SYNTHESIZING DATA FOR MEDIA SIMULATIONS

Hugh M. Cannon, Wayne State University
Karthekayan Ramachandran, Wayne State University
Edward A. Riordan, Wayne State University

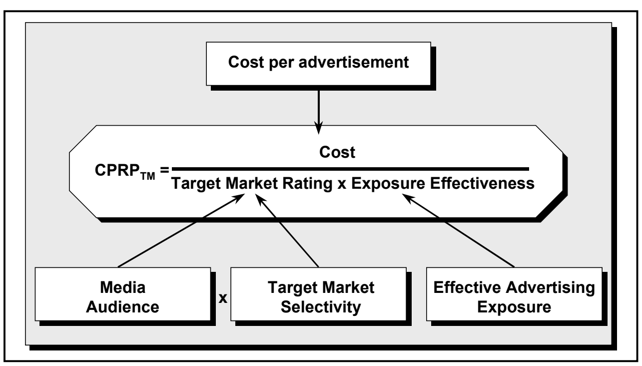
ABSTRACT

Media planning simulations offer enormous potential for increasing the efficiency of training professionals in the area of media planning. However, the simulations themselves can be very complex, requiring large quantities of data. Even with the vastly improved capacity of electronic storage devices, the requirements can be overwhelming, both in terms of space and the physical requirements of locating and transferring data. This paper discusses practical procedures for synthesizing the large data sets required for a media simulation using data that are readily available to simulation developers. These procedures are illustrated through examples from television media.

INTRODUCTION

The data requirements of any simulation are driven by the nature of the model around which the simulation is based. Our model suggests that

EXHIBIT 1: EVALUATING MEDIA ALTERNATIVES



media alternatives can be evaluated on two levels. The first is at the level of individual media vehicles, where the criterion of effectiveness is the cost per thousand effective target market rating point ($CPRP_{ETM}$).

The second level involves the evaluation of combinations of media vehicles, resulting in a distribution of advertising exposures). The relative efficiency of the schedule can be expressed in terms of the frequency value per dollar spent on advertising (Exhibit 2).

EXHIBIT 2: FREQUENCY VALUE ANALYSIS

% of target receiving each level of exposure

% of the theoretical optimal response achieved by this level of advertising exposure

Level of Exposure	Frequency Distribution	Response Effectiveness	Exposure Value
0	36%	0%	0.00%
1	49%	40%	19.60%
2	14%	50%	7.00%
3	1%	55%	.55%
Total			27.15%

% of the theoretical optimum delivered by the schedule

Exhibit 3 combines these two levels of analysis into a comprehensive media planning process (See Cannon, Leckenby and Abernethy 1996). This will provide the basic model for which the paper seeks to address the data requirements.

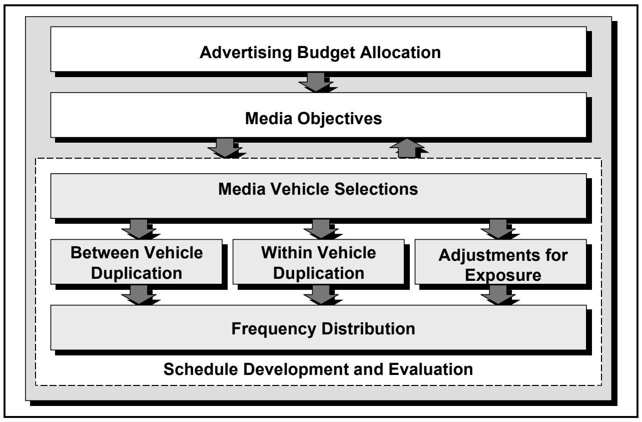
DATA REQUIREMENTS

The evaluation presented in Exhibits 3 suggest the need for several different kinds of data:

1. cost of ads placed in alternative media vehicles, including various media options (formats and size ads within each vehicle);
2. media audience (the number of people exposed to a particular media vehicle);
3. target market selectivity of alternative media vehicles (including both product-usage and de-

mographically defined targets);
4. media vehicle exposure effectiveness relative to advertising objectives;
5. within and between vehicle duplication.

