ESOMAR RADIO SYMPOSIUM - JULY 1997

SYNOPSIS

Title: Radio Audience Measurement - Continuity or Probability Models

Author: Steve Wilcox, RSMB Television Research Ltd

Keywords: Continuous panels; one week diaries; average audiences; extended reach; probability models; cost; sampling error; respondent fatigue.

The widely accepted standard for the measurement of television audiences is with continuous panels and usually an electronic meter system. This is particularly appropriate for high penetration channels, providing a time, date and schedule specific measurement of the audience. A panel also provides the opportunity to track each individual's behaviour over an extended period of time.

In the more fragmented radio market there is less need for a detailed measurement of the audience but a greater need for large sample bases. The RAJAR model in the UK is typical. A series of consecutive one week diary samples are accumulated and published audiences relate to an average week in a three or six month period. Audience measurements which normally require continuity of reporting over longer periods of time - for example extended reach - are generated using probability models.

This paper will describe the probability models currently used in the UK and evaluate the advantages and disadvantages of large sample non-continuous surveys against a smaller continuous panel measurement in terms of cost, sampling error and weaknesses in the probability models. The relative importance of extended reach and average audiences will be considered. Could two or four week diaries offer significant advantages or is there too much of a respondent fatigue problem?

Analyses to support the paper will be based upon the RAJAR special four week diary study and the BARB TV audience measurement panel.

Radio Audience Measurement
- Continuity or Probability Models

by Steve Wilcox, RSMB Television Research Ltd.

This paper reports on an investigation of the relative performance of one, two and four week diaries for radio audience measurement. A special four week diary study carried out in 1992 by Research Services Limited for RAJAR has been re-analysed to evaluate the trade-offs between respondent fatigue, the performance of the extended reach model, sample efficiency and relative cost.

1. Introduction

The widely accepted standard for the measurement of television audiences is with continuous panels and usually an electronic meter system. This is particularly appropriate for high penetration channels, providing a time, date and schedule specific measurement of the audience. A panel also provides the opportunity to track each individual's behaviour over an extended period of time.

In the more fragmented radio market there is less need for a detailed measurement of the audience but a greater need for large sample bases. The RAJAR model in the UK is typical. A series of consecutive one week diary samples are accumulated and published audiences relate to an average week in a three month period. Audience measurements which normally require continuity of reporting over longer periods of time - extended reach for example - are generated using probability models.

Whilst this model appears to provide the optimum trade-off between sample size and cost, it may be that smaller samples of two or four week diaries provide a larger effective sample size, although any gain must be set against the issue of respondent fatigue. It is probably also the case that a longer recording period for each individual will provide a better basis for extended reach projections. Therefore the main purpose of this paper is to reconsider the advantages and disadvantages of the one week diary methodology against two or four week diaries in terms of the following criteria:

- Respondent fatigue
- The performance of the extended reach probability model
- Sample efficiency and effective sample size
- Cost

In making decisions concerning the design of the survey it is of course important to consider the full range of audience measurements required and their relative importance. This is often the most difficult part of the specification exercise. For the purposes of this paper and to illustrate the principles, the following measurements have been considered:

- Weekly average hours of listening
- Weekly reach
- Projected four week reach

All the analyses used in this evaluation are based upon a special one-off four week diary study commissioned by the UK's commercial radio industry and carried out by Research Services Limited in 1992. The sample comprised nearly 1400 adults and over 250 children aged 4-14.

2. Respondent Fatigue

The issue of respondent fatigue was debated extensively prior to the launch of the new RAJAR service in 1993. There are many ways in which respondent fatigue can manifest itself, therefore the following data can only give an overview.

Table 2.1 is based upon weekly average hours of listening to all radio as measured in the four consecutive weeks of the diary study. For ease of interpretation the average hours for weeks two, three and four have been indexed on the first week.

Table 2.1 Here

Overall there is a small fall-off in weeks two and three but a more significant decrease in week four. Of the larger sample groups - Men, Women, Adults ABC1 and Adults C2DE - only Adults ABC1 are seen to deviate from this pattern with a larger decrease in week two. The age breaks analysed are quite narrow and therefore produce more variable data. However, the fall-off in average hours for Adults 15-24 also starts in week 2 and is more significant. The results are consistent with the hypothesis that fatigue starts earlier and is more of a problem for young and upmarket adults.

Table 2.2 shows the equivalent indexes for weekly reach to all radio for All Adults and the two demographic groups highlighted above.

