DATA INTEGRATION: VALIDATING TELMAR'S MULTIBASING TECHNIQUE

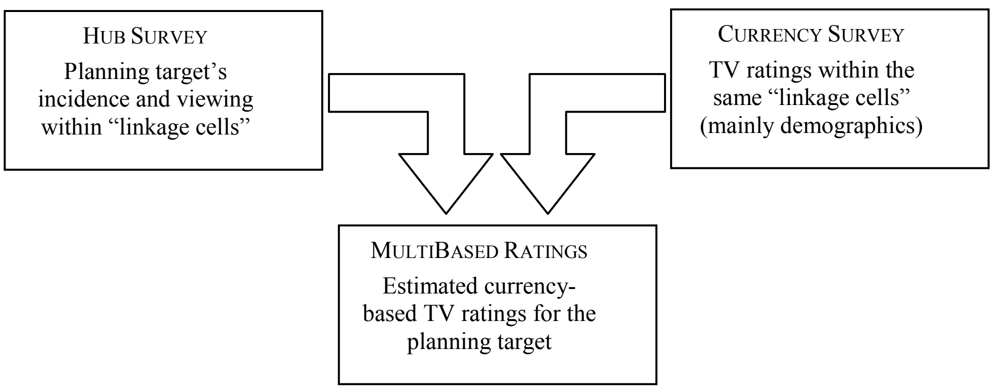
Peter Walsh, Telmar Group Inc.
Barbara Zack, PHD USA

Background

Media planners expend considerable time and effort developing sophisticated target audience definitions using syndicated consumer research such as MRI, Simmons, Mendelsohn and JD Power in the USA, custom studies, etc. Yet the television ratings "currency" that is used by media buyers allows only demographic breaks, because that's all that is collected. So, for example, a strategic planning target might be "upscale women who are independent, career-oriented, confident, and fashion-focused"; the television buying target might get translated into "women 25-49 with household income $75,000+".

The disconnection between these two realities is frustrating, to say the least. Media professionals have long sought ways to transfer the insights gleaned from consumer research onto the television ratings, so that the subtle nuances of viewing behavior for their specific target audience can be reflected in program rankings, negotiations, and buys.

To that end, various data integration techniques have been attempted over the years, ranging from Profile Matching (also called Indexing) to Fusion. Telmar has developed a new data integration technique called MultiBasing. This is used to estimate TV vehicle ratings (the "currency" we're missing) based on the planning target's profile and TV viewing behavior as reported by a broad consumer survey such as MRI. This is depicted by the schematic below:



MultiBasing is a way of inter-relating two separate surveys such as MRI and Nielsen, making use not only of demographic and other behavioral predictors (e.g. weight of viewing) but also the kinds of nuances that are related to lifestyles, interests and other psychographic factors. On this basis we believe MultiBasing provides audience estimates for planning targets that are closer to the truth than ratings for demographic targets.

Project Overview

This paper documents an evaluation of MultiBasing based on a "foldover" test. This kind of test involves splitting a survey into matched halves and using information in one half to predict certain values in the other, and then comparing these predicted values against the actuals in the second split-half.

This is analogous to one split-half being a product and media survey such as MRI (the "hub", in MultiBasing terms) and the other being a currency ratings survey such as Nielsen TV.

MRI kindly provided two replicated sub-samples from among respondents who completed both the personal and self-administered product interviews in the 2002 MRI Doublebase (i.e. unascribed data). The split was stratified so that the halves would be well-matched on Census and Claritas demographics.

An overview and descriptions of data integration techniques including MultiBasing appear in the appendix. In brief, MultiBasing involves a 2-phase procedure. In phase 1, initial estimates are obtained of currency vehicle audiences within the target group by (a) multiplying the population of each linkage cell on the currency survey by the incidence of the target group in the population of the corresponding cell on the hub survey (i.e. thus estimating the target population in each cell on the currency survey); (b) multiplying this estimate of the target population by each vehicle's true rating in the same cell on the currency survey (i.e. thus estimating the currency audience in thousands in each cell); and then (c) summing audience estimates across cells and dividing by the total target population to get an overall estimate of the vehicle's rating in the target group.