EXHIBIT 3:
THE MEDIA SCHEDULING PROCESS



If the simulation were limited to a relatively few alternatives, these data could be simply entered into the simulation in the form of data matrices. However, in order to create an educational tool that offers a high level of creative flexibility, we would like a simulation that includes a much broader range of alternatives. Indeed, we would like the flexibility of incorporating a virtually unlimited number of alternatives. Even more important, our simulation sould allow students to select media that might not have been anticipated by the simulation designer.

Media Vehicle Costs

Media cost data can be accommodated with simple list of costs. However, the data problem becomes more complex when we consider the multitude of alternative options available. For instance, in television, these include 30-second versus 60-second commercials, station-break commercials versus those embedded within an actual program, commercials aired during different seasons of the year, and so forth.

As it turns out, most of the variance in cost is explained by distinctions among the options themselves rather than variations across media vehicles. That is, the relationship (i.e. ratio) between the cost of a 30-second and a 60-second television commercial will be relatively constant across programs and time slots, even though the actual cost varies considerably. This suggests that the cost data problem can be simplified by using cost estimates, starting with a basic cost-per-rating-point (CPRP) figure, then modifying it based on media attributes. For instance, a standard 30-second commercial would have a cost adjustment index of 1.0, while a 60-second commercial would have an index of 1.8, and so forth (Exhibit 4).

EXHIBIT 4:
USING ATTRIBUTE COST INDICES

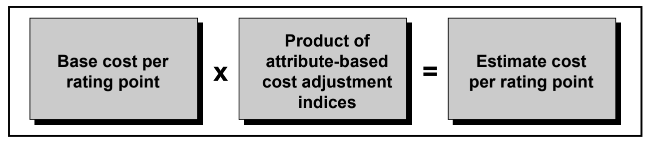


Exhibit 5 provides a summary of indices for different media attributes. Note the versatility of the approach. They can be used to represent everything from different formats within a particular medium (30- versus 60-second or four-color versus black & white) to different media classes (television versus radio, etc.). In each case, we use the attribute indices to adjust the cost per rating point. Therefore, any variation in price must be due to factors other than the size of the audience. To get the estimated cost, we simply multiply the CPRP by all of the relevant indices. For example, if the basic CPRP is $6,000, the daytime CPRP would be ($6,000 x 0.50 =) $3,000. A daytime, 60-second, imbedded commercial, during the third quarter of the year, in a prestigious, mass-appeal program would be ($6,000 x .5 x 1.8 x .8 x .7 x 1.5 x .7 =) $3,175.

The indices shown in Exhibit 5 provide a useful set of attribute weightings for developing a media simulation. Indices such as these are often published in pocket media planning guides developed by various advertising agencies. However, they can also be developed for special applications by simply taking a sample of values from published sources, such as *Standard Rate and Data Service*, and averaging ratios of attribute costs relative to the base.

EXHIBIT 5: SAMPLE MEDIA COST INDICES

Television:		Magazines:		Vehicle Quality:	
10-second:	0.5	4-color:	1.0	Prestige:	1.5
30-second:	1.0	Black & white:	0.7	Specialty:	2.0
60-second:	1.8	2nd cover:	1.3	Mass appeal:	0.7
Embedded:	1.0	3rd cover:	1.3		
Station-break:	0.8	4th cover:	2.0	**Other Media:**	
Daytime:	0.5	Half page:	0.6	Cable:	0.3
Primetime:	1.5	Quarter page:	0.3	Newspaper	0.4
Late evening:	1.0	Business:	2.0	Supplements:	0.3
Weekend:	0.7	**Radio:**		Out-of-home:	0.2
1st Quarter:	1.0			Dir mail (pkg):	2.0
2nd Quarter:	1.1	60-second:	1.0	Dir mail (ind):	60.0
3rd Quarter:	0.7	30-second:	0.6		
4th Quarter:	1.2	1st Quarter:	0.4		
Spot:	1.0	2nd Quarter:	0.5		
		3rd Quarter:	0.5		
		4th Quarter:	0.4		