Table 2.2 Here

Again there is fall-off by week four and some evidence that this starts earlier amongst 15-24's and ABC1's. However, given that weekly reach to all radio is around 95%, it is not surprising that a possible fatigue effect is not really apparent in these data. What is more relevant is the effect on each individual station's weekly reach.

Although the four week study recorded listening to all available channels the sample size was only sufficient to itemise all seven national stations and eleven of the local stations. These are listed in Table 2.3 together with their average weekly reach estimates, shown to indicate the size of each station.

Table 2.3 Here

The best way to summarise the changes from week to week in reach for a number of channels is to calculate the average number of channels listened to at all in a week. This is a measurement of the listening repertoire. It is calculated simply by summing the individual stations' reach estimates then dividing by 100. The calculation is not quite perfect here because a large number of local stations are covered by "All Other Stations". However, the distortion will not be large because listening repertoires are quite small. Table 2.4 shows the weekly repertoires and their indexes on the week 1 repertoire.

Table 2.4 Here

In terms of the fall-off in viewing repertoire there is little difference between the demographic groups. However, there is a significant decrease in week two followed by a step change down to weeks three and four. Decreases in the number of stations recorded is logically consistent with a respondent fatigue hypothesis.

These results can be re-interpreted to indicate that the average station will suffer a 4% drop in weekly reach (number of listeners) in week two followed by a step change down to 11% below week one for weeks three and four.

Finally, in order to investigate differential effects by station, Table 2.5 looks at station share.

Table 2.5 Here

There is perhaps some evidence of a shift towards national BBC radio and in particular BBC Radio 2, but really none of the movements are significant.

To summarise, the effects which are consistent with respondent fatigue are:

- Recorded average hours of listening are 1% lower in weeks two and three and 5% lower in week four. This effect is greater for the young (15-24) and the upmarket (ABC1).
- The viewing repertoire (number of stations listened to per week) is 4% lower in week two and 11% lower in weeks three and four. The weekly reach for an average station suffers the same declines.
- There is no evidence of a differential effect by station.

The evidence is that a two week diary does not suffer as much as a four week diary. Given that any fatigue effect would be averaged out in reported audience estimates, the two week diary fatigue effect may be an acceptable trade-off. However, the four week fatigue effect for station reach is probably too large to be acceptable.

3. The Extended Reach Probability Model

The RAJAR model for the prediction of extended period station reach is probability based. It uses data for reach build within a single week as measured by the latest quarters' surveys to predict reach build over an extended period beyond a week. As the relationship between average daily reach and weekly reach varies between stations and demographic groups, it follows that the pattern of extended reach build will differ between stations and between demographic groups.

The pattern of reach build for a particular station will not be fixed but may alter over time as the relationship between average daily reach and weekly reach changes in the RAJAR results.

The Negative Binomial Distribution (NBD) is a probability model that describes the frequency distribution which is the percentages of people who listen to a station on 0,1,2,3,..... etc. days. All that is needed to determine the whole frequency distribution are two pieces of actual survey data:

One week reach
Average daily reach

expressed as percentages of the population. The probability model then fills in the rest of the frequency distribution, including extended reach estimates. The success of the model is obviously dependent upon how well it fills in the rest - the goodness of fit.

The NBD is a two parameter model ('a' and 'b' in the following formula):

$$\text{Proportion listening on i days} = \frac{\Gamma(k+i)}{\Gamma(k)\Gamma(i+1)} \cdot \frac{a^i}{(1+a)^{k+i}}$$

Two much simpler relationships can be derived:

One week reach = $100.[1-(1+a)^{-k}]$
Average daily reach = $100.a.k/7$

Then if we know one week reach and average daily reach for a particular station, we can calculate values for 'a' and 'k'.

This same probability model is used to predict the frequency distribution for any number of weeks by simply multiplying the 'a' parameter by the number of weeks:

$A = a \times n$ where n = number of weeks
n week reach = $100.[1-(1+A)^{-k}]$

The four week diary study has been used to validate the model. The prime concern is how well the model can predict four week reach from a one week diary survey. However it is important to take note of the inherent weakness of the results from the four week diary study which are believed to suffer from respondent fatigue as discussed in section 2 above. By fitting the NBD model to weekly and average daily reach as measured by the average of the four weeks, the fatigue effect is ironed out to a certain extent. A comparison of the actual and predicted four week reach figures is shown in Table 3.1 below, together with the sampling error for each four week reach estimate. (Reach figures for local stations have been rebased to form percentages of those who receive the station.)

