

IMPROVED PLANNING AND ROI MEASUREMENT FOR NEWSPAPERS: ESTIMATING CAMPAIGN REACH & FREQUENCY OVER TIME

Andrew Green, Ipsos MediaCT and Heather White, News Limited

An essential role of readership surveys is to measure the rate at which advertising campaigns in the printed media build audiences over time. Most campaigns run for several weeks or months, so media planners need to be able to estimate with as much confidence as possible what proportion of their target population will be reached and how often after one, two or more weeks – depending on how many insertions they have run and when those insertions have appeared. In short, they need to know whether they are building reach or frequency as they add insertions to a campaign.

We know from countless studies¹ that magazines take time – often several months – to build their average issue audience as reported in readership studies. Monthly titles in particular can take upwards of six months to reach most of the readers they are ultimately likely to be read by, as they are passed on from reader to reader or picked up in public places like doctors' waiting rooms and so on.

Accounting for this time dimension is vital if the market mix models used increasingly by advertisers are to make any sense. In these models, advertising weight (sometimes, erroneously, measured in money terms, but more properly in GRPs or reach) is compared with sales performance over the same period (or a subsequent period). If the entire readership of every magazine on a schedule is assumed to accumulate in its week or month of issue (when a significant proportion of it will certainly *not* have taken place) then clearly the model will misrepresent the connection between magazine audience reach and sales performance. A marketer looking at this may well conclude that his investment in magazine advertising is not paying off when he should be looking at both readership build and sales performance over a far longer period.

Newspapers present a different challenge. Most titles have a relatively short lifespan. Few people will want to read a daily newspaper beyond its day of publication (although inserted magazines and Sunday newspaper sections are often retained for longer periods). Yet reading patterns may vary considerably. Some people, for example, may be loyal to a given newspaper and read it every day or most days. Others may be loyal to newspapers on particular days – perhaps always buying the Wednesday edition of a particular title, for example, because of its coverage of cars or a Friday edition because of a sports supplement. Others still may just dip in and out of several titles over the course of a week or month.

Capturing this disparate behaviour and representing it properly is one of the most important jobs of a readership survey. Ideally, the survey should be able to help users estimate how each newspaper builds readership over one or more weeks and to distinguish between the reach and frequency achieved for multiple insertions within a week or for the same number of insertions spread over several weeks.

The key variable in this calculation is known as readership 'turnover' or, in Australia, 'casualness' – the rate of readership build achieved by multiple additional insertions in a title over time. A publication with a low proportion of casual or occasional readers will tend to add readers only slowly as time goes on; conversely, one with a high proportion of casual readers will add new readers quickly from issue to issue. The longer the period of time under consideration, the greater the probability that unique or new readers will see an issue.

Multiple Approaches

As Richard & Frankel pointed out in their paper at the Montreal Symposium (although the paper was concerned with magazines)², there are several ways in which reading frequency – and thus casualness or turnover can be measured:

1. Empirically – with two or more interviews over a period of time, enabling reading to be measured directly;
2. By means of a frequency question (asking people to estimate how often they usually read a title) and then developing personal probabilities for each respondent from their answers to the question;
3. By including a screening question, as well as a question about average issue reading in a single interview.

The authors examined data from a large-scale US study and concluded that, for magazines at least, using a frequency question in a single interview tended to *under*-estimate the rate of readership build over time (i.e. that the number of casual or occasional readers was being under-counted), perhaps because people tended to forget some of their occasional or casual reading.

Over the years, a large number of papers at the Symposia have discussed alternative ways of asking the frequency question, debated the relative merits of using numerical or verbal scales and looked at the use of the frequency question as a way to estimate Average Issue Readership itself³.

In Erhard Meier's *Summary of Current Readership Research* in 2009, 65 out of the 98 surveys listed employed a general filter or screening question of some kind – e.g. asking whether respondents had read a publication in the past 6 or 12 months. This can be used to 'cap' the maximum likelihood of somebody reading any given title. Those screening out can effectively be assigned a reading probability of zero.

94 of the 98 surveys ask frequency questions of one form or another in a single interview. Some ask respondents about a specific period in the past; others ask about ‘usual’ behaviour. Some ask about the number of issues they read out of four; others about how many out of the last 5, 6 or 7 issues. Some use verbal scales, some numerical, others mixed. And so on.