The implications of this analysis is that a users can enter a new media vehicle and the cost will automatically be calculated, using the indices presented in Exhibit 5. The user need only specify the required media characteristics

Media Audience

Media audience estimates, like media costs, are relatively straight-forward, with data readily available. Unlike media costs, they do not vary with media options. However, there are seasonal and regional variations, for some media, at least. Again, these can be handled as attributes. For instance, the incidence of television viewing is much lighter during the summer months, when people typically spend more of their time out of doors. It is also lighter in southern climates, where the climate lends itself to outdoor activities – especially during the winter.

For mainstream network programs, the problem is simple. The modeler need only extract ratings data from syndicated sources, providing a national average. From a very practical perspective, these data can be easily obtained from Simmons or MRI, which provide a comprehensive source, and are widely available. While these sources are not generally used for television media planning, this is primarily because they are gathered nationally and averaged over a two-year period. Hence, they do not account for regional or seasonal variations. However, this problem can be addressed by including a regional and/or seasonal adjustment index.

The greater problem arises for non-network programs. For instance, what about reruns? The programs are often the same as those carried by network stations, but they can be aired during any number of different dayparts, on different stations or different networks. It does not make sense that the ratings would be the same as it is for primetime network viewing. While the program material might be the same, the ratings would depend on the reruns' dayparts and network/station placement.

HUT		50.0%
Seasonal adjustment	x	.8
Share	x	30.0
Estimated national rating		12.0%
Regional adjustment	x	1 .1
Local population adjustment	x	.0170
US Population base (000s)	x	185,000
Local TV audience (000s)		415

EXHIBIT 6: ESTIMATING LOCAL TV AUDIENCE

Exhibit 6 illustrates a structure for establishing the estimates. It begins with households using television (HUT). HUT varies by daypart. Estimates of these variations are readily available from syndicated data sources. Among the most accessible to game developers are Simmons and MRI. While they are limited to national and biannual averages, they nevertheless provide a useful starting place for simulation development. Exhibit 7 provides set of useful values.

The first adjustment is to address seasonality. As we have noted, audiences vary with the time of

EXHIBIT 7:
HUT ESTIMATES BY DAYPART

Monday-Friday		Monday-Sunday	
6-7 AM	12%	6-7 PM	49%
7-8 AM	18%	7-8 PM	53%
8-9 AM	21%	8-9 PM	59%
9-19 AM	22%	9-10 PM	60%
10-11 AM	22%	10-11 PM	56%
11-noon	23%	11-midnight	41%
Noon-1 PM	26%	Midnight-1PM	25%
1-2 PM	29%	1-2 AM	15%
2-3 PM	29%	2-3 AM	10%
3-4 PM	30%	3-4 AM	7%
4-5 PM	33%	4-5 AM	5%
5-6 PM	40%	5-6 AM	5%

year, a fact that reflects the impact of competing activities and variations in personal interests. For instance, summer provides a host of outdoor alternatives that distract people from watching television. These are important to consider when estimating national ratings. Hence the need for a seasonal adjustment index. Exhibit 8 provides seasonal adjustment indices for television media. HUT times the seasonal adjustment index provides an estimated of the effective national HUT for any given daypart.

