Toward Total Audience – Integrating Magazines' Hardcopy and Internet Site Audiences Using Dynamic Segmentation Fusion

Risa Becker and James Collins, Mediamark Research Inc.
Mainak Mazumdar, Nielsen // NetRatings

Introduction

Increasingly media brands are investing substantial resources to extend their offerings beyond their traditional, established channels. Concomitantly, advertisers and their agencies are seeking to more fully understand, plan and measure cross-media audience behaviors the better to reach their consumer bases more efficaciously.

Confronting these efforts to develop (the media) and exploit (the agencies and advertisers) cross-media brands stands much traditional media audience measurement. In the United States and for much of the rest of the world the currency measures of the different major media – television, radio, print, the internet and outdoor – are founded on quite different methodologies. Each of these separate methodologies is designed and executed to optimize the measurement of its particular media and to do so to the highest quality standards reasonably attainable. To this end, media currency measurements generally focus on measuring a single media within a single population sample using a media specific methodology, the better to insure measurement quality.

Hence the dilemma for media, agencies and advertisers each pursuing their respective cross-media strategies – they want and need currency quality measurement for each media, but also the utility of an information source capable of illuminating cross-media consumer behaviors.

Fortunately, there are resolutions to this dilemma; all very much related to what media researchers have been doing more or less formally for decades and that is integrating the separate media measures using known relationships among them. One such alternative is data fusion – matching respondents from two or more separate (media) data sources using measures common to each of these sources to create a quasi-single-source database.

In the case at hand, Mediamark Research Inc. (*National Media and Marketing Survey*) and Nielsen//NetRatings (*NetView*) undertook a fusion of their respective media currency databases. The fusion of these two media currency databases affords media researchers – planners, buyers, marketers, etc. – the opportunity to analyze the relationships among magazines (MRI), internet properties (Nielsen//NetRatings), other media such as television and radio and an extensive range of consumer behaviors (MRI) using a single data source. Because of the complexity of the internet and related media consumption behaviors Dynamic Segmentation Fusion was developed drawing on features previously employed in both static and runtime fusion techniques. Following is a brief review of the 1) data sources involved in the fusion, 2) classical fusion techniques upon which Dynamic Segmentation Fusion is founded, 3) complexity of the internet site consumption, 4) basic process of Dynamic Segmentation Fusion, and 5) prototypical analyses of the sorts of media schedules available through the resulting fused database.

Data Sources

As noted above the two databases involved in this fusion are Mediamark Research Inc.'s *National Media and Marketing Study* and Nielsen//NetRatings' NetView.

Mediamark conducts an ongoing, comprehensive study of the adult population of the United States. Employing a strict area probability sample design the study surveys the demographics, product usage, and media exposure, with particular emphasis on magazine readership, of all persons aged 18 and over in the contiguous 48 states. Released twice yearly –

a rolling average of the two most recent six-month survey periods – the database contains approximately 26,000 respondents and serves as the ratings currency for the planning and buying of magazine advertising. Of particular relevance to the fusion, in addition to an extensive set of demographic measures the *National Media and Marketing* survey includes relatively substantial batteries of questions regarding internet access, generic internet activities and particular site consumption.

Nielsen//NetRatings' *NetView* is a random-digit-dial recruited panel of households (each household member aged 2 or older) and business employees having Internet access at home and/or at work. Each household or business computer has installed on it a software "meter" which tracks each panel member's internet behavior (pageviews, time spent, etc. by site). Released monthly with an average active panel size of approximately 20,000 *NetView* aggregates internet URLs and applications into three levels: the company-level (Parents) and two site/advertising-entity levels (Brands and Channels) for more recognizable and relevant site analysis.

In the MRI / Nielsen//NetRatings fusion under review, the latter's internet brand and channel information is fused to the former's database yielding a single information source integrating the repertoire of measures available in each separately.

Although employing quite different methodologies and modes of information collection the substantial number of measures (e.g., demographics, internet related activities and sites, etc.) common to both databases goes a substantial way toward insuring that sufficient discriminating measures are available to the underlying matching process.

Dynamic Segmentation Fusion - Background

While the particulars of individual fusions and fusion techniques vary, fusion is generally individuated from among other data integration techniques by its use of record/respondent matching. In particular, fusion integrates data from two or more separate sources through matching records/respondents from one database with those from the other(s). Essential to this process is the presence of variables common to all of the databases as it is on these variables that records/respondents are matched.

The presence of common matching variables, while essential to fusion, is for practical purposes generally insufficient. Simply because two or more databases possess common measures does not insure that 1) the distributions of these variables' values is uniform across the data sources and 2) that the correlations between any particular common variable and those unique to each database are sufficient to insure that the resulting relationships between the latter are authentic in the fused database.

In practice these two conditions are usually related. As usually not all respondents from the separate databases can be matched exactly on the common measures, conventionally fusion employs some sort of distance-based algorithm reflecting some matching hierarchy – e.g. it is more important to match respondents on Sex than exactly on Age assuming an exact match on both is not possible for all respondents.

