Neil Shepherd-Smith
Telmar Communications Ltd
London, UK

6.7 Respondent probability models – a fatal flaw?

Media models for schedule evaluation or schedule construction can be broadly divided into two types: respondent probability models which treat every individual in the survey data base separately, and formula (analytical) models which use aggregated data such as percentage readerships and pairwise duplications between publications or issues.

Respondent probability models

The respondent probability models, having established the probability of an individual (or group of individuals) being exposed to an issue of a given publication then 'expand' such probabilities using the binomial theorem. In other words the probability of exposure to r out of n issues is assumed to be the coefficient of t^r in the expansion of $(q + pt)^n$ where p = personal probability of seeing an average issue, and q = 1 − p, (ie the probability of *not* seeing an average issue).

For example, if there are 100 individuals with a probability p = 0.4 of seeing an issue of a publication then it may be stated that 40 of them see an issue and 60 do not. $100 \times (.6 + .4t)^1 = 60 + 40t$. Note that the coefficient of t^0 gives those seeing 0 issues and the coefficient of t^1 gives those seeing 1 issue.

For two issues the calculation becomes:

$$100 \times (.6 + .4t)^2 = 100 \times (.36 + .48t + .16t^2)$$

The coefficient of t^0 gives those seeing no issues, the coefficient of t^1 gives those seeing one issue only and the coefficient of t^2 gives those seeing two issues.

The resultant frequency distribution looks like this:

Seeing 0	36%
1	48%
2	16%
	100%

Where two publications are concerned, and the respondents have a probability of 0.4 of seeing one publication and 0.3 of seeing the other, the calculations are similar:

$$100 \times (.6 + .4t) \times (.7 + .3t)$$
$$= 100\ (.42 + .46t + .12t^2)$$

The resultant frequency distribution for two publications looks like this:

Seeing 0	42%
1	46%
2	12%
	100%

Formula models

Formula models on the other hand do not go back to individual behaviour, but use summarised data tabulated from the database (usually average issue readership and pairwise inter-issue and inter-publication duplications) as parameters for the formula. Once the tabulations have been done, formula models are extremely fast. They use very much less data than the respondent probability models; it will be noted for ten publications there are only ten readerships, ten inter-issue duplications and 45 inter-publication duplications totally 65 data items. A respondent probability model would have to cope with ten probabilities for each of the respondents, which could be as many as 250,000 data items depending on the survey.

Probability calculation

There is a practical difficulty for respondent probability models in establishing the probabilities themselves for different target markets. A probability is established for a given frequency claim for a given publication by dividing the total respondents making the claim into those found to be 'average-issue' readers. The ratio of readers to claimers (ie the probability) is likely to be different for each subcategory of the population.

For example for all adults claiming to read four out of six issues of the *Daily Mail* the probability of seeing an average issue is 0.37. When these claimers are analysed by subgroups, the probabilities are not necessarily the same in every case.

The probability of adults in each socio-economic group claiming to see four out of six issues of the *Daily Mail* is shown in **Table 1**.

Ideally therefore, in a respondent probability model, the computer should make two 'passes' of the survey data, firstly to establish the probabilities by frequency claim within publication for the specified target subgroup, and secondly to 'attach' those derived probabilities to each of the respondent members of the subgroup. In practice this would add significantly to the cost of running what is already an inherently costly model and the operators solve the problem by storing 'standard' probabilities for each publication and then applying those probabilities to the respondents in the target subgroup in question. Where the actual probabilities for the subgroup respondents differ significantly from the 'standard' probabilities, errors result in the readership estimates. In an attempt to improve an admittedly unsatisfactory situation, the more conscientious operators will therefore establish the 'standard' probabilities by sex and/or age

Respondent probability models – a fatal flaw?