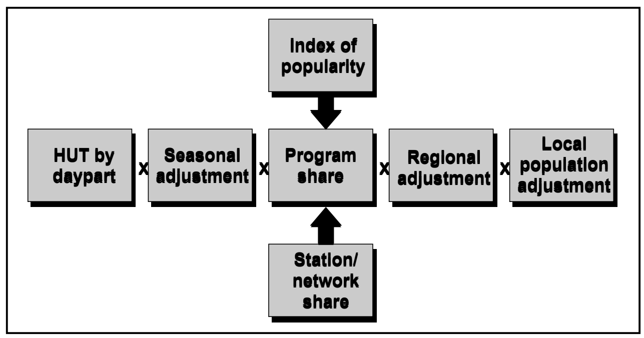
EXHIBIT 8:
SEASONAL ADJUSTMENT INDICES*

1st Quarter	1.2
2nd Quarter	0.9
3rd Quarter	0.8
4th Quarter	1.1

*Adjustments do not apply to early morning (1:00-7:00 AM) and weekday daytime (10:00AM-4:00PM) dayparts.

In order to estimate ratings, one need only multiply the effective HUT by the share of available television watchers who are viewing the target program. Exhibit 9 suggests that the share, in turn, is a function of two media attributes: (1) an index of program popularity and (2) the average share of the station or network. As with HUT, station and network share is readily available from syndicated sources. Program popularity is more difficult to estimate. It can be developed using the judgment of the simulation developer, or students who are adding a media alternative to the media plan. Or it can be estimated from syndicated data by comparing the alternative's rating (from prime time, or other slot where it is normally viewed) with the average rating for other programs. Thus, a program whose average share is half again that of other programs would have a popularity index of 1.5. Typically, popularity indices will vary between .5 and 2.0 for non-prime-time programming.

EXHIBIT 9:
ESTIMATE MEDIA AUDIENCE INDICES



While the preceding steps are sufficient for national media, we would also like a simulation to address local media as well. Exhibit 6 uses a regional adjustment index to modify local HUT. Exhibit 10 provides a set of adjustment indices. that reflect the variation from the national average due different region/season interactions.

The final factor in local adjustment is the local population adjustment. The model assumes that the rating will be the same for the local market situation as it is for the national market, subject to the urban viewership and regional adjustments. However, the base is different. The actual audience can be obtained by multiplying the rating by the proportion of the total United States population in the local market. This can be converted to an actual audience figure by multiplying it by the total U. S. population. Note that the effect of regional and season does not have a significant impact on early morning and daytime dayparts, as suggested by the footnote in Exhibit 10.

EXHIBIT 10:
REGIONAL ADJUSTMENT INDICES

	Quarters 1st	2nd	3rd	4th
New England	0.9	0.9	0.8	0.9
Middle Atlantic	1.2	1.1	1.1	1.2
West Central	0.9	0.9	0.8	0.9
South East	0.9	0.9	0.8	0.9
South West	1.0	1.1	1.2	1.0
Pacific	1.0	1.1	1.1	1.0

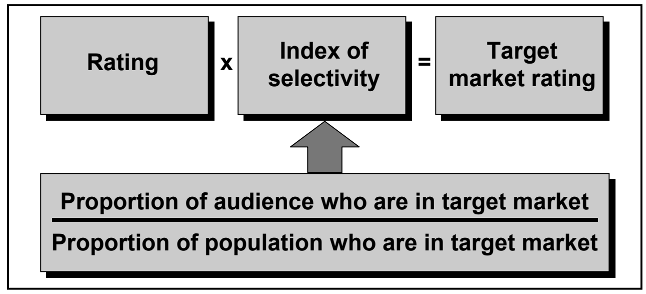
*Adjustments do not apply to early morning (1:00-7:00 AM) and weekday daytime (10:00AM-4:00PM) dayparts.

From a practical perspective, this discussion suggests that a user can add media to the simulation at will by simply estimating the relevant attributes. In the case of television, these would include HUT, seasonality, and share. If it is a local program, they would also need to include the station/network share, program popularity, MSA classification, region and market size, expressed as a percentage of the total US.