The development of such a matching hierarchy is usually predicated on some formal understanding of the correlation structure between the common and the unique measures in the databases – e.g., it is more important to match respondents on Sex than on Age assuming an exact match on both is not possible because the correlations between Sex and the unique variables to be fused are stronger than those for Age, etc. In short, some common variables are more predictive of the measures unique to the datasets than are others and thus they should place more prominently in the hierarchy of matching variables.

While a variety of particular data fusion techniques exist they generally settle into one of two categories – static fusion and analysis-time fusion. When understood as a matching problem, the distinction between these two approaches is at what point and with what data is the hierarchy of common matching variables developed and the matching process undertaken.

With static fusions – where a persistent database results – data integrators usually spend considerable effort to develop an a priori matching hierarchy reflecting the discriminating power of all potential common matching variables. Generally, this a priori

hierarchy results from the analysis of the relationships between the common variables to be used for matching and the measures unique to the databases to be fused. Optimally, the result is a single database in which the relationships among the various unique measures are resilient and authentic across the multiplicity of possible analyses to which the fused dataset could be potentially subjected.

An alternative strategy, analysis-time fusion, (e.g., "Just In Time Data Modelling" [Raimondi, 1997], "Fusion-on-the-Fly for multimedia applications" [Soong, 2003]) is a technique where two or more separate data sources are fused as part of some other analysis, e.g., crosstabulation, etc. The advantage of this technique over static strategies is that the hierarchy of common matching variables can be optimized for each particular analysis. The disadvantage of analysis-time fusion is that the relationships between two unique variables in one analysis may be altered in subsequent analyses involving a different constellation of variables due to changes in the hierarchy of matching variables. Moreover, analysis-time fusion requires the availability of specific software at the time of the analysis consequently limiting access to the integrated data and restricting the availability of such data from use in standard media analytic applications.

In the fusion of the MRI and Nielsen//NetRatings NetView currencies, given the relative complexity (see next section) of both internet and print consumption behaviors the use of a static, monolithic hierarchy of matching variables, as employed in conventional static fusions is compromizing. Moreover, clients' need to have a stable dataset yielding consistent levels using existing analysis systems forecloses on analysis-time fusion. Hence, Dynamic Segmentation Fusion – a technique deriving from prior and established data fusion techniques - capitalizing on their respective strengths while avoiding their primary vulnerabilities.

Complexity of Internet Behaviors

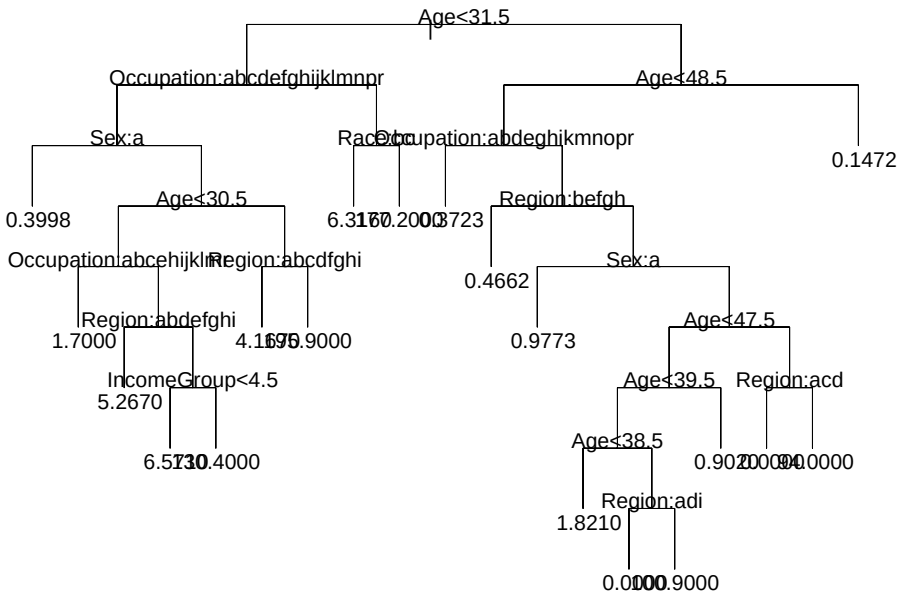
As noted previously, media consumption behaviors are relatively complex, especially so with respect to the internet wherein the number, variety and specificity of site choices is enormous. By way of example consider a gaming site and an epicurean one. Regression tree analyses using Nielsen//NetRatings NetView data for March, 2007 with a constant set of independent variables (Sex, Education, Occupation, Household Income, Hispanic, Race, Region, Age, and Presence of Children) such as are common to both the MRI and Nielsen//NetRatings datasets, and page views for the respective sites as the separate dependent variables suggest quite different discriminating configurations. Considering just the first two levels of each regression tree, for the game site (Graph #1) Age is the most discriminating measure leading to Occupation and another break within Age, whereas for the epicurean site (Graph #2) Household Income leading to Sex and Race are most discriminating – very different configurations of discrimination.

Moreover, the arrangements of discriminators are not themselves ordered into a simple hierarchy as for example in the epicurean site regression tree wherein Sex (Female) is most discriminating among the left split/lower income branch whereas Race is more discriminating among the right split/higher income branch.

