MODELLING IN A MULTI-MEDIA ENVIRONMENT

Peter Masson and Paul Sumner, Bucknull & Masson

INTRODUCTION

The purpose of media behaviour models is to provide the media planner with a decision making tool. We have had models for several decades that work at the simplest level to support questions like 'should I buy title A or title B', or more complex 'should I buy this combination C of newspapers/insertions or combination D of newspapers/insertions'. But the complexity of the model compounds at the multi-media level in supporting questions like 'should I buy this combination E (TV channels/spots) or combination F (magazines/insertions) or 'should I mix media by taking a percentage out of combination E (TV) and placing it in magazines and if so what percentage'?

The first step in modelling media behaviour is normally to provide the planner with comparative statistics on the reach and frequency (OTS) of selected schedules. In the case of multi-media evaluation this becomes a multi-dimensional matrix with as many dimensions as there are media. From the matrix we will know what proportion of the target receive only TV OTS, what proportion receive 3 print OTS, 2 TV and 6 radio OTS and so on. The interpretation of such a matrix is no longer the simple choice of the schedule that delivers the most at (3+) OTS.

The second and essential step therefore is to evaluate the 'effect' of those given (multi-dimensional) 'doses' of advertising mapped out in the OTS analysis. Without the second step the planner has limited means to judge between the extremely large and complex OTS distribution maps of different schedules. The issues and analysis mechanisms involved have been described in previous papers[1] where TIMING of the advertisement exposure is shown to be a key determinant of 'effect'(the rate and maintenance of 'learning' and the level of synergy effects between media). Our OTS analysis we will need therefore not only to evaluate the total quantity of OTS delivered medium by medium but also the TIMING of those exposures (day by day) if we are to create a realistic model.

In such a complex effect environment it is vitally important that the first multi-media schedule estimates, in terms of the distribution of OTS, are performed as accurately as possible and account for differences in the definition of OTS inherent in the research for each media type. This paper will focus on the issues involved in this area.

Nowadays very few commercially available models use 'aggregate' or 'formula' models for schedule OTS evaluation; the 'individual informant' approach is almost universal. But most models of schedule performance are very simple, and simplistic, Binomial Expansion models.

This model is fairly adequate for Press, though even here the accumulation of the reach of daily newspapers across the week and Parent/Supplement relationships are not correctly represented by the Binomial's independence assumption. As we look further at providing a timed response model, it is becoming increasingly clear that even in Press the independence assumption in the Binomial leads to week-on-week estimates which are not as accurate or realistic as we would like.

We believe that the independence assumption is quite simply *wrong* for TV and Radio. Nobody watches two different TV channels at the same time, though the independence model implies exactly that. Cinema, Internet and Posters also need careful examination of the independence assumptions. It is clear, for example, that a Poisson distribution model (which deals better with a large number of very small probabilities) is more appropriate for the Internet, and conditional probability changes must be made to the basic binomial for Cinema and Posters.

PRESS

Today there is one model which has gained total dominance for the production of Net Reach and Gross Reach statistics for Press schedules: this is the individual informant Binomial Convolution model. This model has been in place for more than 30 years now.

[1] WRRS Venice 2001: Time related measures of relative 'effect' - an essential ingredient for Multi-Media campaign evaluation', by Peter Masson, Bucknull and Masson, London and Peter Callius, Research International, Stockholm

It is obviously a great advantage to have a universal model. The down side of this is that there is a tendency to regard the model as 'correct'; to its adoption as a 'reference' standard. Perhaps only now that computing is so cheap can we properly examine the foundations of our 'universal' model. It must be recognised that it is a 'model': it is not reality.

The basic assumption of the Binomial Convolution (after just Binomial) is the *independence of probabilities*. The chance of reading one publication is unaffected by the chance of reading any other. It is easy to adduce examples of individual behaviour which are very obviously *not* independent. However, over all informants, the independence assumption is assumed to hold fairly or very well. Nor is there any clearly better assumption to use in the general *Press* case.