Target Market Selectivity

Target market selectivity can be expressed in an index of selectivity, as shown in Exhibit 11. As the exhibit suggests, the index is based on the ratio of target market concentration in the media audience, as compared to the concentration in the population as a whole. As a rule, media planners generally begin their analysis by defining target market membership in terms of product usage., assuming that past usage is the best indicator of the kinds of people who are likely to use the product. Assuming that target market membership is defined in terms of product usage, if 20% of the media audience were product users, whereas only 10% of population as a whole used the product, the index would be 2.0, or 200%.

EXHIBIT 11:
TARGET MARKET SELECTIVITY

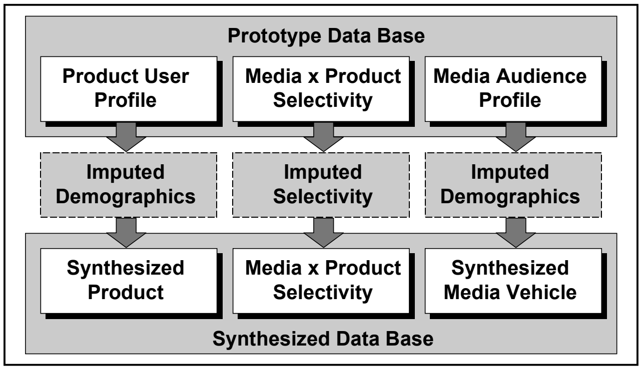


In order to get an index of selectivity, one needs a data set that contains both product usage and media audience membership. This, of course, is demanding indeed, not only for the simulation designer, but for actual media planners as well. Consider the thousands of media alternatives available, and the even greater number of products. Where would one get access to a data matrix consisting of thousands of media crossed with thousands of products?

As it turns out, a considerable amount of research has been done to identify various approaches for addressing the data problem. The most elegant solution, of course, would be *single-source data*, where the data base actually contained both product and media usage. Such data bases have existed since the early 1960s, first in survey form (Garfinkle 1963), and later through the combining of electronically gathered data (Garrick 1984, 1986). We have already mentioned Simmons and MRI as modern successors to the survey data approach. However, these sources involve enormous amounts of data, and even so, they are by no means comprehensive. Much of the motivation for this article is to provide principles for developing simulations that overcome these limitations.

Alternative approaches generally seek to synthesize the information provided in single-source data by linking product and media-usage data sets using demographics. (See Cannon and Seamons 1995 for a review). While the approaches are extremely popular, none of them performs consistently well. Furthermore, they are not useful for our data problem, since we are synthesizing both our product and media usage data. Therefore, they do not have any demographic correlates.

EXHIBIT 12:
ESTIMATING FROM "PROTOTYPES"



An alternative uses the logic of "prototyping" (Baron 1990/1; Cannon, Williams and Doyle 1992; Cannon, McGowan and Yoon 1995). That is, it draws on the characteristics of a limited number of variables (media and product usage categories), using them as "prototypes" for other variables of interest. According to this method, the selectivity of a given media vehicle relative to a product usage category can be predicted by the selectivity of the media and product prototypes, as suggested by Exhibit 12. The simulation, then, must include a matrix of representative product and media prototypes. As the exhibit suggests, the selectivity indices for any synthesized product usage or media variables we be imputed to be the same as those of corresponding prototypes.

While direct matching of product-usage and media-usage variables is the preferred approach to resolving the targeting problem, demographics are still the accepted language of media planning. As a result, the simulation should also include relevant demographic variables (generally age and sex). This suggests that the simulation should include a matrix of prototype products and media with the age and sex demographics. The demographic profile of product users and media audience members will be imputed to be the same as that of the prototypes.

The most broadly used demographic categories in media planning are males 18-24, males 25-34, etc. through females 65 and over. There is also some work done on categories of prototypic media, particularly magazines (e.g. Cannon, Williams and Doyle 1992). However, there is virtually no work done on the categorization of products, and the classification work on media falls far short of what we need for purposes of the simulation. Pending better research, game developers will be left to create their own systems for prototyping.