TABLE 1
Those 'reading' the *Daily Mail* among those claiming to see four out of six issues. Analysed by socio-economic group

	All adults	*A adults*	*B adults*	*C1 adults*	*C2 adults*	*D adults*	*E adults*
Claimers	408	7	57	144	128	66	6
Readers	151	0	27	61	35	25	3
Probability	.37	0	.47	.42	.27	.38	.50

Source: UK NRS July 1979–June 1980. All adults.

and/or other category for each frequency claim for each publication. But it is of course impossible to cater for every possible subgroup, particularly with the increasing trend to weighted target universes. As a result, unless the target subgroup happens to fit the 'standard' probabilities exactly, the result of applying the probabilities in this way is to distort the readerships of the publications. That is the reason why the gross impressions or gross OTS calculated from the frequency distribution of a respondent probability model often differ significantly from the known gross readership of the component publications in the schedule. In this context it is perhaps worth noting that when the readerships, inter-issue and inter-publication duplications are tabulated for a formula model, they should be based precisely on the specified target subgroups, incorporating market weights if defined. This will preserve the consistency between the 'input' readerships of the publications and the 'output' frequency distribution.

The basic respondent probability model assumption

Leaving aside for a moment the errors in respondent probabilities models from the application of incorrect probabilities, the statement has often been made that such models are more 'accurate' than formula models because they use every individual in the sample. As we have seen, the underlying principle of the probability model is that the readership probabilities are independent. In other words, the assumption is made that if a respondent has a probability of 0.4 of seeing publication X then his probability of seeing two issues of publication X is 0.4 × 0.4 ie, 0.16. Similarly if his probability of seeing publication Y is 0.3 then his probability of seeing both publication X and publication Y is 0.4 × 0.3 ie, 0.12. To my knowledge this assumption has never been questioned, which is surprising when it forms the basis for models which are meant to be accurate if expensive.

A test of the assumption

There is of course no precise yardstick for testing the validity of results from a model evaluating a multi-publication, multi-insertion schedule. However, we can test the assumption that readership probabilities (as currently calculated and used in respondent probability models) are independent by using the probabilities to calculate the duplication between the average-issue readerships of two publications, and then comparing the results with a tabulation from the survey itself.

I have based this exercise on the JICNARS NRS for July 1979–June 1980, using All Women probabilities (as published in the report) to avoid any possible distortions from variation in probability among different target subgroups.

Two weekly publications, *Woman's Own* and *Woman*, were used. The 'claimers', 'readers' and probability for each frequency claim are shown in **Table 2**.

The two publications were then analysed in terms of their frequency claim cells in each case. The complete matrix is as **Table 3**.

It can be seen from the table that, for example, 112,000 women claimed to see two out of four issues of *Woman's Own* and three out of four issues of *Woman*.

The next stage was to establish the combined probability for each frequency claim cell in the matrix, ie a probability in each case of seeing both *Woman's Own* and *Woman*. The calculation was done by multiplying the appropriate probability for the frequency claim for one publication by the frequency claim probability for the other publication.

For example, the probability of seeing an issue of both *Woman's Own* and *Woman* for those claiming two out of four issues of *Woman's Own* (probability: .401) and also three out of four issues of *Woman* (probability .585) is the product of the two probabilities, ie .401 × .585 = .235. This calculation was carried out for each combination of frequency claims to produce the matrix of probabilities in **Table 4**.

TABLE 2
Claimers, readers + probabilities. Universe: all women ('000s)

	Claimed number of issues read						
	4	*3*	*2*	*1*	*<1*	*0*	*Total*
Woman's Own							
Claimers	4452	722	2272	2098	1152	11598	22294
Readers	3804	388	912	511	135	0	5750
Probability	.854	.537	.401	.244	.117	0	.258
Woman							
Claimers	4064	698	2130	1997	1151	12254	22294
Readers	3496	408	895	536	142	0	5477
Probability	.860	.585	.42	.268	.123	0	.246

Source: UK NRS July 1979–June 1980.