At the 2001 Symposium in 2001, Gary Morgan⁴ re-visited some of the conclusions drawn by Richard & Frankel, arguing that the “Gold Standard” approach was to ask the question in two separate surveys. He also made a number of other points in the paper, including:

- Supporting Richard & Frankel’s finding that use of a frequency question in a single interview tended to underestimate the rate of reach build by under-counting casual readers;
- That asking people about how many out of the last two, rather than the last four issues they read (0,1 or 2) would generate a more accurate estimate of the proportion of casual reading; and
- That it was important to distinguish between ‘within-week’ and ‘between-weeks’ casualness for daily newspapers. In other words, knowing how often people read a title during any particular week will not necessarily predict how often they read on a week to week basis.

Based on experience with his own survey in Australia, Morgan argued that to measure casualness, it was vital to capture two separate reading occasions for each respondent in the sample. In the case of the Roy Morgan Single Source study, respondents are interviewed face-to-face about their reading of newspapers and magazines and then asked to complete a leave behind ‘diary’ the following week which captures reading behaviour each day of the week.

Because the second survey is completed fairly soon after the face-to-face survey, it is arguably most applicable to newspaper and weekly magazine reading, as only limited additional data on monthlies is likely to be forthcoming.

In practice, there is a question mark over response bias – although because details of samples and response rates are not currently released to the market in Australia, it is impossible to draw any firm conclusions. Face-to-face response is known to be very low in the country and, on top of this, less than half of the initial respondents are thought to return the second survey, compounding the problem. It is not known to what extent bias exists or is compensated for in the second survey and whether that bias would affect the accuracy of the results. The fact that the first survey is face-to-face and the second self-completion may also result in modal effects that have never been explored publically.

All this said, Morgan’s paper did draw attention to the expense of administering two surveys to each respondent. As it is obviously cheaper to run only one survey, his company sought to find out if the results of a two-interview study could be replicated in a single survey. He proposed a new frequency question which, he opined, minimised some of the problems inherent in the more commonly used frequency questions. In these latter, respondents who claim to read either zero or one out of four issues tend, the argument goes, to *under*-estimate their frequency of reading; those claiming to read four out of four issues tend to *over*-estimate their frequency.

This line of thinking is based on observations (attributed to American researcher Willard Simmons from a paper published in 1969) that a lot of people in the habit of reading all or most issues of a publication will report that they read every issue (4 out of 4) even though they sometimes only read 3 out of 4 issues.

The lightest readers, on the other hand, may report reading none out of the last four issues when in fact some of them did see at least one of the issues. In other words, people often report more consistent behaviour than might actually be occurring. If true, both of these tendencies will lead to an under-estimate of the number of casual readers and therefore to slower reach build over time than is in fact the case.

The new question restricted people to answering about how many out of the last two issues they read – with the choice of zero, one or two. Arguably this was an easier time period for people to recall accurately and Morgan’s tests showed the results of the question to be closer to those of the two-interview procedure. Details of the tests were not published.

The Ipsos/TRW Validation Study

In 2010, Ipsos was awarded a contract by The Readership Works (TRW), an organisation in Australia consisting of representatives of newspaper publishers, magazine publishers and media agencies, to create a new readership survey, which is due to be launched in 2012.

As part of the pilot work preparing for the survey, we decided to test different ways of estimating reach and frequency build from a single survey specifically for the major metro newspapers and to compare the results of these different approaches with actual reading behaviour measured using a diary.

The current approach to measuring newspaper readership in Australia, as in many other countries, is to ask separately about Monday-Friday, Saturday and Sunday titles. For weekday publications, a slightly different method to that used in other countries has been adopted in the market.

For titles screened in by the respondent, a question is asked about reading ‘yesterday.’ They are then asked separately about each of the four weekdays prior to yesterday. So for example in a Friday interview, respondents will be asked to indicate reading of any weekday newspapers ‘yesterday’ (Thursday) and then, in turn, about reading of the title on Wednesday,

Tuesday, Monday and on the previous Friday. This enables a direct measure of reading frequency in the most recent week to be determined.

These questions are followed for each newspaper by asking about the most recent occasion *apart* from the past seven days when the publication was read or looked at – whether it was one, two, three or more weeks before the latest week. For whichever period is chosen, respondents are asked to estimate how often they read the title in question during that particular week (the question is not asked if reading was more than four weeks ago).

In other words, instead of being asked to fill out a second questionnaire in the week following the main interview, respondents are asked to try to remember behaviour from one or more weeks beforehand.

We decided to validate this questionnaire sequence by comparing the reach and frequency estimates it generated with the approach more commonly employed around the world of asking people simply to estimate how often, in general, they read particular newspapers. Both approaches were then compared with the reading that actually occurred.