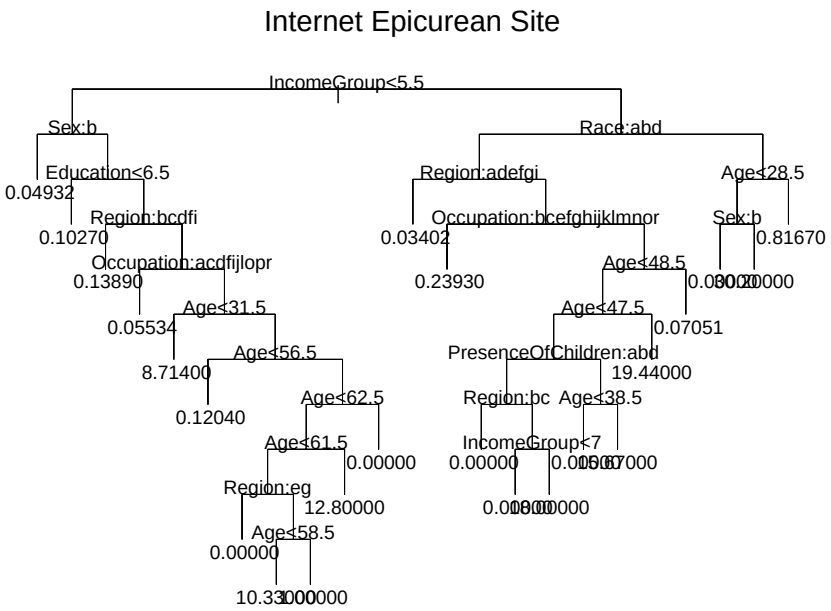
Graph #1
Internet Game Site Tree Analysis

Mediamark Research Inc.
A Member of the GfK Group

Internet Game Site

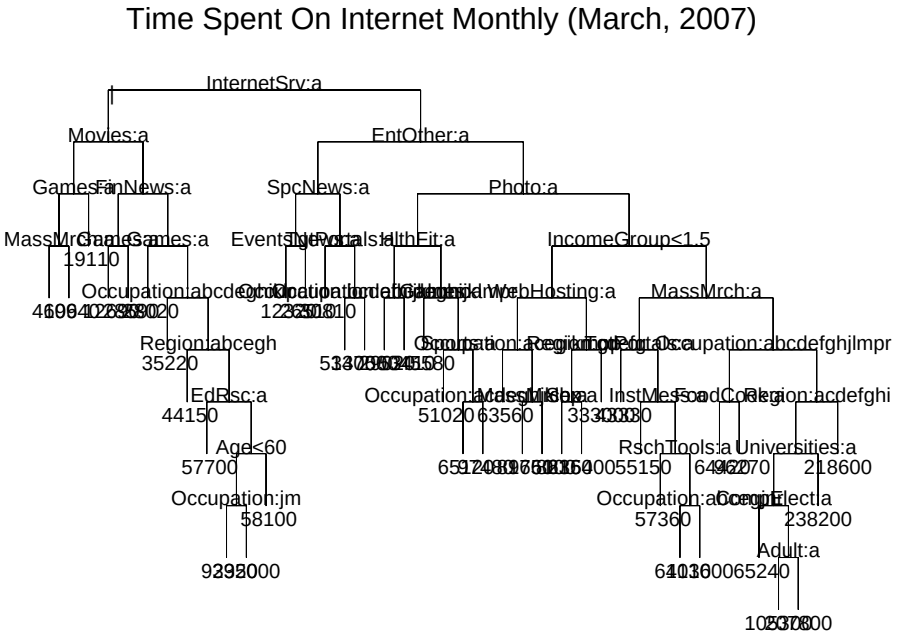


Graph #2



Internet Epicurean Site Tree

Graph #3
Time Spent on the Internet Monthly



This second point is especially important with respect to the entire fusion process, for it suggests that a monolithic hierarchical approach to matching respondents from the two databases is problematic. Chart #3 is the output from a regression tree analysis of NetView data. The dependent variable is Time Spent on the Internet Monthly (March, 2007) and an extensive variety of demographic and internet site category related measures comprise the independent ones. (This set of measures and the regression tree are very analogous to the components of the actual MRI / Nielsen//NetRatings fusion under review.) Just as was apparent in the regression trees for the epicurean and gaming sites, the independent/predictor measures are very much not monolithically arrayed into a simple hierarchy, but track a very rugged and complex terrain.

Dynamic Segmentation Fusion – Details

Dynamic Segmentation Fusion is designed to facilitate respondent matching in contexts wherein a simple hierarchy of common matching measures is inadequate given the complexity of the data being fused, but where nevertheless a single stable fused dataset is required. The premise of Dynamic Segmentation Fusion is that while there may not be a simple single hierarchy of matching variables that relatively discrete homogeneous sets of respondents can be systematically identified and that this homogeneity is in respect to a particular and unique hierarchy of discriminating common variables and associated values.

The use of regression/classification tree analysis in prior sections of this work foreshadowed the statistical technique employed as the classifier in this particular fusion. However, there is a variety of other classification and clustering techniques that would be adequate, e.g., CHAID, MARS, rule-induction, etc. among the guided methods, and potentially even unguided clustering such as

k-means or even self-organizing neural-networks. The distinct advantage of guided over unguided techniques is that in their use of a dependent variable control is exerted over the precedence of the common matching variables.