TABLE 3
Analysis of frequency claim cells. Universe: all women ('000s)

Woman's Own	***Woman*** **claimers**						
claimers	*4*	*3*	*2*	*1*	*<1*	*0*	*Total* Woman's Own
4	3064	88	199	105	41	955	4452
3	63	375	114	29	6	135	722
2	160	112	1237	161	39	563	2272
1	78	38	147	1186	68	581	2098
<1	40	6	30	59	737	280	1152
0	659	79	403	457	260	9740	11598
Total *Woman*	4064	698	2130	1997	1151	12254	22294

Source: UK NRS July 1979–June 1980.

TABLE 4
Combined probabilities. Universe: all women

Woman's Own	***Woman*** **probabilities**						
probabilities	*.860*	*.585*	*.420*	*.268*	*.123*	*0*	
.854	.734	.500	.359	.229	.105	0	4
.537	.462	.314	.226	.144	.066	0	3
.401	.345	.235	.168	.107	.049	0	2
.244	.210	.143	.102	.065	.030	0	1
.117	.101	.068	.049	.031	.014	0	<1
0	0	0	0	0	0	0	0
	4	*3*	*2*	*1*	*<1*	*0*	

Source: UK NRS July 1979–June 1980.

6.7 Respondent probability models – a fatal flaw?

Having established the probability of seeing both publications for each cell, it was then possible to apply the probability to the relevant number of claimers in each case (see **Table 3**) to obtain the estimate of those reading both publications (**Table 5**).

The 'calculated' results could then be compared with the actual average-issue readers of both publications tabulated from the survey (**Table 6**).

It can be seen that there are significant differences in virtually every cell between the calculated readers of both publications using the probabilities and the actual readers tabulated from the survey. The percentage variation of the calculated figures from the actual are shown in **Table 7**.

In every case the estimate calculated by the probability method is much too low, producing a total underestimate of 861,000 or 22% from the tabulated figure. This error will of course produce a corresponding variation in the reach estimate of the schedule.

Inaccuracy

What are the reasons for the inaccuracy? First, to soothe unnecessary worries, it is *not* being suggested that the theory of independent personal probabilities is no longer valid! On the contrary, if it *were* possible to find for each respondent in the database his or her precise probability of reading every publication, then such probabilities could indeed be multiplied together to produce the probability of reading two or more publications. The problem is of course in establishing the probabilities in the first place.

A 'probability' as currently used is in fact the mean probability of reading for a whole group of respondents who may have a range of probabilities between nought and one. We have already seen how the probabilities can vary between different socio-economic groups, and the variation between individuals is likely to be greater still. The more all-embracing a group is, the greater the range of individual probabilities within that group, the less valuable the mean probability will be. To illustrate this

TABLE 5
Calculated readers of *Woman's Own* and *Woman*, ie claimers × probability
Universe: all women ('000s)

***Woman's Own* claimers**	***Woman* claimers** *4*	*3*	*2*	*1*	*<1*	*Total*
4	2249	44	71	24	4	2392
3	29	118	26	4	0	177
2	55	26	208	17	2	308
1	16	5	15	77	2	115
<1	4	0	1	2	10	17
Total	2353	193	321	124	18	3009

Source: UK NRS July 1979–June 1980.

TABLE 6
Readers of *Woman's Own* and *Woman* tabulated directly from the survey
Universe: all women ('000s)

***Woman's Own* claimers**	***Woman* claimers** *4*	*3*	*2*	*1*	*<1*	*Total*
4	2555	56	75	29	8	2723
3	30	175	41	3	0	249
2	68	35	418	28	3	552
1	21	9	20	218	4	272
<1	5	3	4	10	52	74
Total	2679	278	558	288	67	3870

6.7 Respondent probability models – a fatal flaw?

TABLE 7
Percentage variation of calculated on actual readers. Universe: all women

***Woman's Own* claimers**	***Woman* claimers** 4	3	2	1	<1	Total
4	−12%	−21%	−5%	−17%	−50%	−12%
3	−3%	−33%	−37%	+33%	0%	−29%
2	−19%	−26%	−50%	−39%	−33%	−44%
1	−24%	−44%	−25%	−65%	−50%	−58%
<1	−20%	–	−75%	−80%	−81%	−77%
Total	−12%	−31%	−42%	−57%	−73%	−22%

Source: UK NRS July 1979–June 1980.

point by its logical extreme, it would be possible to multiply the overall mean probability of reading *Woman's Own* (0.258) by the overall mean probability of reading *Woman* (0.246) to derive the mean probability of reading both (0.258 × 0.246 = 0.063468). However, this combined probability when multiplied by the total 'claimers' (in this case the All Women population of 22,294) gives 1415 as opposed to the actual figure of 3870 (all three numbers in thousands).