Given the degree of latitude required to allow for fatigue effects and also considering the sampling error involved , the model fit was deemed to be acceptable by RAJAR. However, it is logical to hypothesise that a longer recording period for each individual would provide a better basis for extended reach projections. The same NBD probability model would be used but the inputs would be:

Two week reach
Average daily reach

The formulae for calculating the parameters 'a' and 'k' become:

Two week reach = $100.[1-(1+a)^{-k}]$

Average daily reach = 100.a.k/14

The formulae for calculating extended reach become:

$A=a \times n/2$ where n = number of weeks

n week reach = $100.[1-(1+A)^{-k}]$

Using these formulae, the four week diary study has been used to determine how well the model can predict four week reach from a **two** week diary survey. These alternative four week reach predictions are also shown in Table 3.1.

Table 3.1 Here

Based upon the differences (shown in brackets) from actual reach, the overall impression is that the predictions based upon a two week diary survey are indeed more accurate. The relative performance is summarised by calculating the average of the differences (an indicator of model bias) and the average of the absolute differences (an indicator of the average prediction error). These summaries are shown in Table 3.2.

Table 3.2 Here

Whilst there is some evidence that the two week diary base tends to under-estimate four week reach (larger negative average differences) the gains in prediction accuracy (smaller average absolute differences) are important. Their significance is demonstrated more clearly by showing the differences as percentages of the average reach figures as in Table 3.3.

Table 3.3 Here

In summary, whilst a one week diary base does provide acceptable predictions of four week reach which are not statistically significantly different from the actual four week diary results, a two week diary base is seen to provide important gains in prediction accuracy.

4. Sample Efficiency and Effective Sample Size

If each respondent is asked to complete a two or four week diary as opposed to a one week diary then there will be an increase in the cost per respondent. However this will not be a pro-rata increase. Relative costs are discussed in section 5.

On the surface this is not an attractive option because for a fixed total expenditure, the survey sample size would have to be reduced. However, in order to fully understand the trade-off it is necessary to consider the **effective sample size** which is the real determinant of sampling error.

Following the original work by Arbitron (1974) it has been demonstrated many times that for a given sample of individuals and a particular audience measurement, the sampling error decreases as the data is averaged over longer periods of time. This is because all individuals are, to a greater or lesser degree, inconsistent in their listening behaviour from week to week. Therefore averaging each individuals' listening over two or four weeks results in a regression to the mean (ie. individuals become more similar) and therefore there is a reduction in sampling error. This increase in sample efficiency is best measured by calculating the effective sample size for two or four week averages compared to the bench-mark one week average.

The effective sample size is the notional sample size for a one week diary survey that would give the same sampling error as a two or four week diary survey. The formula for calculating the effective sample size is as follows:

$$\text{Effective Sample} = \left(\frac{\text{Sampling Error on One Week Average}}{\text{Sampling Error on Two (Four) Week Average}}\right)^2 \text{x} \begin{pmatrix}\text{Sample}\\ \text{Size}\end{pmatrix}$$

In this particular application the "Sample Size" is the sample size of the four week diary study from which the sampling errors of one, two and four week averages are calculated. It

should be noted that these sampling errors do take account of the fact that respondents are clustered within households.

As an example, consider the average weekly hours of listening by All Adults to BBC Radio 1:

Sample Size = 1397 Weekly average hours = 7.44

% Sampling error
- One week diary = 7.05%
- Two week diary = 6.86%
- Four week diary = 6.72%

Effective sample size
- Two week diary $= \left(\frac{7.05}{6.86}\right)^2 \times 1397 = 1474$
- Four week diary $= \left(\frac{7.05}{6.72}\right)^2 \times 1397 = 1538$

So a two week diary benefits from a 5% increase in effective sample size and a four week diary benefits from a 10% increase in effective sample size.

In practice, each audience measurement, each station and each demographic group will be affected differently. Table 4.1 shows the gains in effective sample sizes for the measurement of weekly average hours and weekly reach for All Adults and each station. (For local stations the calculations are based upon receivers only.)

Table 4.1 Here

As expected there is quite a variation by station and the gains are different for average hours and weekly reach. Some of the variations by station may well be random effects. However, the results for BBC Radio 3 and BBC Radio 5 are interesting. Listening to these two stations is more programme led and it is more likely that individuals are more inconsistent from week to week. This would result in the higher gains in effective sample size which are apparent for weekly reach using the two and four week diary bases.