For the MRI / Nielsen//NetRatings fusion, Dynamic Segmentation Fusion begins by systematically developing a classification tree-based model using the NetView dataset – the internet currency. The dependent variable in the tree model is Time Spent on the Internet in a Month, chosen because of its broad relationship to the intensity of internet use. (In instances where there is no natural dependent variable unguided strategies would be employed.) As independent variables the various demographics and internet-related measures (over 100) common to both datasets are used.

Each NetView person/record is classified into a particular node of the tree with the result that persons/records occupying the same tree-node are relatively homogeneous with respect to the most discriminating common measures defining that node. The persons/records from the MRI dataset are next classified through this NetView based tree model resulting in two corresponding sets – NetView and MRI - of relatively homogeneous groups of internet users who are then matched node-to-node. In addition to matching respondents across the corresponding nodes of the two databases additional controls are imposed with respect to Sex, Age, Hispanic Origin and Race. Moreover, split-weighting/Transportation Algorithm is employed to constrain the distribution of projection weights and thus preserve currency levels from the individual datasets in the final fused database.

The advantages of this strategy over its predecessor methods are:

1) No a priori assumptions with respect to the hierarchy of common matching variables need be made
2) Persons/records are conveniently organized for the purposes of the fusion matching by the classification procedure
3) If there is not a single exhaustive segmentation of persons/records between the two datasets, then the classification can proceed iteratively with one set of persons/records being fused using the dominant segmentation scheme, with the classification tree being rebuilt and/or pruned to reclassify the remaining persons/records.
4) A stable database usable in a variety of different analysis systems results

The third item above deserves special note as it is related to an issue identified previously - "Simply because two or more databases possess common measures does not insure that 1) the distributions of these variables' values is uniform across the data sources ". The consequence of this for Dynamic Segmentation Fusion is that not all respondents are necessarily exhaustively matched with the initial classification, especially as Transportation Algorithm/Split Weighting constrains respondent distribution. Corresponding nodes of the classification tree do not necessarily contain identical population sums as the distributions of the values for those variables/values defining the particular nodes may be different. Thus, as a practical matter the matching process proceeds iteratively, with the classification tree being prune/trimmed and/or rebuilt until matching is exhausted.

While the quality of the matching strategy employed by a fusion technique is a necessary condition for insuring the quality of a particular fusion it is not sufficient. As important as is the technique is the extent and relevance of the common variables underlying the match, which in the case of Dynamic Segmentation Fusion comprise the classification/regression tree model's independent variables.

The MRI / Nielsen//NetRatings fusion employs an extensive range of both demographic and internet site related measures:

1) Demographics
    a. Sex, Income Group, Education, Hispanic Origin, Race, Census Region, Household Size and Occupation
2) Internet Site Category Use
    a. Automotive
        i. Automotive Information
        ii. Automotive Manufacturer

iii. Multi-Category Automotive

b. Computer Hardware and Software
   i. Computer & Consumer Electronic News
c. Education & Careers
   i. Career Development
   ii. Multi-Category Education & Careers
d. Entertainment
   i. Broadcast Media
   ii. Events
   iii. Gambling/Sweepstakes
   iv. Multi-Category Entertainment
   v. Music
   vi. Online Games
   vii. Sports
   viii. Video/Movies
e. Family & Lifestyles
   i. Family Resources
   ii. Health & Fitness & Nutrition
   iii. Kids, Games & Toys
   iv. Multi-Category Family & Lifestyles
   v. Personals
f. Finance/Insurance/Investments
   i. Credit Card
   ii. Financial News & Information
   iii. Financial Tools
   iv. Online Trading
g. Home & Fashion
   i. Food & Cooking
   ii. Real Estate / Apartments
h. Shopping
   i. Classifieds/Auctions
   ii. Mass Merchandiser
   iii. Shopping Directories & Guides
i. News & Information
   i. Current Events & Global News
   ii. Directories/Local Guides
   iii. Multi-Category News & Information
   iv. Research Tools
   v. Weather
   vi. 4", "CatSubCat.11.6"
j. Portals & Communities
   i. General Interest Portals & Communities
   ii. Member Communities
   iii. Search
k. Telecommunication / Internet Services
   i. E-Mail
   ii. ISP
   iii. Long Distance / Local Carrier

iv. Instant Messaging

l. Travel
   i. Airlines
   ii. Cruise Lines
   iii. Destinations
   iv. Ground Transportation
   v. Hotels/Hotel Directories
   vi. Maps / Travel Info
   vii. Multi-Category Travel

In addition a number (approximately 100) of magazine and national newspaper internet sites are also measured in common and employed as part of the fusion process.