Put simply, phase 1 involves the "weighted averaging" of currency ratings, where the weight is simply the number of target group members within each cell. Hereafter this is referred to as Weighted Profile Matching, which indeed is a data integration technique in its own right. In other words, MultiBasing incorporates Weighted Profile Matching.

Phase 2 detects the presence of any residual correlation between the target group and a media vehicle that might be due to lifestyles, interests or psychographic factors. For example, there would very probably be something in common between people who are actively involved in home improvements and people who regularly read home and garden magazines, and this common factor would be "residual to" (i.e. left over after accounting for) whatever demographic similarity these two groups might have.

MultiBasing employs "surrogate" media measures on the hub survey to detect such residual relationships. In this test, the magazine screen-in question (i.e. read in the past 6 months) was treated as the surrogate measure on the hub survey for each corresponding vehicle (average issue readership) on the currency survey. That is, residual correlation between the target group and the surrogate measure on one split-half was used to predict the relationship between the target group and currency measure on the other.

A challenge in any foldover test is that unavoidably there are random differences between the split-halves. How should this be taken into account when judging the accuracy of the predicted results? We have taken the approach of comparing MultiBasing against the alternatives available to media planners, as well as evaluating absolute accuracy – in other words, absolute and relative accuracy.

Two alternatives available to media planners are simple demographic targeting – i.e. using the buying target, which is also sometimes referred to as "single cell profile matching" – and Weighted Profile Matching (i.e. the MultiBasing phase 1 calculation only), which we believe produces results that should be comparable with respondent-level fusion. Results are compared between MultiBasing and these two alternatives.

Audience estimates were generated for 204 magazine titles against each of fifteen planning targets, for each of which was identified an appropriate buying target. This generated 3,060 (i.e. 15 x 204) observations. The results are reported for all 3,060 comparisons, both overall and segmented by leverage index where mentioned in the findings.

	Planning Target	**Buying Target**
1	Heavy users of canned dog food	Household heads or homemakers
2	Decision-makers in purchase of a new SUV	Males 25-54 earning $50,000+
3	Decision-makers in purchase of a new truck	Males 25-54 earning $50,000+
4	Males 25-44 heavy/medium users of imported beer	Males 25-44
5	People using 2 or investment services in past year	People with HHI $75,000+
6	Buyers of homemaker magazines in past month	Female homemakers 25-44
7	Males play golf 1+ times/month	Males 25-74 with HHI $50,000+
8	Took 2+ international business trips in past year	Professionals, managers, self-employed
9	Digital camera owners	Adults 18-34
10	Attend movies 2+ times per month	Adults 18-34
11	Watch Discovery Channel 2+ times per week	Parents
12	Decision maker home PC purchase	Adults 18-34
13	Exercise 2+ times per week	Adults 18+
14	Primary users of Kentucky Fried Chicken	Adults 18+
15	Homemakers who spent $2500+ on remodeling	Homemakers 25-54

Table 1: Fifteen planning and buying targets used for validation test

Findings

An example of results from the first target (heavy users of canned dog food) is shown in Table 2 below, for a subset of all magazines (20 magazines are shown of the 204 that were analyzed).

- Column A is the "true" planning target audience in the currency split-half. This is the unknown to be estimated.
- Column B shows ratings for the Demo Target in the same currency split-half. This is what would be used to buy media against the planning target in lieu of data integration.
- Column C shows the audience estimate produced by MultiBasing.
- Column D shows estimates based on Weighted Profile Matching (i.e. phase 1 of MultiBasing).