It is for this reason that the frequency claims are used to segment the 'readers' of a publication into smaller and more useful probability groups (including of course a large group of zero probabilities). As we have seen, the increase in the number of probability cells improves the accuracy of the estimate, but not unfortunately to a satisfactory level. One of the problems is that the frequency claims establish for example the mean probabilities of *all* women readers of *Woman's Own* and *all* women readers of *Woman*. Yet what we need in this case are the probabilities of that *subgroup* of *Woman's Own* readers who also read *Woman* and the subgroup of *Woman* readers who also read *Woman's Own*. Because we have *not* got the accurate subgroup probabilities and make do with the 'all-reader' probabilities in each case, the result inevitably must be subject to error.

Another way of looking at the problem is that using the 'all-readers' probability in each case to calculate the 'readers' of both publications is in effect to assume random duplication between the cells. In fact, duplication between *Woman's Own* readers and *Woman* readers is very far from random; if a female respondent is a reader of *Woman's Own* she has a much higher likelihood of being a reader of *Woman* than a non-reader of *Woman's Own* has. That is why the probability method *underestimates* rather than overestimates the readership of both publications in this case.

The practical considerations

Does the inaccuracy matter in practice? It depends whether the user minds about accuracy or not! It is of course true to say that the larger the schedule (with a corresponding larger number of mixed publications) and the higher the reach, the less the reach errors produced by the respondent probability method will matter, on the broad general basis that in a large enough schedule you reach almost everybody. However, it is difficult to see the logic in using such an expensive method of producing figures of such dubious value.

For small schedules, as we have seen, the results can be dangerously misleading. It would of course be possible to avoid the problem of inaccuracy when there is only one insertion in each publication, by tabulating the results straight from the survey, ie without using the respondent probability method. However that would simply postpone the problem because inconsistencies would occur as soon as another insertion was added to any publication on the schedule.

It is clear from the above example that respondent probability models, as currently used, have a serious flaw in the basic assumption underlying their calculations of reach. If formula models (which have in the past been somewhat unfairly attacked in uninformed quarters for the alleged inaccuracy of their results) were to produce estimates as demonstrably incorrect as the *Woman's Own/Woman* illustrated in this paper, they would be regarded as quite unusable for any practical media planning purposes. As I have said earlier, there is no real yardstick of the reach of a multi-insertion, multi-publication schedule, but one might suggest that the reach estimate of a respondent probability model is not a good guide to the true reach. At the very least it is hoped that the practice of suggesting that respondent probability models give the true results against which formula models can be judged should now cease.

Formula models

There is a very strong case to be made for using a formula model particularly one based on the beta-binomial frequency distribution. The speed and therefore the benefits of reduced cost have already been referred to earlier in this paper, but it is important to realise that the cheapness is not achieved at the expense of accuracy.

The close correspondence of the beta-binomial distribution to the actual distribution of exposures from several insertions in a given publication has been amply demonstrated many times in the past.[1, 2]

An example based on six issues of *Life* magazine is as **Table 8**.

To produce such a distribution, the beta-binomial formula needs as its input parameters the readerships of one and two issues of a given publication. In some

TABLE 8
Six issues of *Life*

Number of issues out of six	*Estimated (Beta-binomial) %*	*Observed %*
0	52.91	52.92
1	14.92	15.29
2	9.54	9.55
3	7.25	6.85
4	5.93	5.15
5	5.05	5.88
6	4.40	4.36

Source: Politz survey 1966.