The frequency question tested was as follows:

In general, for each of the following Monday to Friday newspapers, which of the following best describes how often you, yourself, look at or read a printed copy?

- 1 Always (5 out of 5 issues)
- 2 Most days (4 out of 5 issues)
- 3 Several days (3 out of 5 issues)
- 4 Some days (2 out of 5 issues)
- 5 Occasionally (1 out of 5 issues)
- 6 Less often than this

Actual reading behaviour was measured by asking a little over 1,000 adults to complete a weekly reading diary over a period of 6 weeks beginning on 13th June 2011 and ending on the 3rd August 2011. Once they had done this, respondents were asked to replicate the standard newspaper readership questions by recalling their reading behaviour from the previous 4 weeks. This included a day by day record of the most recent week and questions about reading frequency in prior weeks.

Figure 1: Diary Questionnaire Day by Day for Last Week - Monday to Friday Titles

Ipsos Let's start with newspapers that come out on Monday through to Friday. For each of the following printed Monday to Friday newspapers, did you, yourself, look at or read a copy on each of the following days?

(this includes glancing at or flicking through any part of a printed newspaper, anywhere, for any length of time, any copy, old or new) Please select a single response for each publication on screen.

Please review the dates below carefully as you complete your response, as they are in reverse date order.

	Friday 17th June	Thursday 16th June	Wednesday 15th June	Tuesday 14th June	Monday 13th June
Masthead A A Monday - Friday	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>
	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>
	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>
Masthead B B Monday - Friday	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>
	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>
	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>
Masthead C C Monday - Friday	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>	<input type="button" value="Yes"/>
	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>	<input type="button" value="No"/>
	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>	<input type="button" value="Not Sure"/>

Your progress in the survey: %

Diarists were drawn from the Ipsos MyView panel to represent people over 14 years old living in the one of Australia’s main states. 1,573 respondents were recruited, 554 of whom dropped out over the course of the study (see Table 1).

Table 1: Response Rates to Weekly 6 Week Reading Diary

	Completes	Dropped Out	Attrition %
Screenener	1573	0	0.0%
Diary 1	1366	207	13.2%
Diary 2	1253	113	8.3%
Diary 3	1171	82	6.5%
Diary 4	1110	61	5.2%
Diary 5	1069	41	3.7%
Diary 6	1019	50	4.7%

Results

Can People Recall Newspaper Reading Accurately?

There are many ways we can look at the data. In this paper we concentrate on the results for weekday newspapers – but we will also include some observations on weekend newspapers in our final presentation.

Overall observations include the following:

- Almost without exception, all the techniques which rely on an element of recall *under*-count the most casual readership (those reading 1 out of 5 issues) compared with what we know from the diary. This was apparent even when people were asked about the most recent seven days, but got progressively worse as time went on – with a lower and lower proportion of people claiming to read 1 out of 5 issues (though we knew they were doing just that)
- In most cases, a greater proportion of people claimed to read 2 out of 5 issues than was apparent in the diary – but this was insufficient to make up for the general shortfall in 1 out of 5 issue readers
- The proportion of respondents claiming to read 3 out of 5 issues was appreciably higher than diary data indicated (possibly a symptom of the common market research problem with the middle of the scale?)
- Diary data showed that reading 4 out of 5 issues of weekday newspapers was by far the least common frequency. Yet it was a common option picked by respondents when asked about it later.
- The standard frequency question asked in most surveys worldwide generated *lower* numbers of people claiming to read 5 out of 5 issues than diary data indicates is the case; which goes against the hypothesis out forward by Simmons in the 1960s and referenced by Morgan. It therefore – in combination with its under-estimate of casual readers – points to lower reach build for daily newspapers than is in fact the case.
- The more detailed and lengthy day by day readership question asked in Australia generates the most accurate estimates of reading frequency for the most recent week; however when asked about periods prior to this, memory effects are apparent.

Not every title experienced the same effects in the same way. In Table 3, for example, we take a look at what people remembered about their reading in the weeks prior to the most recent week (for the most recent week, as noted above, we ask detailed readership questions for each day of the week). The table compares the newspapers people recorded having read in their diaries in particular weeks with what they remembered having read when asked about those same weeks - after the final diaries were returned. Responses are compared using a numerical scale for weekday newspapers.

Recall of behaviour in more distant periods (i.e. 21-28 days before the interview) showed the greatest differences from the diaries. The most regular readers in our study (those reading every issue) tended to remember less reading than they had actually done. And this under-stating got more pronounced the longer in the past they were asked to remember.