	(A)	(B)	(C)	(D)
	True Planning Target Audience (currency split-half)	Demo (i.e. buying) Target Audience (currency split-half)	MultiBasing Estimate of Planning Target Audience	WPM Estimate of Planning Target Audience
Allure	2.1%	2.2%	2.7%	2.6%
American Baby	3.2%	4.0%	4.0%	4.0%
American Hunter	2.3%	2.0%	2.1%	2.5%
American Legion	1.2%	1.6%	1.7%	1.6%
American Photo	0.2%	0.9%	0.7%	1.1%
American Rifleman	4.8%	2.7%	3.0%	3.4%
American Way	1.3%	0.6%	1.0%	0.7%
Architectural Digest	1.2%	2.6%	1.9%	2.9%
Arthritis Today	2.5%	1.3%	1.3%	1.5%
Atlantic Monthly	0.7%	0.6%	0.1%	0.5%
Attache	0.9%	0.8%	0.3%	0.9%
Audubon	1.3%	0.8%	1.0%	0.9%
Automobile	1.1%	2.2%	2.7%	2.3%
Baby Talk	1.4%	3.1%	3.6%	3.0%
Backpacker	0.7%	0.6%	0.6%	0.6%
Barron's	0.4%	0.6%	0.5%	0.5%
Baseball Weekly	1.1%	1.0%	1.2%	1.1%
Bassmaster	2.4%	2.0%	2.3%	2.6%
BHG	23.0%	20.6%	24.4%	22.7%
Bicycling	1.6%	0.9%	1.0%	1.0%

Table 2: Sample results for 20 of the 204 magazines, for one planning target (heavy users of canned dog food) and an appropriate buying target (household heads or homemakers)

So, which estimate gets us closest to the "truth" – MultiBasing, Demo Targeting (typically used by buyers), or Weighted Profile Matching? In order to evaluate the relative "goodness of fit", we first compared the methods in these three different ways:

1. After ranking the audience estimates from each of the three methods, which shows a higher correlation to the rank order of the currency audience estimates?

2. In an absolute sense, which of the three methods provided an estimate that was closest to the unknown "true" currency audience?

3. How often did each method provide an estimate that was not significantly different (higher or lower, at 95% confidence or "2 sigma tolerance") than the actual currency audience estimate?

	Correlation of MB rank to actual	Correlation of Demo rank to actual	Correlation of WP rank to actual	MultiBasing was closest to actual	Demo rating was closest to actual	Wtd Profile was closest to actual	MB not significantly different from acutal	Demo not significantly different from actual	WP not significantly different from actual
Heavy users of canned dog food	88	88	**91**	**37**	28	34	78	**79**	**79**
Decision-makers in purchase of a new SUV	86	49	**89**	**43**	16	41	**72**	32	**72**
Decision-makers in purchase of a new truck	87	59	**90**	35	22	**43**	77	44	**84**
Males 25-44 heavy/medium users of imported beer	**83**	82	82	**39**	**39**	22	63	**72**	66
People using 2 or investment services in past year	**88**	82	86	**44**	26	30	**77**	66	69
Buyers of homemaker magazines in past month	**98**	91	97	**46**	18	36	**79**	47	69
Males play golf 1+ times/month	**89**	86	87	**50**	25	25	**70**	57	56
Took 2+ international business trips in past year	**81**	64	76	**39**	30	31	64	58	**70**
Digital camera owners	**89**	70	88	**46**	15	39	**78**	39	77
Attend movies 2+ times per month	**97**	90	96	**40**	24	36	71	59	**74**
Watch Discovery Channel 2+ times per week	**98**	91	**98**	**48**	16	36	**80**	48	79
Decision maker home PC purchase	**95**	77	94	**43**	15	42	79	38	**83**
Exercise 2+ times per week	**99**	97	98	**59**	13	28	**92**	42	80
Primary users of Kentucky Fried Chicken	92	89	**95**	29	**40**	30	77	75	**85**
Homemakers who spent $1500+ on remodelling	91	86	**92**	**38**	32	30	69	63	**77**

Table 3: Comparison of results by target

In terms of preserving the rank correlation between the estimated and actual vehicle ratings, Weighted Profile Matching and MultiBasing both performed well. In an absolute sense, MultiBasing gets us closer to the correct answer more often than Weighted Profile Matching and Demo Targeting. In terms of statistically significant differences however, MultiBasing and Weighted Profile Matching perform comparably well overall.