In order to iron out what may be random fluctuations between stations, Table 4.1 also shows the averages of the gains for the seven national stations and the eleven local stations. These are of the same order of magnitude as the gains in effective sample size for "Other Local" and "Total Radio". Therefore the average gain for the national stations provides a reliable guide to the average gains in effective sample size whilst BBC Radio 3 and BBC Radio 5 are a good guide to the gains for programme led stations.

The guidelines are summarised in Table 4.2.

Table 4.2 Here

To explore variations by demographic group, Table 4.3 shows the gains in efficiency for Total Radio.

Table 4.3 Here

There is some evidence of larger gains for Adults 15-24 and Men but in general there are no dramatic differences between demographic groups.

Given that two or four week diary surveys produce larger effective sample sizes, it follows that smaller actual samples can be used to generate the same level of precision as a one week diary survey. Table 4.4 is an interpretation of Table 4.3 and shows the relative sample sizes required to produce equivalent precision from one, two and four week diary surveys.

Table 4.4 Here

Dependent upon the relative costs of one, two and four week diary surveys, it could be that the longer recording periods offer significant improvements in sample efficiency. This trade-off is explored in the next section.

5. Relative Costs

This has to be the most speculative part of the paper because an extension to the diary recording period may require a number of unforseen changes to the survey methodology. For

example, a longer diary period may require additional interviewer contacts and respondent incentives. In arriving at the cost increase guidelines we have taken advice from Research Services Limited (the contractor for RAJAR). However, it is important to stress that this was not a rigorous costing exercise and the information presented here is our interpretation of their views. Again the intention is to demonstrate a principle which is worthy of further investigation.

If there were no change in sample size, the costs of two and four week diary surveys would be respectively 30% and 90% higher than the cost of a one week survey. Table 4.4 in the previous section shows the relative sample sizes required to produce equivalent precision from one, two and four week diary surveys for average hours of listening and weekly reach for an average station and for a "programme led" station. Table 5.1 shows the relative cost of the alternative designs against the existing one week diary design.

Table 5.1 Here

Obviously these cost comparisons are less extreme than if the longer recording periods were based upon the same sample sizes. To summarise:

- For the measurement of station weekly reach and average hours of listening, two or four week diary surveys are **less** cost effective than a one week diary survey.
- Four week diary surveys are significantly more expensive and suffer from a large respondent fatigue effect.
- The increased cost of two week diaries is much smaller, particularly if weekly reach is considered to be the more important measurement which is used to determine the sample size. A two week diary suffers only a minimal respondent fatigue effect and offers a significant improvement in the model performance for extended reach prediction.

6. Conclusions

All the conclusions drawn in this paper are based upon survey performance in terms of station reach and average hours of viewing as measured by one, two and four week diaries.

It is believed that longer term diary recording periods suffer from respondent fatigue. Based upon the findings of the four week diary study, the evidence is that a two week diary does not suffer as much as a four week diary. Given that any fatigue effect would be averaged out in reported audience estimates, the two week diary fatigue effect may be an acceptable trade-off. However, the four week fatigue effect for station reach is probably too large to be acceptable.

The RAJAR model for the prediction of extended period station reach is probabilty based. Whilst a one week diary does provide acceptable predictions of four week reach, a two week diary base is seen to provide important gains in predicition accuracy. Whilst not explored in this paper, the advantages for the modelling of schedule reach and frequency are likely to be greater.

Analysis of the four week diary study shows that sampling error decreases as the data is averaged over longer periods of time. This is equivalent to increases in the effective sample size and means that for equivalent precision, two and four week diary surveys can be smaller than one week surveys. Thus the increased cost of longer diary recording periods can be offset by the reduced cost of a smaller sample. Whilst the evidence is that a four week diary survey is significantly more expensive and suffers from a large respondent fatigue effect, the increased cost of a two week diary survey is much smaller and suffers only a minimal respondent fatigue effect.

Overall it seems that the costs and fatigue effects associated with a four week diary would rule out this option as an alternative to the existing one week diary survey. However, a two week diary is a viable alternative which provides significant improvement in the model performance for the prediction of longer term reach.

When considering the survey design for a future radio audience measurement service, it is to be recommended that a full evaluation of the performance of one versus two week diaries is made. This should embrace all applications of the survey and be based upon more rigorous costings.