People also failed to remember casual reading, even when it was quite recent. As Table 3 also shows, the proportion of respondents saying they had not read a title at all in a given week (or were not sure, which we classify the same way) was far higher than the reading diaries had indicated, especially when we asked them to remember their behaviour further back in the past. This finding was consistent across all titles examined.

So people, according to our diary findings, under-report their most regular reading, but also under-report casual reading (by claiming not to read the title at all, even though they stated in their reading diary that they *had* read the title in question).

Table 3: Frequency of Reading Weekday Newspapers: Diary v Recall Method

	1-2 Weeks Ago		2-3 Weeks Ago		3-4 Weeks Ago	
	Diary	Recall	Diary	Recall	Diary	Recall
Title A	%	%	%	%	%	%
5 out of 5	24	21	26	21	26	8
4 out of 5	5	3	3	3	4	2
3 out of 5	6	6	6	6	5	3
2 out of 5	10	6	10	7	9	4
1 out of 5	14	8	15	9	14	8
Not sure/ Not Read	42	56	40	54	41	75
Title B	%	%	%	%	%	%
5 out of 5	8	6	8	6	8	2
4 out of 5	1	1	1	2	1	1
3 out of 5	2	2	3	2	3	1
2 out of 5	7	4	5	4	5	3
1 out of 5	9	7	10	6	12	6
Not sure/ Not Read	73	80	73	80	71	87
Title C	%	%	%	%	%	%
5 out of 5	1.1	0.8	0.7	1.0	1.1	0.4
4 out of 5	0.4	0.3	0.6	0.3	0.1	0.5
3 out of 5	0.5	0.5	0.5	1.0	0.4	0.5
2 out of 5	0.8	0.3	1.2	0.5	1.4	0.3
1 out of 5	3.2	1.5	3.8	1.6	2.9	1.4
Not sure/ Not Read	94	97	93	96	94	97

All this occurred even though there was likely to be some sort of conditioning effect on our sample: the process of filling in a reading diary week after week having probably at least heightened respondents' awareness of their reading behaviour.

We see this in Table 4, where we compare peoples' recall of their reading frequencies before and after they embarked on the six-week diary. But we also see that there is a definite decay in memory between the most recent and most distant weeks asked about – which sees fewer and fewer of them claiming to read every issue of a title and many more saying they didn't read it at all.

Table 4: Frequency of Reading Weekday Newspapers: Recall Pre- and Post-Diary

	Recall Prior to Diary			Recall After Diary		
	1-2 weeks ago	2-3 weeks ago	3-4 weeks ago	1-2 weeks ago	2-3 weeks ago	3-4 weeks ago
	%	%	%	%	%	%
Title A						
5 out of 5	21	21	5	21	21	8
4 out of 5	10	9	4	3	3	2
3 out of 5	8	8	3	6	6	3
2 out of 5	9	10	4	6	7	4
1 out of 5	8	9	7	8	9	8
Not sure	2	3	2	4	3	3
Not Read	42	40	76	52	51	73
Title B						
5 out of 5	7	6	2	6	6	2
4 out of 5	3	3	1	1	2	1
3 out of 5	3	4	2	2	2	1
2 out of 5	5	5	3	4	4	3
1 out of 5	6	6	5	7	6	6
Not sure	1	2	2	2	2	1
Not Read	75	74	86	78	79	87
Title C						
5 out of 5	0.6	0.6	0.1	0.8	1.0	0.4
4 out of 5	0.2	0.4	0.3	0.3	0.3	0.5
3 out of 5	0.3	0.7	0.5	0.5	1.0	0.5
2 out of 5	1.0	1.2	0.8	0.3	0.5	0.3
1 out of 5	1.2	1.6	1.5	1.5	1.6	1.4
Not sure	0.0	0.3	0.1	0.2	0.3	0.4
Not Read	97	95	97	96	95	97

Multi-Week Reach - Recall versus Diary

For all newspapers the diary captured more incremental multi-week reach than the last 4 week reading recall question did. The degree of this incremental reach build varied from title to title. Title C, the title with the lowest market penetration, benefited most from the diary method, as reach over 4 weeks was double that of the recall method.

Table 5: Incremental Reach Build from Week 1 -4

	Title A	Title B	Title C
Recall	22%	33%	50%
Diary	24%	40%	100%

In all cases depicted in Figures 2 - 3, while the accumulation curves follow a similar pattern, accumulated reach is higher for the diary.