However there are (at least) two situations where the independence assumption is violated for Press. These are Parents/Supplements and (so called) Daily Reach of Dailies[2]. If, for example, a parent and a supplement are both attributed 0.5 reading probabilities, it is much more likely that both are read together than that they are read independently. It is sensible for any model, including the Binomial, to adjust its assumptions in these cases.

There is a clearly better conditional probability treatment for both the above situations. However these assumptions, while definitely better than independence, are far from perfect. To improve on the model fit we need more information from the Readership Survey which clarifies issue-on-issue or week-on-week dependence. Probability attribution methods too need careful attention.

We have so far been discussing the press schedule Net Reach (and Gross Reach) only. When we turn to the Frequency Distribution there is not the same consensus as exists with Net. Calculating the full Frequency Distribution for a medium to large schedule requires billions of calculations for its full and accurate assessment. Few if any commercial programs include this full calculation as routine; it takes too long. Various (commercially sensitive) approximations are used.

These approximation methods can, even when we are discussing Press only, have a big impact on the shape of the Frequency Distribution produced. When we turn to automatic Press Schedule Building (AutoBuilding; also and *wrongly* known as Optimisation) the shape of the Frequency Distribution becomes a key factor in the schedules produced. A clearly poor Frequency Distribution approximation is a very unreliable basis for AutoBuilding. Approximation methods which cluster the Frequency Distribution too much, such as concentrating all of the Frequency Distribution at the quartiles of the distribution, *are* clearly poor.

Now the predictive accuracy of a model is not only a function of the model assumptions but also the nature and the accuracy of the data on which it is based. There needs to be a much greater interaction between modeller and researcher than currently exists.

Probability attribution is the key. This is dependent on the nature of the reading frequency scale. While an enormous amount of attention has been given over the last 3 decades to the recent reading scale and model very little has been given to the reading frequency scale. Its ability to distinguish different reading frequency groups, especially amongst the high frequency groups and the low frequency groups is very important to the final shape of the FD (and hence the outcome of any automated scheduling procedure). The UK NRS reading frequency scale is very 'blunt' in this respect in comparison to the Swedish Orvesto Consumer reading frequency scale. The latter (although not calibrated to RR) is able to produce much greater discrimination while being easily understood and completed by the respondent. The gap in the probability distribution in the UK is very evident in the following table:

Probability attribution

UK NRS		**Always**			**Quite often**	**Only occasionally**	**Not last year**
Daily X							
Probability		91.5			28.3	5.4	0
Profile %		9.8			5.0	15.4	69.8
Sweden OC	**Read all**	**Nearly all**	**3 of 4**	**2 of 4**	**1 of 4**	**Nearly none**	**Never**
Daily Y							
Probability	100	92.5	75	50	25	7.5	0
Profile %	6.0	3.8	1.4	1.8	4.8	21.4	60.8

[2] ESOMAR/ARF Cannes 2002: Measuring the 'daily reach of dailies' and newspaper sections. Data modelling for realistic reach and frequency evaluation of schedules. Ingemar Lindberg, Paul Sumner and Peter Masson

Modifying the basic press audience measurement

Print research normally measures the exposure to ***an issue*** of a publication (sometimes a section of it) but not to ***an advertisement*** within it (the advertisement audience). In a multi-media comparison, where (TV) media are measured close to the advertisement exposure level, it is important that the press OTS should be 'modified' to the same level. This advertisement audience level can be expressed as the proportion of the issue audience that is likely to pass an average page (from page traffic studies) or claims of the proportion of the issue read. This proportion, expressed as a probability can be applied within the binomial calculation.

However this probability data is seldom available 'single source' at the individual level and tends to be applied as an overall aggregate factor. This will lead to a poor estimate of the effect on advertisement reach and schedule audience profile. In the first instance we know that frequent readers spend more time reading and claim to read higher proportions of the issue, yet such differences are ignored when applied as an overall or aggregate factor.

