brand health are measured:

1. General Impression
2. Quality
3. Value
4. Customer Satisfaction
5. Loyalty/Recommendation
6. Corporate Reputation
7. Buzz

For each of the seven measures, respondents are asked to identify the brands within a certain industry sector (i.e. – banks and financial service companies, broadcast and cable television networks, etc.) to which they have a positive response.

Respondents are then prompted to select those brands within the same sector to which they have a negative response. The resulting score is the “net rating” (the positive minus the negative).

The survey is comprised of 14 questions regarding specific brands, bookended with questions regarding current affairs, and respondents’ lives and personal views on a range of topics.

BrandIndex interviews 5,000 U.S. adults 18+ each weekday. Each respondent answers questions for all seven indicators on only one sector per survey, generating 125 responses per sector, per measure, per day. Respondents are drawn from an online panel of more than 1,000,000 individuals. Respondent data may be broken out according to gender, age, ethnicity, education, income and/or geography. For this exercise we used BrandIndex data for gender and age.