The importance of behavior-relevant measures – in the case at hand, internet behavior in the MRI database – as common matching variables is especially high and the availability of a large number of them is critical to the success of the fusion. Emphasizing this point in commenting on several media and product related fusions Suzanne Rassler in Statistical Matching comments:

> "The importance of the common variables forming a link between the specific variables of the donor and recipient sample was emphasized for conducting an efficient statistical match…
> Within media and consuming data the typical demographic and socioeconomic variables will surely not completely explain media exposure and consuming behavior. Variables already concerning media exposure and consuming behavior have to be asked as well. Thus, the common variables also have to contain variables concerning television and consuming behaviors…Roberts (1994) reports better results using such 'specific common' variables than the usual demographic and socioeconomic issues alone." [Rassler, 2002]

Cross-Channel Media Analysis

While the soundness of the fusion technique and the discriminating power of the common measures are essential to the quality of the resulting database, ultimately it is the information and insights it offers up with respect to cross-media and related behavior that constitute the genuine measure of its value. To begin to explore the possibilities offered by a multi-media database brief examination of three prototypical media research and planning scenarios will be offered:

1) Assessing the extent to which a media brand reaches its audience through a variety of different channels
2) Multi-media reach for a variety of different demographic targets
3) Developing a media plan for a consumer attitude/preference target

For the purposes of these analyses we have chosen to use schedule reach/frequency rather than tabulations of net audiences insofar as reach/frequency allows for a more realistic assessment of how the media brand actually delivers audience for a particular advertiser's schedule.

Multi-Channel Media Brand Exposure

As noted in the Introduction, most significant media brands are aggressively developing multiple delivery channels the better to reach their traditional and emerging audiences with their content – an anytime, anywhere strategy. To begin to develop a

measure of the general utility of an integrated database in analyzing and assessing a significant media brand's audience exposure across multiple channels we will consider a media franchise with substantial cable, magazine and internet properties.

Table #1 below contains reach/frequency results from the MRI / Nielsen//NetRatings fused database for a prototypical single week schedule utilizing this media brand's various channels for total adults and four age categories. Clearly, the parent cable network and its extensions together constitute the primary audience delivery channel, and as an established cable brand generates relatively consistent audience levels across the age cohorts. But perhaps most important with respect to cross channel delivery, both the magazine and the internet site generate relatively substantial incremental reach, particularly for the 18-34 and 35-49 age cohorts. In branding terms the cable channel is mature and relatively stable, but the newer channels, the magazine and the internet are bringing new, younger audiences to the brand.

Table #1
Media Brand – Multi-Channel Reach

		Reach				
Media Brand	**Schedule**	**Total**	**Age 18-34**	**Age 35-49**	**Age 50-64**	**Age 65+**
Cable Network - Primary (Spots)	**10**	13.5%	14.3%	13.9%	12.1%	13.2%
Cable Network - Extension (Spots)	**10**	7.4%	7.8%	7.8%	6.2%	7.3%
Cable Network - Historic (Spots)	**10**	2.3%	2.6%	2.7%	2.0%	1.4%
Cable Network - News (Spots)	**10**	4.0%	4.9%	4.3%	3.5%	2.8%
Total Cable	**40**	***18.9%***	***19.6%***	***20.0%***	***17.0%***	***18.1%***
Magazine (Inserts)	**1**	5.5%	9.3%	6.0%	2.6%	1.1%
Internet Site (Pageviews)	**2,000,000**	0.8%	0.8%	1.0%	0.4%	0.1%
Total Target Reach		**48,144,834**	**16,650,990**	**15,608,847**	**9,301,557**	**6,583,440**
Total Target Reach%		**22.0%**	**24.6%**	**23.6%**	**18.7%**	**18.5%**

Multi-Media Reach

A fundamental media planning analysis is schedule reach/frequency. In the context of multi-media campaigns and using discrete media currency databases, conventionally separate schedules with associated reach (and frequency) results are developed for each of the media types. To integrate these separate results some estimation must be made with respect to duplication – how much unique reach is contributed by each media type to the total multi-media schedule and how much frequency developing duplication?

A variety of different approaches to determine inter-media duplication have been developed – some formal, some more rough-and-ready – but as the true duplication relationships among the individual vehicles from different media types in the separate currency measures is not known all of these strategies involve estimation. This holds true for fusion in general and Dynamic Segmentation Fusion in particular and thus the quality of the inter-media duplication estimates is a primary artifact of the fusion's quality. As the MRI / Nielsen//NetRatings fusion involves 1) an extensive set of common matching variables – the independent variables in the regression tree model – with many related to specific internet site behavior, 2) discerns fairly discrete homogeneous respondent groups with respect to these behaviors when matching and 3) preserves the currency levels for the total populations and generally for primary Age/Sex groups, the quality of these duplication estimates is minimally fairly good.

(As part of the validation process duplication relationships in the fused MRI / Nielsen//NetRatings database were assessed against comparable ones obtained from other single-source measures (non-currency) and there was general alignment albeit with some modest attenuation to the mean.)

Considered as a source the MRI / Nielsen//NetRatings database offers a fairly straightforward approach to multi-media reach/frequency. Table #2 summarizes results from a prototypical one week news oriented schedule including print (MRI magazine audience currency estimates), internet (Nielsen//NetRatings internet site pageviews currency) and television (MRI's non-currency television viewing). (While many more pageviews could have been scheduled for the internet sites, it is unlikely that any media plan would include the purchase of a week's worth of pageviews for a site.)