Suppose a segmentation of our page traffic data showed the following differences of average page traffic score between frequency of reading groups:

Average % of pages passed	All	Nearly all	3/4	2/4	1/4	Nearly none	Average overall
Title A	.9	.8	.7	.6	.5	.3	.74
Title B	.9	.8	.7	.6	.5	.3	.69

We then apply these page traffic scores differentially within reading frequency group and as one overall aggregate factor. (6 insertions in title A and title B). The impact on the R & F results is shown below:

	Net Reach %	Gross Reach %	Profile (Net % earn 350K+
Page traffic score applied within frequency	23.5	92.0	17.4
One aggregate page traffic score overall	28.5	92.0	16.2

While the gross reach is unaffected the net reach, when the page traffic scores are applied within frequency, is significantly lower and the schedule net profile (e.g. by personal income) changes (improves), those earning SEK350K+ increase from 16.2% to 17.4%.

But advertisement audience estimation is far more complex than this. The editorial of newspapers and magazines is very often sectionalized. Readers will go to favourite sections by 'appointment' (in the same way that people chose to be around to view favourite programmes). The nature and the interests of the reader will play a large part in his or her own individual page selection. We really need the 'single source' link between the target person and the reading of the specific sections of the publication. (These issues are explored in far greater depth by Peter Callius in his paper at this symposium entitled 'The opening of Pandora's box')

This factor (proportion likely to pass the advertisement page) is one of the elements of an MPX/PEX score. On the one hand the ***proportion of issues*** read ***reduces*** the print issue audience (to the advertisement audience) but on the other the ***repeat pick ups*** of the same issue will ***increase*** the OTS at the advertisement level. In a multi-media evaluation this should be taken into account – since a TV or radio spot offers only one exposure possibility.

The same comments apply to 'multiple pick ups' as to 'proportion of issue read'. In the absence of single source data with multiple pick ups reported at the individual level, this multiple pick up data would also be better applied within reading frequency groups since frequent readers are more likely to have higher levels of multi-pick up than infrequent ones, but again this not a complete solution.

The gross effect of say 3 pick ups (times parallel reading[3]) is trivial to model (gross contacts increase 3 fold), but the net effect needs careful consideration. First the advertisement net reach can never exceed the issue net reach. But should it, once reduced by the proportion of issue read, increase again as a result of a second or subsequent pick up and if so at what rate? This depends on the question asked in the research. For example in the UK QRS the proportion of issue read question is asked specifically of the last of any issues read on the last day of the reading days claimed during the last issue period. The proportion so claimed is then applied to ***all*** the claimed reading occasions (as is the number of different (Parallel) issues read) on any other day such that it represents the total proportion of the issue read. However even if this was the correct proportion of pages read on each occasion respondents will not necessarily (indeed are unlikely) to read the ***same*** pages on each pick up. The net advertisement audience should increase with pick up to pick up. The independence assumption is unlikely to apply. Readers will generally not re-read the pages that they have already read but of course may pass by some of the same to get to new ones.

[3] While it not the purpose of this paper to critique the PEX methodology we do not concur that the inclusion of the parallel reading factor is justified. A number of the average issue claims in the period arise from replicated reading. Of course replicated reading (itself a further pick up in a later issue period) gives rise to further pick ups which in themselves can legitimately be included in an average issue period claim ***so long as they are being counter balanced by parallel reading in the period.*** To count parallel reading again as further OTS completely negates the approximate balancing effect of the 'errors' in the average issue model.

In addition, if the question is answered as posed, it is likely to lead to an underestimate of the start point for the net advertisement audience accumulation. One might reasonably suppose that the 'proportion of pages opened' claim will decrease after the first or second pick up and that the application of the score for the last occasion overall will lead to an underestimate. In practice this may not be serious, as we suspect that a high proportion of respondents will generalize and answer for the proportion of pages that they usually read ***in total*** combining all reading occasions. This poses a problem for modelling as we do not know precisely the relationship of the last 'proportion of issue read' claim for the last pick up to earlier pick ups or in which way the question has been answered.