Figure 2: Incremental Multi-Week Reach – Title A

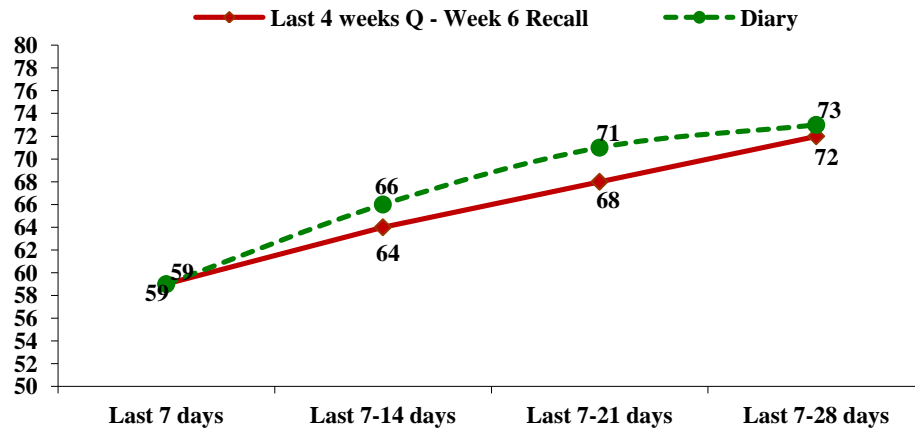


Figure 3: Incremental Multi-Week Reach – Title B

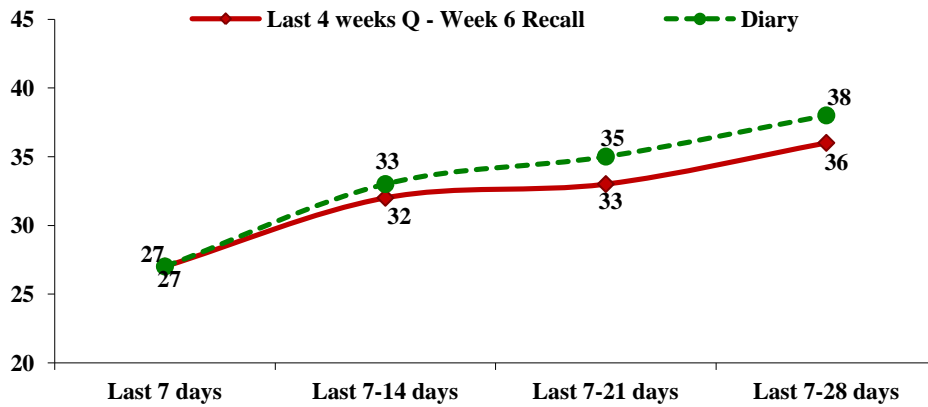
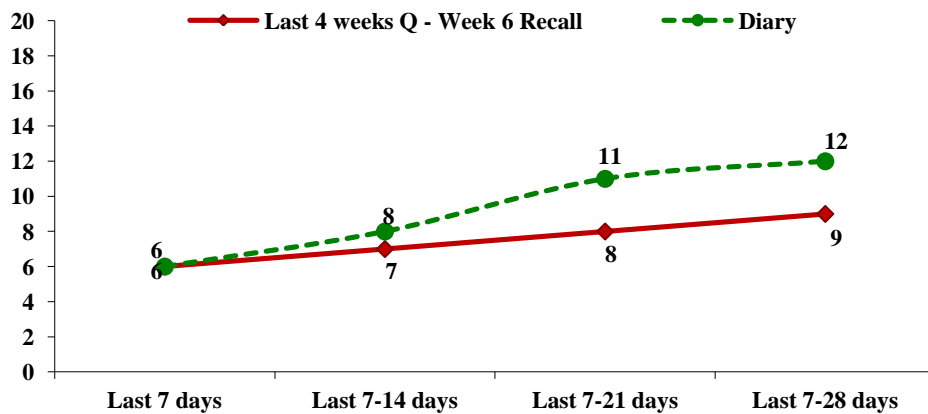


Figure 4: Incremental Multi-Week Reach – Title C



Discussion

Readership research – like a lot of other market research – depends on respondents being as accurate and truthful as they can be when answering questions. It is our job to try to make their task as straightforward and unambiguous as we can.

But memory is rarely perfect and it is very natural with an activity like newspaper reading for people to try to report what they ‘usually’ do rather than what they might have done at a particular period in time.