Table 2.1

	Total Radio Weekly Average Hours Index on Week 1			
	Week 1	**Week 2**	**Week 3**	**Week 4**
Adults 15+	100	99	99	95
Children 4-14	100	107	99	94
Adults 15-24	100	89	83	79
Adults 25-34	100	100	97	99
Adults 35-44	100	101	97	95
Adults 45-54	100	101	102	97
Adults 55-64	100	98	102	96
Adults 65+	100	105	110	104
Men	100	99	98	95
Women	100	99	100	95
Adults ABC1	100	96	97	93
Adults C2DE	100	101	100	97

Table 2.2

	Total Radio Weekly Reach Index on Week 1			
	Week 1	**Week 2**	**Week 3**	**Week 4**
Adults 15+	100	101	100	99
Adults 15-24	100	101	98	99
Adults ABC1	100	99	98	98

Table 2.3

Average Weekly Reach for All Adults					
National Stations		**Local Stations**		**Local Stations**	
BBC Radio 1	41%	Trent FM	2%	Jazz FM	1%
BBC Radio 2	22%	BRMB FM	2%	Kiss 100	1%
BBC Radio 3	6%	Capital FM	6%	LBC Talkback	2%
BBC Radio 4	19%	Capital Gold	4%	LBC Newstalk	1%
BBC Radio 5	7%	Power FM	2%		
Atlantic 252	10%	Clyde 2	2%	All Other Stations	58%
Classic FM	10%	Melody FM	3%		

Table 2.4

	Weekly Listening Repertoire = Number of Stations			
	Week 1	**Week 2**	**Week 3**	**Week 4**
Adults 15+	2.13	2.05	1.89	1.89
Index	*100*	*96*	*89*	*89*
Adults 15-24	2.15	2.03	1.87	1.89
Index	*100*	*94*	*87*	*88*
Adults ABC1	2.29	2.19	2.04	2.06
Index	*100*	*96*	*89*	*90*

Table 2.5

	Weekly Station Share for All Adults			
	Week 1	**Week 2**	**Week 3**	**Week 4**
BBC Radio 1	22%	22%	21%	21%
BBC Radio 2	12%	13%	14%	14%
BBC Radio 3	2%	2%	1%	2%
BBC Radio 4	4%	4%	5%	5%
BBC Radio 5	1%	1%	1%	1%
All National BBC	**41%**	**41%**	**42%**	**42%**
Atlantic 252	4%	4%	3%	4%
Classic FM	2%	2%	2%	2%
All National Commercial	**6%**	**6%**	**5%**	**5%**
All Local	**53%**	**53%**	**53%**	**53%**

Table 3.1

	All Adults Four Week Reach Predictions %			
	Actual Reach	**1 Week Diary Prediction**	**2 Week Diary Prediction**	**Sampling Error**
BBC Radio 1	56.6	59.0 (+2.4)	56.3 (-0.3)	±1.7
BBC Radio 2	34.5	33.2 (-1.2)	33.0 (-1.5)	±1.6
BBC Radio 3	12.7	10.4 (-2.4)	11.6 (-1.1)	±1.1
BBC Radio 4	25.1	27.2 (+2.1)	25.5 (+0.3)	±1.6
BBC Radio 5	16.0	14.1 (-1.8)	14.8 (-1.2)	±1.2
Atlantic 252	18.3	16.6 (-1.7)	16.7 (-1.6)	±1.3
Classic FM	18.9	17.1 (-1.8)	17.5 (-1.4)	±1.3
Trent FM	46.5	46.0 (-0.6)	43.1 (-3.4)	±5.9
BRMB FM	36.2	37.7 (+1.5)	35.8 (-0.4)	±5.1
Capital FM	43.5	43.6 (+0.1)	43.2 (-0.3)	±4.0
Capital Gold	34.2	30.3 (-3.9)	32.0 (-2.2)	±3.7
Power FM	38.0	36.3 (-1.7)	34.8 (-3.2)	±5.9
Clyde 2	41.0	45.4 (+4.4)	41.2 (0.2)	±6.9
Melody FM	22.3	20.6 (-1.8)	20.7 (-1.6)	±3.3
Jazz FM	13.0	10.3 (-2.8)	11.0 (-2.0)	±2.4
Kiss 100	13.4	11.7 (-1.7)	12.4 (-1.0)	±2.5
LBC Talkback	15.2	13.3 (-1.9)	13.8 (-1.4)	±2.8
LBC NewsTalk	11.9	9.4 (-2.5)	10.4 (-1.5)	±2.7
Other Stations	74.7	76.4 (+1.8)	73.6 (-1.0)	±1.5
Total Radio	97.4	99.7 (+2.3)	98.7 (+1.3)	±0.6