Table #2
Multi-Media News Schedule Reach/Frequency

Media Vehicle	Schedule	Total (18+)	Age 18-34	Age 35-49	Age 50-64	Age 65+
News Magazine A (Inserts)	1	8.2%	7.9%	8.8%	8.7%	6.8%
News Magazine B (Inserts)	1	10.6%	10.4%	11.7%	10.6%	9.1%
News Magazine C (Inserts)	1	5.2%	4.8%	5.2%	6.0%	5.3%
Total Print Reach %	***3***	***19.1%***	***18.4%***	***20.3%***	***20.0%***	***16.7%***
Newspaper Site A (Pageviews)	200,000	0.1%	0.0%	0.1%	0.1%	0.1%
Newspaper Site B (Pageviews)	500,000	0.2%	0.1%	0.3%	0.2%	0.1%
Newspaper Site C (Pageviews)	1,000,000	0.3%	0.5%	0.3%	0.3%	0.1%
Internet Portal A - News (Pageviews)	2,000,000	0.7%	0.4%	1.0%	1.0%	0.2%
Internet Portal B - News (Pageviews)	2,000,000	0.7%	0.6%	1.0%	0.8%	0.4%
Internet Portal C - News (Pageviews)	2,000,000	0.4%	0.4%	0.6%	0.5%	0.2%
Total Internet Reach %	***7,700,000***	***2.3%***	***2.0%***	***3.1%***	**2.6%**	***1.0%***
Network A Evening News - Sat. (Spots)	1	7.0%	3.5%	6.6%	8.5%	12.0%
Network A Evening News - Sun. (Spots)	1	6.3%	3.0%	5.8%	8.3%	10.6%
Network B Evening News - WkEnd. (Spots)	1	2.3%	1.5%	2.3%	2.8%	3.1%
Network C Evening News - Sat. (Spots)	1	5.9%	2.5%	5.5%	7.0%	11.8%
Network C Evening News - Sun. (Spots)	1	6.0%	2.3%	5.7%	8.0%	10.6%
Network D Evening News - Sat. (Spots)	1	7.0%	3.6%	5.5%	8.8%	13.7%
Network D Evening News - Sun. (Spots)	1	7.8%	4.5%	6.1%	9.5%	14.6%
Network A News Interview (Spots)	1	2.5%	1.0%	1.9%	3.3%	5.3%

Network C News Interview (Spots)	1	2.2%	0.8%	1.7%	3.4%	4.4%
Network C Morning Show - Sat (Spots)	1	1.9%	1.1%	2.2%	2.4%	2.0%
Network A Evening News - WkDay. (Spots)	5	15.1%	11.2%	14.8%	18.4%	18.5%
Network B Evening News - WkDay. (Spots)	5	2.9%	2.7%	2.6%	4.0%	2.4%
Network C Evening News - WkDay. (Spots)	5	14.9%	10.5%	13.6%	18.3%	20.9%
Network D Evening News - WkDay. (Spots)	5	13.4%	8.6%	11.4%	16.6%	21.7%
Network C Morning Show - Sun (Spots)	1	3.6%	1.4%	3.6%	5.4%	5.3%
Total TV Reach %	***31***	***42.6%***	***28.0%***	***40.5%***	***50.8%***	***62.8%***
Total Target Reach		**117,795,987**	**28,349,359**	**34,947,503**	**30,309,735**	**24,189,389**
Total Target Reach%		**53.8%**	**41.8%**	**52.9%**	**61.1%**	**68.1%**

As individual respondent-level probabilities for each media vehicle can be calculated fairly straightforwardly a personal probability reach/frequency model is rather naturally recommended and thus employed in this exercise as with the other two. In the case of the magazines and television programs, as measured by MRI, the underlying respondent probabilities are derived from the frequency of exposure measures (theoretical probabilities for television, empirical probabilities for magazines) for each vehicle. In the case of the Nielsen//NetRatings internet sites the probabilities are derived from the respondents' pageviews in relation to the ratio of planned to total pageviews for the site.

With respect to the utility – information and insights - offered by the fused database it is clear that different Age cohorts are exposed at quite different levels to news across the three different media types (see also Chart #1).

1) News consumption in total is higher among older Age cohorts
   a. 41.8% for Age 18-34 v. 68.1% for Age 65+
2) News consumption from internet sites is lowest among the oldest Age cohort, probably as a result of the lower level internet penetration and use among this cohort
3) The pattern of news consumption for internet and news magazines is roughly parallel albeit with higher levels for the latter
4) The overall duplication rates among the media types for news are fairly constant despite the very differences in total reach and reach for the different media type (Table #3)