We know that for many consumer targets, some media vehicles will show an extremely strong positive or negative skew (e.g. golfers are more likely than anyone else to read golf magazines; teenage boys are extremely unlikely to read Mature Outlook). So which method most reliably identifies the strongest positive and negative skews?

For this purpose a "leverage index" was calculated for each vehicle (i.e. target rating ÷ all 18+ rating x 100) for the true currency target audience. The following table summarizes the results across all 3,060 observations by leverage index, showing the percentage of cases where the estimate was ***not*** significantly different from the actual rating. For example, in 62% of all cases with a leverage index of 350+, MultiBasing provided an acceptably accurate estimate whereas Demo Targeting did so in only 19% of those cases.

Leverage Indices	Number of Cases	**Demo Targeting**	**MultiBasing**	**Weighted Profile**
350+	53	19%	**62%**	21%
250-350	106	41%	**67%**	42%
150-250	530	56%	75%	**76%**
125-150	469	50%	83%	**85%**
75-125	1,252	60%	81%	**86%**
0-75	650	44%	**56%**	48%
Total	3,060	53%	**74%**	73%

Table 4: Comparison of "correct" results (within 2 sigma of actual), grouped by leverage index

Conclusions

- All the methods perform better in the mid-ranges than at the extremes.
- MultiBasing is much more likely than the other methods to correctly identify vehicles with very high (index 250+) or very low leverage (index <75).
- The phase 2 MultiBasing calculation adds nothing to the accuracy of Weighted Profile Matching in the mid-range – i.e. for vehicles that index between 75 and 125.
- Both MultiBasing and Weighted Profile Matching, as expected, outperform Demo Targeting by wide margins.

This analysis shows that MultiBasing (specifically the phase 2 calculation) provides a worthwhile improvement upon Weighted Profile Matching, and that both of these are far better than using a Buying Target alone. MultiBasing's overall accuracy was creditable when we allow for the fact that random differences between the split-halves created a considerable degree of fluctuation in the currency ratings.

APPENDIX: Overview of Data Integration Techniques

This section summarizes the main features of the various data integration techniques. Essentially, all data integration techniques attempt to achieve the same goal. For a target audience selected on a consumer survey – i.e. the users of a particular product – the goal is to predict the audience levels *that would have been measured* had the consumer survey target definition question(s) been asked on a 'currency' media survey such as a TV ratings panel.

For simplicity the following discussion is limited to the integration of TV ratings with product data. Of course the same techniques are used with other media data such as radio ratings. The techniques briefly summarized below include the following:

- Profile Matching
- Clustering
- Fusion (various forms)
- Calibration
- MultiBasing.

Profile Matching

Profile matching is the oldest data integration technique, having been practiced in effect for as long as the concept of an advertising target audience has existed. There are two forms of profile matching: single cell and weighted.

In the single cell approach, the first step is to identify the demographic group within which the incidence (or volume) of usage is greatest. The demographics of interest here are the same ones for which TV ratings are available – e.g. age-groups by gender.

So, if usage of the product is greatest among (say) women 18 to 29, then women 18-29 is defined as the target audience. Single cell profile matching is as simple as that – i.e. it is demographic targeting.

In weighted profile matching, product usage is analyzed to determine how much of the total each demographic accounts for. E.g. women 18-29 account for 30%, women 30-39 for 25%, and so on. It is usually found that the peak usage group accounts for only part of total usage – usually less than half – and so it makes sense to base media planning on *all* users rather than just one group.

Next, weighted average ratings are calculated for all the demographics. E.g. if a program has a 2% rating among women 18-29 and 3% among women 30-39 then its weighted average will fall somewhere between 2% and 3% and this will depend on relatively how much of total usage each group accounts for. The weighted averages therefore represent estimated ratings among all users (or total usage volume, if each demographic is weighted by that).

Profile matching does not attempt to determine whether special correlations exist between the target audience and any media vehicles. For example, people in the target audience for a gardening products (i.e. active gardeners) are more likely to watch gardening shows on TV than other members of the same demographic groups. This interest-based correlation cannot be accounted for by profile matching.