Media Vehicle Exposure Effectiveness

Media vehicle exposure effectiveness is by far the most difficult type of data to estimate. As a first step, many media planners try to estimate the the proportion of people who are media audience members who are likely to be exposed to a given advertisement. One survey suggests the figures suggested in Exhibit 13 (Kreshel, Lancaster and Toomey 1985). We can use the same numbers, adjusting them up or down, using the cost indices shown in Exhibit 5. Thus, if the rating of a television program were 20%, the estimated proportion of the population actually seeing a 10-second station-break commercial would be (20 x .5 x .8 =) 8%.

EXHIBIT 13:
AD/VEHICLE EXPOSURE RATIOS

Direct Mail	65.0%	Outdoor	16.0%
Magazines	52.5%	Radio	16.0%
Television	32.0%	Newspaper	16.0%

Rather than try to estimate effective exposure, we simply settle for advertising exposure. We then estimate effective exposure rate through the advertising response curve implied by the third column of Exhibit 2. Based on the of evidence of published studies, we can assume these to be concave in shape (Cannon and Riordan 1995). We can estimate the actual response levels by plotting the curve using a modified exponential function (Cannon, Leckenby and Abernethy 1996). In fact, the function can be embedded in a simulation game, thus enabling the simulation to automatically determine the exposure value of any given frequency distribution.

Within and Between Vehicle Duplication

The problem of within and between vehicle duplication is similar to that of target market selectivity. The question is what proportion of a media vehicle audience is also being reached by that of another media vehicle. The only difference between this and target market selectivity is that we are substituting a second media vehicle for product users. The same problems and solutions also apply. That is, there is no way we can include duplication data for all possible media vehicles, nor can we even use demographics

EXHIBIT 14:
USING PROTOTYPES TO ESTIMATE AUDIENCE DUPLICATION

	M_1	M_2	M_3	M_4	M_5
M_1	$M_{1,1}$				
M_2	$M_{2,1}$	$M_{2,2}$			
M_3	$M_{3,1}$	$M_{3,2}$	$M_{3,3}$		
M_4	$M_{4,1}$	$M_{4,2}$	$M_{4,3}$	$M_{4,4}$	
M_5	$M_{5,1}$	$M_{5,2}$	$M_{5,3}$	$M_{5,4}$	$M_{5,5}$

Key: M_1 = Audience for media vehicle 1
$M_{1,1}$ = Within vehicle 1 cumulative audience
$M_{2,1}$ = Between vehicle 1 and 2 cumulative audience

to match them. We can, however, use prototypes. Cannon, McGowan and Yoon (1995) provide a detailed description of how this might be done for magazine media. They develop a matrix of selectivity indices for prototypic media types (Exhibit 14). To get duplication, they estimate random duplication, and then multiply it by the appropriate selectivity index.

In oder to develop the initial matrix, duplication data are readily available for magazines from sources such as MRI and Simmons. Lancaster, Lee and Katz (1988) suggest a method by which they can be estimated for other media, using a regression model.

SUMMARY AND CONCLUSIONS

The objective of this paper has been to describe a process through which a large number of different target markets (both product usage categories and geographic markets) and media might be incorporated into an educational media simulation. Our solution has been to assume the that variance in the values of and relationships among media and target market variables can be explained in terms of key attributes, such as daypart, station, and program type in the case of television media, and product type in the case of target market variables that are defined by product usage. By including the effects associated with the attributes in a data base that accompanies a simulation, the user may add new media vehicles are markets at will by simply specifying their attributes.

REFERENCES

Note that the references have been omitted for the sake of brevity. For a complete manuscript, please contact:

Hugh M. Cannon
Adcraft/Simons-Michelson Professor
Department of Marketing
Wayne State University
5201 Cass Avenue, Suite 300
Detroit, MI 48202-3930

(313) 577-4551(o)
(313) 577-5486(f)
hughcannon@aol.com
http://cannon.busadm.wayne.edu