Chart #1
Patterns of News Consumption

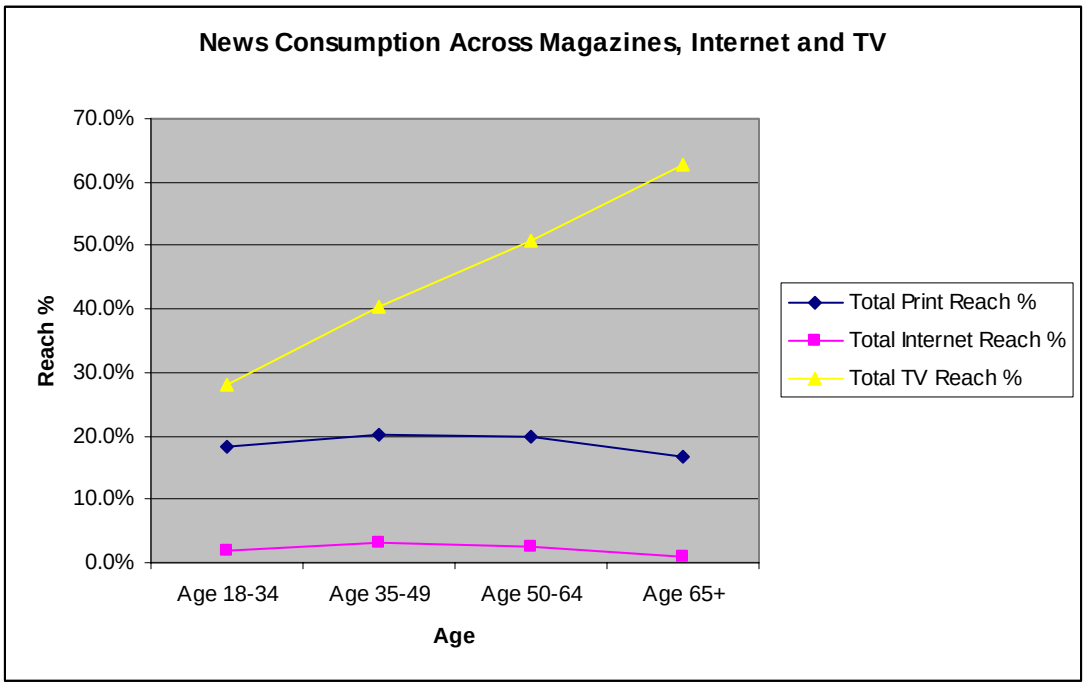


Table #3
Aggregate to Total Reach Ratio

Total Print Reach %	19.1%	18.4%	20.3%	20.0%	16.7%
Total Internet Reach %	2.3%	2.0%	3.1%	2.6%	1.0%
Total TV Reach %	42.6%	28.0%	40.5%	50.8%	62.8%
Aggregate Reach %	**64.0%**	**48.4%**	**63.9%**	**73.4%**	**80.5%**
Total Target Reach%	**53.8%**	**41.8%**	**52.9%**	**61.1%**	**68.1%**
Aggregate / Total Reach Ratio	**1.19**	**1.16**	**1.21**	**1.20**	**1.18**

Developing a Media Plan for a Consumer Attitude/Preference Target

The prior two analyses have focused on demonstrating the potential of a fused media currency database to support the analysis of cross-channel delivery of a single media brand and a multi-brand/multi-media reach/frequency schedule. However, the potency of such a database can be more fully exploited if one or more of the constituent databases to the fusion have consumer behavior measures, and even attitudinal/preference data. Print currency databases, such as MRI's, have traditionally offered not only their currency measures – magazine audiences – and other media on a non-currency basis, but also a plethora of consumer-relevant measures. As such, with the MRI / Nielsen//NetRatings fusion analysts are able to explore not only cross-media and media/demographic relationships but also ones involving rich varieties of consumer behaviors and preferences.

Indeed, many of the fusions undertaken in Britain and Europe beginning in the 1980's and continuing through to the present, have focused on integrating just such consumer measures (and print) to enhance television planning and buying.

In the prototypical example at hand an analysis of travel related media – magazines, cable channels, and internet sites – in relationship with travel preferences is developed.

Table #4 summarizes reach results for four different travel-oriented preferences for a prototypical schedule employing cable, magazine and internet media. Comparing and contrasting two broadly antithetical preferences "Rather Travel in U.S." with "Foreign Travel is a Good Way to Learn About Other Cultures" the print and internet components of the schedule offer stronger reach against the latter preference than the former. Moreover, given the comparatively weaker travel media profile of the "Rather Travel in U.S." target – relatively weak on all media reach measures relative to the other three targets and even Total – this group might better be reached through a more expansive media schedule involving travel and non-travel media.

Separate from the question of cross-media scheduling, the relationship between these travel targets – measured in the MRI study – and the travel-related internet sites – measured through the Nielsen//NetRatings service – supports the general integrity of the fusion process. As "Foreign Travel is a Good Way to Learn About Other Cultures" is not a measure common to the MRI and Nielsen//NetRatings databases the fusion does not explicitly control for it. Nevertheless, the availability of matching variables either directly associated with internet/travel (see Travel items listed in Dynamic Segmentation Fusion – Details above) or separately correlated with travel and internet use (e.g., Education, Income, etc.) insures prima facie validity to the results.

