L Liebman, R Makely & E Lee
Interactive Market Systems Inc
New York, USA

6.4 A multi-national reach and frequency comparison

INTRODUCTION

This report was organised in two sections. The first, published here contains a description of the analyses performed and a summary of preliminary conclusions.

The second, separate section, which could not be included here for reasons of space, contains the detailed statistical data from which all conclusions were derived. It is available from IMS, as are copies of two previously published articles (Metheringham's 1964 article and the Lee/Liebman ARF 1974 article).

STATISTICAL DATA

The statistical data available as mentioned above are organised by country in alphabetic sequence. For each country the following tables are provided:

(**1**) *OTS distributions* - for each schedule and for each model the OTS distribution.

(**2**) *Coverage summary* - for each schedule, the coverage estimated by each model.

(**3**) *Peaking comparison* - for each schedule the number of times peaks appear in the OTS distribution.

(**4**) *Schedule summary* – insertion levels for each media in each schedule.

In addition to the above data, for the United States analysis additional data are provided:

(**1**) *OTS distribution comparison* – the average absolute difference and relative error of each model are presented.

(**2**) *Coverage summary* – the empirical, tabulated coverages are provided. Also, the difference between each model's estimate and the empirical estimate is tabulated along with the percentage error.

(**3**) *Confidence limits analysis* – for each schedule and each model, the model's coverage estimate is compared to 95% confidence limits for each tabulated coverage estimate.

GOALS

This study has two primary goals: to establish a data base of schedules and models which can be used by others to compare and contrast the performance of alternative models on alternative data bases; and to compare and contrast for conference attendees the performance of the models available to IMS.

Work on the study began in November, 1980 and continued until February 6 1981 when the last data were received and data analysis began.

DATA SOURCES

Studies from five countries were used. The countries and the studies, used were: (**1**) Canada – PMB-II; (**2**) Germany – AG.MA; (**3**) Great Britain – April–September, 1980 NRS; (**4**) Netherlands – NOP; (**5**) USA – SMRB.

Country and data selection was entirely pragmatic – each source was running on IMS. In the United States we could have used either SMRB or MRI. We selected SMRB because we could empirically tabulate the coverage and OTS distributions on schedules of up to two insertions for any number of media. This empirical tabulation could not be performed on MRI.

MODELS AND SCHEDULES ANALYSED

Seven sources of coverage and OTS estimates were available. These included:

(**1**) ADPRO - a personal probability model developed by IMS.

(**2**) CRC/GSI's personal probability model.

(**3**) Holborn Research Service's personal probability model.

(**4**) METREX - a formula model offered by Telmar.

(**5**) MODAL – a formula model offered by IMS.

(**6**) 'Old Metheringham' - a formula model based upon Richard Metheringham's initial work in 1964 and still in commercial use in the United States, Canada and Great Britain.

(**7**) empirical data – actual tabulations of schedules on the United States' SMRB data.

The analysis summary of **Figure 1** below shows which were evaluated in each country. The only source of empirical data was the United States. In Great Britain we evaluated all models other than empirical. In Canada we did not evaluate CRC/GSI or HRS. For Germany and the Netherlands only those models available on IMS were evaluated. 27 schedules were analysed in each of Canada, Germany, Great Britain and the Netherlands, and 57 in the United States.

THE THEORETICAL MAXIMUM COVERAGE LIMIT

A critical concept which has confused the evaluation of coverage and OTS models is *the theoretical maximum*

6.4 A multi-national reach and frequency comparison

FIGURE 1
Analysis summary

	Canada	*Germany*	*GB*	*Netherlands*	*USA*
ADPRO	X	X	X	X	X
CRC/GSI			X		
HRS			X		
METREX	X		X		X
MODAL	X	X	X	X	X
'Old Metheringham'	X	X	X	X	X
Empirical					X

coverage limit. The coverage limit is an issue in all studies having both an average issue and a frequency claim question. The coverage limit states that if a person did not read one or more of the last n issues of a publication (where n is a property of the study), then, the potential reader cannot become an average issue reader.

This assumption is employed, for example, in Great Britain's NRS survey. Its extension to reach and frequency suggests that regardless of the number of insertions in a magazine, a person who did not claim to read one of n can *never* become a reader, even if n becomes *very* large.

A great deal of the dispute between personal probability and formula models is, in reality, a dispute about the appropriateness of the coverage limit concept.

Figure 2 displays, for all available reach and frequency models, those which are formula and those which are personal probability, tabulated against those which can exceed the coverage limit concept and those which can not. It is important in all analyses to keep this differentiation in mind.