Many people, for example, buy the same newspaper regularly from a newsstand and read it every day (in markets like the United States, much of a newspaper’s circulation is sold via subscription, so different patterns of readership are likely to occur). Or at least that is how they would characterise their behaviour. In fact, they may miss the odd issue for a variety of reasons and either read no paper or read or buy a different one.

From time to time they may read a colleague’s newspaper or pick one up in a waiting room or on the seat of a train. Again, because this is not part of their ‘normal’ behaviour, they may forget it.

There are other people who do not have strong loyalties to any particular title, but prefer to choose from amongst several over the course of a year. Others still may have preferences for particular editorial sections or columnists which lead them to buying certain titles on some days and other titles on other days.

Whatever the particular circumstances, it is always going to be difficult to achieve perfect accuracy in a readership survey depending, as we do, on peoples’ memories and circumstances at the time of the interview.

The underlying assumption behind much of what we do has been to assume that if we ask enough people about their reading, ‘outliers’ in their responses will tend to cancel each other out so that, in aggregate, we can present a fairly accurate picture of reading behaviour for a country or market.

Another fly in the ointment is that of economics. The perfect readership survey may be unattainable simply because of resource constraints. A good example is that of the two-interview survey highlighted by Richard & Frankel at the Montreal symposium and again by Morgan in Venice. Combined with economics is the whole challenge of response. Every year it is getting harder to persuade people to take part in surveys. The longer and more complex the survey, the less easy it will be to persuade people to take part.

Face-to-face interviews – traditionally seen the Gold Standard of readership research – are becoming increasingly difficult and expensive to administer.

In practical terms, most readership research is going to have to be contained within a single interview. But we find the simple frequency question used in most surveys around the world to be wanting. According to our data, a simple frequency question about peoples’ normal reading behaviour will under-report both casual reading behaviour and the most regular reading behaviour (the proportion reading 1 out of 5 issues and 5 out of 5 issues).

It also fails to give us enough information to be able to build different kinds of newspaper schedule – e.g. telling us the difference between placing 3 insertions in a single week or spreading them over (say) 3 weeks.

Conclusions

Unfortunately, it is not practical or economical to re-create diaries on a regular basis. So to create accumulation curves for the various titles, we need to look at some measure of ‘ever read’ (e.g. past 12 months), at day by day reading for the most recent week and at peoples’ recall of how often they read newspapers in the recent past – but not going too far back. These are the basic building blocks for an effective reach and frequency approach.

It is also vital in the Australian market, where readership (and very soon circulation) data is published by day of week, to be able to calculate reach build for individual days of the week; this goes beyond projecting weekday average AIRs.

¹ E.g. Shepherd-Smith, N (1988). *Taking Account of the Time Factor*. Proceedings of the Worldwide Readership Symposium, Barcelona; Peeters, S., Debeer, V. & Lanckriet, T. (1999) *Magazines Need Time – The Build-up of Magazine Audiences Over Time*. Proceedings of the Worldwide Readership Symposium, Florence. Pincott, G (1991). An Examination of Wear-out and Exposure Over Time. Baim, J., Frankel, M., Agresti, J. & Becker, R. (2001). *Magazine Audience Accumulation: Discussion of Research Issues, Modelling and Application*. Proceedings of the Worldwide Readership Symposium, Venice.

² Richard, A & Frankel, M (1983). *A Comparison of Reach & Frequency Estimates: Single Versus Dual Interview Approaches*. Proceedings of the Worldwide Readership Symposium, Montreal

³ There are too many papers to cite here, but the PDRF website (<http://www.printanddigitalresearchforum.com>) has a comprehensive list. Amongst the more ‘seminal’ are Cornish, P. (1995). *The Frequency Scale – Problems and Priorities*. Proceedings of the Worldwide Readership Symposium, Berlin; Douglas, S. and Jones, Richard (1995). *Frequency of Reading for AIA: Under-Researched?* Proceedings of the Worldwide Readership Symposium, Berlin; Tennstädt, F. And Hansen, J. (1981). *Frequency of Reading – Allensbach’s Point of View*. Proceedings of the Worldwide Readership Symposium, New Orleans and Brink, C. & Napior, D. (1997). *Calibration of Frequency-of-Reading Scales*. Proceedings of the Worldwide Readership Symposium, Vancouver.

⁴ Morgan, Gary, Levine, M. & Dorofeev, S. (2001). *A New Method to Measure Media Casualness for Magazines and Newspapers*. Proceedings of the Worldwide Readership Symposium, Venice