In effect a print insertion is like a mini schedule. It does not deliver a 'batch' of net OTS on the one day (or even the month) where the X was put on the schedule. There is a (different) time span over which readers come to a copy of each title and then read and perhaps re-read over a period longer than the issue period. Increasingly studies will become available which, using diary type (repeat interview) techniques will monitor the pick up and re-reading of specific issues over time. This will 'time' the repeat pick ups of MPX/PEX and extend them beyond the issue period which means that the data becomes related to specific issues and will create a conflict with the traditional average issue period results.

Few Press evaluations take account of the *timing* of OTS. This is another area which will need attention, especially in the context of multi-media modelling. For example, successive issues of a weekly magazine are likely to be more dependent than successive issues of a monthly or bi-monthly. The basic Binomial model will need considerable modification if it is to deal satisfactorily with the complexities of timing. It is possible, indeed likely, that a new model (possibly a VDiary model, see below) will deal better with these difficulties.

TV and Radio

We will refer below only to TV. The modelling of Radio is very similar to TV. The main difference lies in the interpretation and attribution of probabilities, not the model used.

All so-called Probability models, both aggregate and individual, have at their heart an independence assumption. The problem with using these methods (Binomial Convolution is the commonest) is that the probabilities for TV programmes or spots are not remotely independent. One can just about defend the independence assumption when dealing with a single station. The estimates of Reach & Frequency are not egregiously wrong. But when dealing with a combination of stations then the standard probability methods fail. A very simple example should suffice to illustrate this failure.

Suppose we have an informant, informant weight 1000, who has a particular spot probability of 0.5 for station A and 0.5 for station B. Then under the independence assumption 250 weighted informants will see both stations (at the same time!), 500 weighted informants will see just one of the stations and the balance 250 informants will see neither station. The correct picture is that 500 weighted informants see Station A and 500 Station B. Such absurdities with probability methods occur whether we are discussing spot probabilities, or quarter-hour probabilities, or probabilities over a longer period. Probability methods applied to TV are internally consistent but fail in their applicability to real situations.

A much better way of dealing with TV survey data is the use of reference curves. However this requires that there is some external (to the main survey) reference survey. This reference survey is usually a PeopleMeter survey but may be a Diary survey. This reference survey is used to calibrate the performance of schedules based on the main survey. However, there is very often no suitable reference survey for the small audience stations and the multi-country surveys with which we need to deal. Nor can specific target market definitions be catered for.

To overcome these issues we have developed a Television (and MultiMedia) evaluation method which is based on the use of a Virtual Diary (VDiary). We can no more than sketch this methodology here.

PeopleMeter surveys, or an actual diary, provide schedule Reach & Frequency estimates without the need for any assumptions about mathematical probability distributions. It is simply a counting exercise. (We ignore complications due to inadequate sample size or panel response failures.) Similarly, with our VDiary the production of schedule statistics is also, for the most part, a counting exercise. Schedules over a longer period than the PeterMeter or diary panel need some calculation mechanism. This applies too to our VDiary estimates. This extension beyond the panel or VDiary is a very complex technical exercise and to detail our solution here would obscure the main points of the discussion.

The construction of a VDiary is not a standard specified exercise. Every survey we have dealt with has needed some individual adaptation to produce a good and efficient solution. For every surveyed individual we produce a detailed record *as if* they had completed an actual diary. This detailed record is so constructed that over all informants the accumulated results match exactly the TV statistics we *do have* on the survey. These statistics are usually confined to the weekly and monthly reach and 'yesterday' reach of each station, though they may be as detailed as day part or ¼ hour reach.

With this information we can create VDiary records such that all general statistics are matched exactly. Further the detail of each diary record is as if an *actual* diary had been filled in. The difference between a VDiary record and an actual diary is that no VDiary record is 'extreme'. In an actual diary one or more individuals may have very anomalous patterns of viewing behaviour, while being consistent with the survey aggregate behaviour. In a VDiary such anomalous patterns are not permitted.