To reduce the effect of the coverage limit concept on this comparison study *all schedules selected do not exceed the theoretical maximum coverage limit.*

FIGURE 2
R & F models and the coverage limit concept

	Formula	**Personal Probability**
Cannot exceed limit	Not currently available	Adpro (not US) CRC/GSI HRS
Can exceed limit	Metrex Modal Old Meth	Adpro (US only)

FACTORS INFLUENCING COVERAGE AND OTS ESTIMATES

Two major factors influence coverage and OTS estimates:
(**1**) how the probabilities are calculated for personal probability models and, corresponding to that, how the two issue cumes and pair-wise duplication data are calculated for formula models.
(**2**) the probability model employed to derive the estimates.

Estimating personal probability model parameters

When estimating personal probability model parameters using the SMRB two-issue through-the-book data, it is impossible to assign the probabilities directly. Therefore, the IMS personal probability model for Simmons-type data uses the average issue audience and two issue cumes to fit a beta model; probabilities are derived from this beta model's estimate.

In the United Kingdom, for IMS ADPRO, the probabilities are calculated 'on the fly' for each specific target to be analysed. Therefore, individual respondent's probabilities are not constant irrespective of the analysis but, depending upon the definition of the target market, 'change'. The effect of this is to give up 'additivity' – men plus women will not equal adults for IMS' UK personal probability model. However, in exchange this approach provides a better estimate of the p's for any target group.

In contrast to this the German models as well as the CRC and HRS models, have preassigned probabilities for each respondent. A principal differentiating factor between these is the number of demographic cells for which the p's are tabulated. In the case of CRC we understand this to be two; for HRS we understand this to be 12; for the German models we understand this to be

6.4 A multi-national reach and frequency comparison

up to 16. The greater the number of cells, the 'finer' the cuts at the data base, and the greater the 'reliability' of a personal probability model. Also, keep in mind that additivity is preserved for each of these techniques.

Of all personal probability models examined which rely on preassigned probabilities, those based on data derived using the German procedure will, all other factors being equal, produce the most 'robust' estimates. The German methodology assigns up to 16 probabilities ('p's') to each medium where the cells for the p's are derived separately and uniquely for each medium from a combination of claimed frequency and demographic variables.

In contrast the UK models (CRC/GSI and Holborn) apparently use preassigned probabilities determined for two (CRC/GSI) and 12 (Holborn) demographic groups which are constant across *all* vehicles.

Estimating formula model parameters

All formula models examined rely on average issue audiences (C1) and two issue cumulative audiences (C2) for each vehicle, as well as vehicle pair-wise duplications. These parameters are target specific. As a result, formula models do not generally possess the additivity property.

'Old Metheringham' aggregates C1's, C2's and pairs to create a single 'composite' vehicle representing all vehicles in the schedule.

MODAL employs a sequential application of the beta model to avoid the aggregation problem. MODAL's estimates are order dependent, but 'reasonable' ordering rules based on empirical data analysis have been developed.

All questions about METREX must be addressed to Telmar.

All formula models can produce 'declining reach' due to aggregation.

'Old Metheringham' produces the phenomenon most frequently. It is very rare in METREX and MODAL.

SCHEDULE SELECTION CRITERIA

Using the same selection criteria adopted in February 1973 (see Lee/Liebman ARF article) focused on structural properties common to groups of schedules. They were:
(**1**) magnitude of the average issue audience of each publication, relative to base size.
(**2**) duplication within media.
(**3**) duplication between media pairs, and
(**4**) base size (for schedules run in the US only).

57 schedules were used for the US analysis. The data base chosen was the 1980 SMRB study, to facilitate a comparison with the 1973 study. Furthermore, the only publications selected were those measured 'through-the-book'. Five categories were created for each criterion (1) through (3): high, medium-high, medium-low, low and mixed groupings. Three schedules were created in each grouping. For the groupings in (1) through (3) all analyses were on total population. For (4) three schedules were created. These same three schedules were used for each of four different base sizes. For criteria (1) through (3), no schedules contained any media pairs with extremely high or low duplication (this rule extends as well to the foreign data bases). Schedules were limited to four media, with insertion levels limited to two, so that tabulated data could be used as the criterion for measuring reach estimate accuracy.

A total of 27 schedules were used for each of the other data bases. Again, the number of media were limited to four, with insertion levels also limited to four, to prevent the analysis from becoming overly cumbersome. The first three structural properties were used, with only three groupings created in each. Three schedules were run in each.

Coverage criteria

The US analysis used *tabulated data* as the basis for comparison. Reach estimates from each model were compared to tabulated data to produce three measures:
(**1**) whole differences, percentage differences, and average absolute differences for all 57 schedules.
(**2**) the number of times each method overestimated or underestimated, and
(**3**) the number of times that each estimate was within the 95% confidence limits established around the SMRB tabulated data.

Other countries' data did not permit this kind of analysis because of lack of measured, or actual data. Therefore, reach estimates were simply posted for side-by-side comparisons between techniques. The average of each technique was computed by country for the sum of the 27 schedules run in each, and used in a relative ranking comparison.