Therefore, rather than taking the results at face value, media planners should always consider whether there might be any special correlation (either positive or negative) between the target audience and any 'candidate' media vehicles. Such correlations can exist because of people's interests (as with the gardening example), lifestyles, or underlying psychological factors.

Clustering

Cluster analysis of a ratings survey can be used to develop a viewer typology. For example, one cluster might be comprised of people whose viewing patterns are characterized by heavy viewing of sports programs, another by heavy viewing of daytime soaps, and so on.

The viewer typology can be replicated on a product survey if suitable TV questions are asked (e.g. broad viewing patterns). This provides a basis for ascribing ratings from one survey to the other. Members of (say) a heavy sports viewing cluster on the product survey are assumed to have the same average program and daypart ratings as members of the same cluster in the TV panel.

On that basis, the product survey respondents are ascribed the same program viewing probabilities as their counterparts on the TV survey. These probabilities can then be tabulated for any target audience selected on the product survey.

Because clusters are based on viewing interests, this technique is able to preserve correlations of the sort described earlier – to some extent, at least. Whether it captures the example of gardeners being more likely to watch gardening programs depends on whether a gardening viewer-type is statistically significant enough to emerge from cluster analysis. However, particular clusters can be 'forced' to emerge.

Respondent-Level Fusion

It can be seen that profile matching and clustering work at a group level – i.e. demographic groups and viewer typologies respectively. In contrast, fusion works at the individual respondent level.

The idea is to match every product survey respondent to a TV survey respondent and to then ascribe the measured product usage behavior from the former to the latter. Alternatively, the measured *viewing* behaviour can be ascribed in the opposite direction. Either way, the two survey records are fused as if the two individuals were the same person.

Individuals are matched on whatever variables – primarily demographics – are common to the two surveys. First, a decision is made as to which demographics are mandatory matches (e.g. males can only ever be matched with males and females with females) and which ones will instead be used to construct a measure of similarity.

Suppose the mandatory matches are on gender, age-group, household income and region. Now, both surveys contain some number of males aged 18-29 living in middle-income households in the west. So which particular individuals from these two sets should be matched?

The demographic similarity of any two individuals is expressed statistically in terms of a 'distance measure'. Weights can be used to represent how much influence any particular demographic difference tends to have upon people's behavior – e.g. two people might be the same in many other respects but differ in their educational levels, so how important is that?

In this way, the likelihood of any two individuals being matched depends on how similar they are. Ideally every survey respondent should be matched with the most similar respondent it is possible to find on the other survey, but constraints often require compromises in this regard. For example, obviously it is desirable that the ascribed media ratings or product usage levels (depending on the direction of ascription) should be identical to those in the original survey, and to achieve this it might be necessary to allow some sub-optimal matching.

As with clustering, the question arises as to how accurately the linkage structure (i.e. matching) predicts TV viewing patterns within any target group that might be selected. To use the earlier example once again, does the linkage structure ensure that the estimated ratings for TV gardening programs are accurate for active gardeners? Or that the ratings estimated for TV programs on (say) vacations and travel are accurate for people planning overseas trips?

This problem is addressed by fusion-on-the-fly, where the weights used in the distance measure are set so as to reflect correlations between matching variables and the target audience. In other words, respondents are matched only once the target audience is selected.

The extent to which fusion-on-the-fly improves the accurac depends on whether and the extent to which the matching variables are in fact correlated with the target audience. In other words, if the product usage in question is not very well predicted by demographic characteristics then fusion-on-the-fly cannot improve upon a general purpose fusion.

Calibration

If the product survey includes questions about *all* the vehicles (e.g. TV programs) that are required for media planning, then their audience levels can be calibrated to match the true ratings as reported by a 'currency' media survey. Calibration is usually performed separately within demographic cells (e.g. age-groups of males and females) so that the calibrated ratings are identical to the 'currency' for major demographic target audiences as well as overall.