Table 3.2

	Model Bias and Prediction Error - Reach % Points				
	1 Week Diary Base		2 Week Diary Base		Samp.
	Ave. Diff.	Ave. Abs. Diff.	Ave. Diff.	Ave. Abs. Diff.	Error
7 National Stations	-0.6	1.9	-1.0	1.1	1.4
11 Local Stations	-1.0	2.1	-1.5	1.6	4.1
All 18 Stations	-0.9	2.0	-1.3	1.4	3.1

Table 3.3

	Model Prediction Errors - % of Average Reach		
	Average	**1 Week Diary Base**	**2 Week Diary Base**
	Reach	**% Ave. Abs. Diff.**	**% Ave. Abs. Diff.**
7 National Stations	26.0	±7.3%	±4.2%
11 Local Stations	28.7	±7.3%	±5.6%
All 18 Stations	27.6	±7.3%	±5.1%

Table 4.1

	Percentage Gains in Effective Sample Size			
	Two Week Diary		Four Week Diary	
	Ave. Hours.	Weekly Reach	Ave. Hours.	Weekly Reach
BBC Radio 1	5	12	10	21
BBC Radio 2	3	13	6	21
BBC Radio 3	3	30	4	49
BBC Radio 4	2	8	4	13
BBC Radio 5	11	31	20	56
Atlantic 252	14	14	21	25
Classic FM	10	21	20	38
Average National	**7**	**18**	**12**	**32**
Trent FM	14	11	31	22
BRMB FM	3	10	4	16
Capital FM	5	15	13	26
Capital Gold	10	20	12	34
Power FM	5	14	48	35
Clyde 2	6	4	10	9
Melody FM	4	17	7	29
Jazz FM	12	24	16	43
Kiss 100	11	22	14	41
LBC Talkback	3	10	5	15
LBC NewsTalk	10	15	13	30
Average Local	**8**	**15**	**16**	**27**
Other Local	**5**	**13**	**10**	**24**
Total Radio	**6**	**17**	**11**	**29**

Table 4.2

	Guideline Gains in Effective Sample Size			
	Two Week Diary		Four Week Diary	
	Ave. Hours	Weekly Reach	Ave. Hours	Weekly Reach
Average Station	+7%	+18%	+12%	+32%
Programme Led Station	+7%	+30%	+12%	+50%

Table 4.3

	Percentage Gains in Effective Sample Size			
	Two Week Diary		Four Week Diary	
	Ave. Hours	Weekly Reach	Ave. Hours	Weekly Reach
Adults 15+	6	17	11	29
Children 4-14	10	17	18	34
Adults 15-24	10	20	17	45
Adults 25-34	9	20	14	33
Adults 35-44	6	22	11	37
Adults 45-54	6	21	12	31
Adults 55-64	6	20	10	36
Adults 65+	5	13	9	23
Men	6	23	12	42
Women	7	18	12	31
Adults ABC1	5	18	10	31
Adults C2DE	7	15	12	27

Table 4.4

	Relative Sample Sizes Required for Equivalent Precision		
	One Week Diary	**Two Week Diary**	**Four Week Diary**
Average Station			
Average Hours	100%	93%	89%
Weekly Reach	100%	85%	76%
Programme Led Station			
Average Hours	100%	93%	89%
Weekly Reach	100%	77%	67%

Table 5.1

	Relative Costs Required for Equivalent Precision		
	One Week Diary	**Two Week Diary**	**Four Week Diary**
Average Station			
Average Hours	100%	+20%	+70%
Weekly Reach	100%	+10%	+40%
Programme Led Station			
Average Hours	100%	+20%	+70%
Weekly Reach	100%	0%	+20%

References:

American Research Bureau Inc. (1974), Arbitron Replication: A Study of the Reliability of Broadcast Ratings.

Engel, W.E. (1992), Radio Reach and Frequency Revisited After 25 Years: Do Quintile Probabilities Still Produce Accurate Estimates, Esomar/ARF Worldwide Electronic and Broadcast Audience Research Symposium, Toronto, June 1992.

Gray, Sue and Wilcox, Steve (1995), The UK Reach and Frequency Model, First Esomar Radio Research Symposium, Paris, July 1995.

Twyman, Tony and Wilcox, Steve (1996), The Variability of Audience Measurement Data and How to Live With it, ARF/ESOMAR Worldwide Electronic and Broadcast Audience Research Symposium, San Francisco, 1996.

The author

Steve Wilcox is Technical Director at RSMB Television Research Ltd. UK