Table #4
Travel Media and Consumer Attitude Reach/Frequency

			Reach			
Media Vehicle / Internet Site	**Schedule**	**Total**	**Discounts Influence Travel Plans**	**Willing to Pay More to Use Favorite Airline**	**Rather Travel in U.S.**	**Foreign Travel is a Good Way to Learn About Other Cultures**
Cable Network A (Spots)	**10**	4.1%	4.3%	4.2%	4.0%	4.3%
Cable Network B (Spots)	**10**	5.4%	5.5%	5.7%	5.7%	5.4%
Total Cable	***20***	***8.8%***	***9.1%***	***9.4%***	***9.0%***	***9.0%***
Magazine A (Inserts)	**1**	0.9%	0.9%	0.8%	0.7%	**0.9%**
Magazine B (Inserts)	**1**	1.3%	1.3%	1.2%	1.0%	1.5%
Magazine C (Inserts)	**1**	1.0%	1.2%	1.1%	1.0%	1.1%
Magazine D (Inserts)	**1**	3.4%	3.5%	3.8%	3.0%	3.8%
Magazine E (Inserts)	**1**	2.1%	2.1%	2.2%	1.7%	2.4%
Total Print	***5***	***7.2%***	***7.4%***	***7.6%***	***6.1%***	***8.1%***
Travel Site A (Page Views)	**2,000,000**	0.5%	0.5%	0.5%	0.4%	0.5%
Travel Site B (Page Views)	**2,000,000**	0.7%	0.7%	0.7%	0.7%	0.8%
Travel Site C (Page Views)	**2,000,000**	0.7%	0.7%	0.6%	0.7%	0.7%
Travel Site D (Page Views)	**2,000,000**	0.5%	0.4%	0.4%	0.4%	0.6%
Travel Site E (Page Views)	**2,000,000**	0.2%	0.2%	0.2%	0.2%	0.2%
Total Internet	**10,000,000**	***2.2%***	***2.2%***	***2.2%***	***2.1%***	***2.5%***
Total Target Reach						

	36,657,572	**22,954,046**	**10,298,884**	**20,429,788**	**27,753,659**
Total Target Reach%	**16.7%**	**17.1%**	**17.4%**	**15.9%**	**17.9%**

Conclusion

The availability of separately measured media currencies in a single data source affords the media, advertisers and their agencies broad potential for better understanding complex media and product consumption behaviors. While "there is no free lunch" [Soong, 2004] with fusion or any other data integration method, the ability to analyze high quality currency measures – the best measures of their respective media - in common allows media planning and brand management (of both advertisers and the media) to more fully identify, exploit and validate opportunity. Moreover, the very dynamism of technological change as the source of development of new media channels necessitates the concomitant development of audience measurement methodologies and techniques facilitating integration of their measures.

Endnotes:

Raimondi, Dina and Santini, Gilles (1997). Just In Time Data Modelling. Eighth Worldwide Readership Research Symposium, Vancouver, BC, Canada.

Soong, Roland and de Montigny, Michelle (2003). Fusion-on-the-Fly for multimedia applications. Eleventh Worldwide Readership Research Symposium, Cambridge, MA, USA.

Rassler, Susanne (2002) Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches. Springer-Verlag, New York.

Soong, Roland and de Montigny, Michelle (2004). No Free Lunch in Data Fusion / Integration. ESOMAR, Cross Media Conference, Geneva.

The Authors:

Risa Becker joined MRI in 1999 and was promoted to VP Research Operations in 2002. She is responsible for the day-to-day execution of MRI's national, syndicated media and product study. She has also conducted proprietary studies concerning media usage among Hispanics. Risa previously worked at Audits and Surveys Worldwide on the Los Medios y Mercados en Latinoamerica project. She has co-authored papers on title confusion and on magazine audience accumulation, presented at the WorldWide Readership Symposium. Risa received her B.A. from the State University of New York at Binghamton and her J.D. from American University, Washington College of Law in 1992. She is a member of AAPOR.

James H. Collins is Senior Vice President of Research for Mediamark Research Inc. where he focuses on data integration and related initiatives. Holding a doctorate from Temple University he has taught at Temple University, Baruch College of the City University of New York and the Trenton State College School of Business. A frequent presenter at media and market research conferences, Dr. Collins' work has been selected for inclusion in ESOMAR's Excellence in International Research 2002 and 2003. In addition to broad interest in data integration techniques he has published and presented on multi-media reach/frequency and optimization, the use of artificial intelligence techniques in media research, statistical analysis of text-based information and the economics of cross media consumption.

Telephone: 212-884-9307
e-mail: jim.collins@mediamark.com

Mainak Mazumdar, Chief of Measurement Science, Nielsen//NetRatings

Mainak Mazumdar is the chief of measurement science at NetRatings, a global leader in Internet media and market research. He is a recognized expert in marketing, advertising and media measurement methodologies, and an innovator in the application of traditional market research techniques to the online world.

At NetRatings, Mr. Mazumdar manages Media Products (NetView, @Plan and AdRelevance), measurement panels and all aspects of methodology in the U.S. and EMEA, including Panel-Server data integration and other modeling initiatives

Previously, Mr. Mazumdar held research and product management positions with NPD Group, Jupiter Media Metrix, DoubleClick and VNU.

Mr. Mazumdar frequently speaks about audience measurement and advertising research at industry conferences including ESOMAR, ARF, and IAB. His expertise is also in high demand from news outlets such as CNBC and CNN, as well as publications including New York Times, USA Today and The Industry Standard.

Mr. Mazumdar holds a B.A. in Economics and International Relations from Universitaire American in Switzerland, and both an M.A. in Demography and Sociology and a Ph.D. in Demography from Brown University.

Email:mainak.mazumdar@netratings.com

Ph: 212 703 5931