A limitation of this technique (in addition to the necessity of measuring all media vehicles on both surveys) is that non-random duplications *between* calibrated vehicles are not necessarily preserved, which can have an effect upon the results of reach/frequency calculations.

MultiBasing

Profile matching and respondent-level fusion rely primarily on demographics to predict media and/or product usage. This is ironic because one of the main reasons for data integration is that demographics are inadequate as predictors of consumer behavior in today's society.

By fusing respondents as if they were the same person, based on demographic matching, fusion can close off the opportunity to preserve statistical relationships that actually exist. Unavoidably, any real life correlation that happens to be independent of the matching variables will be diminished or even completely obliterated by the fusion process.

MultiBasing has been designed to solve that problem. It enables a 'hub and spoke' media research architecture. Media 'currency' databases are kept separate while the hub survey collects consumer profile data (product use, lifestyle, psychographics, etc.) and 'surrogate' media measures (described later). For linkage purposes, all surveys must carry a set of standard questions including demographics. The databases can then be interrelated statistically by means of the calculations described below.

1. Linkage Cells

Suppose a product is used more by women than by men – perhaps a cosmetic. This implies that TV ratings among users should be closer to the ratings among women than among men. And if the product is used more by younger women then the ratings among users should be closer to the ratings among (say) women 18-39 than among women 40+. As explained in relation to weighted profile matching, the product user ratings should fall somewhere in between – i.e. in effect they would be weighted averages of the ratings among these two age-groups of women.

With suitable analytical tools, a research analyst could continue to 'drill down' in this fashion, taking into account occupation, education, income and other variables.

But let's go back to the beginning. Clearly, the fact that it is a predominantly female product is important information. But if a TV program's ratings among men and women are nearly the same then gender contributes little or nothing towards an accurate estimate of that program's rating among users. Likewise, if a program's ratings are nearly the same among younger and older women then little purpose is served by taking age-group into account at this point.

On the other hand, suppose education is a strong predictor of TV ratings for certain programs, and suppose further that the product tends to be used by women with higher education. That being so, failure to take education into account could very well result in inaccurate ratings estimates. Therefore it is important to 'drill down' using variables that differentiate TV viewing patterns.

TV ratings survey data is analysed to identify population groups ('linkage cells') in terms of common variables that are found to be predictive of viewing patterns. The variables are not limited to demographics and can include broad media availability and usage measures such as access to cable channels, total weekly hours of television viewing, or viewer typologies. But they must be common to both surveys.

Based on the results of the analysis, mutually exclusive linkage cells are defined by unique combinations of codes on the variables employed. The final linkage structure consists of typically more than 100 cells, depending on sample-sizes. The cells are coded into both the TV ratings survey and the product survey to make ready for cross-tabulation.

An example of a linkage cell is males 25-34, tertiary education, single or married without children, income above $50,000 and light TV viewers.

2. Phase 1 Calculations

MultiBasing's phase 1 calculations are the same as weighted profile matching. That is, the relative contribution of any linkage cell to the ratings estimates for a product target group depends on the relative number of users (or usage volume) in that cell and the ratings that apply. For example, because few men use cosmetics, their ratings carry relatively little weight.

The level of usage in each linkage cell is determined by tabbing product usage against the linkage cells on the product survey.

Next, TV ratings are tabbed against the linkage cells on the media survey.

An initial estimate of a media vehicle's rating among users is now obtained as a weighted average of its currency ratings across linkage cells.

MultiBasing's phase 1 estimates can be accurate in their own right. As with other data integration techniques, this depends on the extent to which ratings among product users are mediated by the linkage structure.

However, within any particular cell there will of course be some people who are more likely than others to watch particular TV shows, listen to particular radio stations and so on. They do so because they are more interested in those media. Perhaps the media vehicle is especially relevant to their lifestyle. Or perhaps there are underlying psychological factors that have resulted in a special correlation existing between usage of the product and viewing of the TV program.