TABLE 9
Reach estimates for 27 schedules (single insertions)

Schedule		*Cosmic %*	*Observed %*
1	W Weekly/My Weekly/W Own/I Home	27.8	28.0
2	Exch & Mart/P Friend/Weekend/W Weekly	24.4	24.5
3	W & Home/F Circle/G Housekeeping/Vogue	19.7	19.8
4	Cosmo/T Romances/House & Garden/Homes & Gardens	16.0	16.0
5	H F Digest/Living/Slimming & Nut/T Romances	13.8	13.8
6	P Householder/DIY/Titbits/Mayfair	12.4	12.4
7	N Mus Express/Autocar/Punch/Economist	6.8	6.8
8	Motor/Economist/Melody Maker/Autocar	6.2	6.3
9	W Journal/Harpers & Queen/Over 21/True Mag	7.5	7.5
10	True Story/She/Over 21/Harpers & Queen	10.9	11.0
11	House & Garden/Vogue/True Story/Homes & Gardens	14.9	14.8
12	Autocar/Economist/Titbits/Melody Maker	8.5	8.5
13	P Householder/Men Only/Mayfair/P Motoring	9.7	9.6
14	True Romance/W Journal/H F Digest/True Mag	10.4	10.4
15	Living/Cosmo/H F Digest/G Housekeeping	17.2	17.2
16	F Circle/Woman & Home/Annabel/Slimming & Nut	17.5	17.5
17	R Times/New Musical Express/TV Times/Woman	38.3	38.4
18	Jackie/W Weekly/TV Times/Weekly News	36.3	36.5
19	Motor/Economist/New Musical Express	5.7	5.7
20	Ideal Home/P Householder/Cars & Car Conversions	10.8	10.8
21	Annabel/Homes & Gardens/Living	11.2	11.2
22	Economist/P Friend/Weekend	12.8	12.8
23	Now!/R Digest/Cars & Car Conversions	21.2	21.1
24	True Mag/Slimming & Nut/Living	10.0	10.0
25	Woman/Men Only/My Weekly	21.1	21.2
26	R Digest/Ideal Home/Hot Car	24.7	24.7
27	Good Housekeeping/True Mag/House & Garden	12.9	12.9

countries this measure can be obtained directly from the survey. In other countries, among them the UK, the two-issue readerships have to be calculated from the frequency questions and the average-issue readerships.

Perhaps more important still is the fact that a good formula is firmly based on observed data, tabulating readerships and duplications, directly from the survey data for the precise specified target market, incorporating market weights if necessary. As a result the formula model is internally completely consistent, with the gross exposures calculated from the frequency distribution equalling the gross exposures calculated from the component publication readerships; as we have seen earlier, that is not necessarily true for a respondent probability model.

The accurate tabulation of duplications of course means that the formula model's evaluation of a schedule like the *Woman's Own/Woman* example will match the observed results exactly; the errors inherent in the respondent probability model simply cannot occur. It is because the formula is based on observed data that it can give such accurate results when more than two publications are used. To illustrate this point, below are the results of 27 schedule evaluations from the COSMIC model compared with the observed results tabulated from the survey. A tabulated yardstick of this sort is only possible for schedules with one insertion in each publication; the publications used were taken from the schedules selected (apparently at random) for another paper in this symposium (see **Table 9**).

Conclusion

The flaw in the basic concept on which respondent probability models are based has been demonstrated in the *Woman's Own/Woman* example above. It is suggested that this flaw is so fundamental that such models should be used only with the utmost caution, particularly with the practical difficulties of establishing accurate readership probabilities for anything but 'standard' target groups. Instead, the user would be well advised to take advantage of the speed and accuracy of a well-established formula model for all schedule evaluation.

REFERENCES

1 Greene, J D, and Stock, J F (1967). *Advertising Reach & Frequency in Magazines* Reader's Digest Association Inc.
2 Simmons, W R (1968). *An examination of alternative methods of estimating media reach and frequency* W R Simmons and Associates Research Inc, New York.