OTS criteria

The US analysis used tabulated data as the basis for comparison. Each level of the OTS distributions for each model was compared to tabulated data. The criteria used in evaluating the distributions were:
(**a**) the absolute differences at each frequency level.
(**b**) the relative error of each schedule (the ratio of the sum of the absolute differences at each frequency level (zero level excluded) to the net reach of the schedule).
(**c**) the shape of the frequency curves, with emphasis on the 'rising characteristics' from one frequency level to the next. A table in the statistical data labelled *Number of Peaks* analyses peaks and/or plateaux in the OTS distribution. Any time the OTS distribution *increases* or

6.4 A multi-national reach and frequency comparison

any time it *decreases by less than 5%* (a plateau) the cell for that schedule and model is incremented by one. Cells with weighted counts of less than 50 were ignored.

Since no measured, or actual data existed for the other countries' analyses, the distributions are posted only for a side-by-side comparison. The shapes of the distributions were compared, with a measure of the number of rises, as described above.

Empirical data (US) analysis results

For the 57 schedules analysed, MODAL had the smallest average absolute coverage estimate error, 'Old Metheringham' the largest. The table below of average comparison based on USA Empirical displays the magnitudes of the errors:

	Average absolute error (%)
ADPRO	0.00
MODAL	0.99
METREX	2.16
'Old Metheringham'	6.24

For the 57 schedules, there were a total of 36 rises or peaks in the OTS distribution. The various models produced the following numbers of peaks:

1 ADPRO – 36
2 MODAL - 19
3 METREX - 15
4 'Old Metheringham' - 0

95% confidence limits were computed for each schedule. The number of times each model's estimate was within the confidence limits for the 57 schedules was:

1 MODAL 57 (100%)
2 ADPRO – 57 (100%)
3 METREX - 40 (70%)
4 'Old Metheringham' – 6 (10%)

We further analysed the coverage data. The number of times coverage estimates were one-half coverage point higher or lower than the tabulated data was computed. MODAL was high 24 times, low 10 times, and within the range 23 times. METREX was high 55 times and low two times. We can conclude, therefore, that there is a tendency for formula models to produce coverage estimates which are biased high.

We examined the OTS distribution to determine if there was a consistent bias by MODAL and METREX. MODAL was high 42 times at the one level, and low or the same 15 times; MODAL was high or equal 20 times, low 37 times. METREX was high 57 times at the one level and low 57 times at the two level. This raised a question which we have been unable to answer. Why are these biases concentrated at the one/two levels and not distributed over the OTS distribution?

From examination of the data we determined that formula models tend to underestimate duplication; METREX more so than MODAL.

Observations on Great Britain data

ADPRO and CRC appear to be the most similar pair of models for all schedules analysed on British data.

Coverage differences between formula and personal probability models are concentrated at the one level. The formula models tend to produce more people at the one level; this is similar to the US.

Personal probability models tend to peak at the two, four, and (sometimes) six levels. MODAL and METREX peak at four. We hypothesise that this is caused by the British personal probability models' tendency to generate probability classes which do not discriminate well between high and low probability readers. This statement applies to CRC/GSI, Holborn, and ADPRO in its UK version.

Observations on German data

We focused our comparison on MODAL and ADPRO. The personal probability estimate was higher than MODAL 14 times, lower 13 times. Therefore, the two models are unbiased with respect to each other's coverage estimates. This is true of the OTS distributions also. This is very different from Canada, Great Britain, and the United States.

We do not yet clearly see why two conceptually different models such as MODAL and ADPRO produce such similar results on German data. However, we suspect that it is closely related to the German methodology for setting the probabilities.

Observations on Canadian data

The Canadian and Great Britain conclusions are identical.

Observations on Netherlands data

The formula techniques produce higher reach estimates than personal probability models in the Netherlands.

The magnitudes are similar to Great Britain and Canada.

With respect to the OTS distributions, the formula models are higher at the one level. However, where the personal probability models tend to peak at the two, four, and six levels in Great Britain and Canada, the Netherland estimates tend to rise at the two, three, four levels and at the six, seven, eight levels.

GENERAL STATEMENTS AND FURTHER EFFORTS

There is one over-riding observation. Do not group formula and personal probability models into separate

6.4 A multi-national reach and frequency comparison

classes. There are major differences within each of the types and major similarities across the two types.

Much more extensive analyses by scheduled groups is necessary. These observations are only a very rough, preliminary, first-cut at the data.

We invite data processing companies in all of the countries measured to contribute evaluations of other models on the same schedules and same data. We will add these data to the data bank and we will extend the analyses performed to date. Of course, at all times the data are available to any organisation or person wishing to do research in this area.