Once again, consider gardeners as a target group. The prime demographic is women 45-64. But of course, not all women 45-64 are interested in gardening. Among those who are, viewing of gardening programs is naturally higher than among the rest. Therefore, although the analysis may have isolated women 45-64 as a linkage cell, it cannot fully have captured the relationship between an interest in gardening and the viewing of gardening programs. So how can it be ensured that real life relationships like this – of which there are endless examples – are fully preserved?

3. Media Surrogates

Media surrogates on the hub survey are shorthand measures for vehicles whose ratings are imported from a currency media measurement system. The use of surrogates is the single most important distinguishing feature of MultiBasing.

Surrogate media measures can be as simple as Cable channels watched for one or more hours per week. Analysis will indicate whether any such question is adequate to capture important real life relationships. Consider the following:

Media Leverage Indices *(relative to all 12+)*	Frequent Gardeners *(40+ days/year)*
Watch HGTV *(1+ hours/week)*	206
Magazines *(average issue)* –	
Canadian Gardening	176
Gardening Life	191
Plant & Garden	176

Source: PMB 2001

This table shows two things. First, people who spend a lot of time gardening – and therefore are more likely to respond to advertising for gardening products – are also more likely to watch gardening programmes on television. This is hardly surprising.

More to the point, it shows that a simple surrogate measure for cable television can be sensitive enough to establish that HGTV has strong leverage with frequent gardeners – in fact, slightly more than gardening magazines. This is not an isolated result; MRI Spring 2001 reveals the same thing.

In short, surrogate media measures enable a hub survey to capture statistical relationships that are not mediated by the linkage structure. Such relationships are of crucial commercial importance especially to specialty television and special interest magazines, not to mention advertisers and media planners.

Other techniques do not fully preserve such relationships because the data integration is based on a limited set of (mainly) demographics. MultiBasing on the other hand employs surrogate media measures on the hub survey to detect and preserve these all-important relationships.

4. Phase 2 Calculations

Whether the phase 1 calculations fully account for a vehicle's leverage in a target group is determined in the following way. If the phase 1 calculations are replicated on the hub survey, using the surrogate measure *instead of* the true ratings, then the result obtained should be equal to the known surrogate-based audience in the target group.

The vehicle's surrogate-based audience in the target group is known because both questions – i.e. product usage and the surrogate media measure – are asked on the hub survey.

So, the surrogate measure is tabbed against the linkage cells, and a phase 1 estimate is calculated in the same 'cellwise' fashion as before. But now the surrogate is tabbed directly against usage of the product, thus providing an observed result against which this expected result is compared.

Note that the surrogate measure is used only to obtain these observed and expected results – i.e. MultiBasing does not report out the surrogate measures themselves (except in the auditing process described shortly).

If the surrogate audience proves to be higher than expected then clearly there is a positive relationship between usage of the product and the vehicle in question – e.g. the kind of relationship that would be expected between gardening and watching HGTV. But of course, the correlation can also be negative.

This relationship is expressed as a Residual Leverage Index. An RLI of (say) 150 means the observed surrogate audience is 50% higher than as predicted by the linkage structure. An RLI of 80 means 20% lower than expected.

If the surrogate measure is higher or lower than expected for the target group then the currency-based rating should reflect this fact. That is, the initial estimate from the phase 1 calculations should be adjusted accordingly.

The adjustment factor is approximately equal to the RLI – e.g. an index of 150 implies an adjustment factor of approximately 1.5. However, it is not exactly equal to the RLI for the following reason.

For any target group, application of the adjustment factor yields the final audience estimate. Suppose we now redefine the target group as the balance of the population – i.e. everyone other than the former target. We repeat the whole procedure and again apply whatever adjustment factor is indicated by the RLI. If the result is subtracted from the vehicle's total audience, we must get the same final audience estimate as before. In other words, the adjustment formula must be 'transitive'.

The transitive adjustment formula employed in MultiBasing is based on tetrachoric correlation.

5. Audit Report

The calculations described above can be audited by means of a special MultiBasing report that sets out the phase 1 and 2 estimates so that it can be seen exactly how any result has been created.

MultiBasing is not a 'black box' model – in principle a research analyst can perform the same calculations in a spreadsheet.