The VDiary so constructed can be counted in exactly the same way as a normal diary. It produces realistic estimates of Reach & Frequency which are not subject to the artificial smoothing induced by the binomial model. If the user wishes to specify, say, 4 spots in three hours, rather than exactly specified spots, then our routines produce an average result from 10 different random allocations of the spots.

A further refinement to the VDiary can be achieved if we have external (to the main survey) information from a PeopleMeter or Diary panel, for some or all of the stations covered. In such cases the constructed VDiary can be *calibrated* to the external data. This ensures that we can match exactly PeopleMeter figures for the total market to which we calibrate, and provide excellent fits to most sub-markets.

This brief overview of the VDiary approach only scratches the surface of a very flexible and accurate method. It does not, unlike probability methods, make any erroneous assumptions about the distribution of viewing probabilities. Finally, the method lends itself to multi-media modelling without introducing further assumptions about the interaction between the media types.

Because we can construct a VDiary using relatively few questions, it becomes possible to use a single source survey to measure several different media at one time and in doing so enables us to preserve well intra- and inter-media duplications.

Other Media

Press and TV have dominated, understandably, the discussion of media modelling. The Press model has been (largely but not totally) the base for model development in other areas such as Posters, Cinema and Internet. Unfortunately the Press Binomial model is a poor base for this development. Posters need a conditional probability adjustment to the basic Press model. The Cinema and Internet are better dealt with by a mixture of Binomial (for regular Internet users and cinema goers) and Poisson (for the irregular majority) models. However we are now investigating the use of VDiaries for all of these areas.

Posters

We have developed a reach and frequency model for the evaluation of poster campaigns and that also works within our multi-media evaluation system. It is different from most available models which focus on counting or estimating coverage from the basis of the poster site (site centric). Instead it is individual informant based and provides for the interaction between the individual (and all his characteristics) using different means (foot, bike, car driver, car passengers, bus etc.) to pass sites of different visibility characteristics (size, position, revolving panels, illumination) and from this calculate R&F.

We can model this on the basis of a complete site database and a traffic survey based on a sample of individuals (pen and paper or GPS). The model is ideal for closed geography situations (subway, town bus shelters, airports, shopping malls).

In wider geographic universes cluster sampling normally leaves gaps in the poster sites geography which leads to an underestimation of poster campaign reach. Using ascription techniques we can ‘fill in’ the sample in missing geographic areas and additionally map respondents journey types *in relation to the sites relevant to geographic area to which they have been ascribed*.

Including Posters as part of the media mix raises again the issue of OTS comparability. The difference with other media is that the passage past a poster site is not purposive in relation to the material offered by the site, in contrast to print where you open pages with a specific purpose to select reading material or TV where you sit down specifically to receive programme material. The relative exposure values assigned to sites in relation to the passage type must not only be acceptable within the poster industry but also in a multi-media mix situation. This complex and contentious area deserves a symposium session to itself.

Multi-Media

We will confine this brief discussion to just Press and TV. Since many TV models are very similar to Press Binomial models there should be no problem in adapting them to a multi-media situation. However doing so just exacerbates the problems caused by the independence of probabilities assumption.

In *Sesame* we have the problem of combining the very different VDiary and Binomial models. In fact, since both models are individual models the combination of the two types is theoretically quite simple. There were numerous practical and technical problems to be solved, mainly concerned with the extension beyond the VDiary period (usually just one week) but we have now resolved all the outstanding issues. We have a multi-media model which is not subject to the faults incurred by the use of the independence assumption and retains the dynamics of the TV and radio media in the same way as a real diary or panel.

We are now turning our attention to the use of VDiaries for all media types. It is the most promising route we have examined for resolving the many problems caused by using inappropriate model types or mixing very different model types.

The VDiary route also provides a much simpler response calculation mechanism than do more mathematical distribution based models. This gives the valuable facility of examining many ‘what if’ scenarios of different media-mix combinations efficiently and economically.

Such methods will require at least a partial rejection or re-evaluation of long held views on appropriate models. We think this re-evaluation is long overdue.