

THE CHOICE OF READERSHIP MODEL: NOW AND IN THE FUTURE

Donald Monk

Introduction

This contribution is broadly concerned with current methods of measuring print readership and possible developments over the next few years.

Developments in print media research are difficult to predict. The problem is not simply that of not knowing what new technologies will become available, or even how technologies such as the internet may have an impact on the nature and economics of print media generally. There is an additional factor: namely, that changes are dependent on their acceptance by publishers, media buyers and advertisers. In the past there have been occasions when the decisions of these important players have puzzled some media researchers.

Some reluctance to change is understandable. Consistency of data over time is preferred by both buyers and sellers. However, increasing demands for more media-market data, new types of publication, and mounting pressure on existing survey methods, may mean that changes in data collection method and the media measurement model will be considered for some media surveys during the next few years.

In this paper I briefly review the developments in media models over the past twenty years, in order to assess our current options if any changes are contemplated. Next, to illustrate the conflicts that can occur between theory and practice, I review a specific example of a recent change to an industry survey, namely, the MMMB survey of doctors in Canada, and some of the problems that can arise. Finally, some conclusions are drawn.

Readership Research Development Over the Past Twenty Years

Over the past twenty years there has been a considerable increase in both the amount of readership research carried out and knowledge of the dynamics of readership research. Until the 1970's media surveys were isolated. Each country conducted its studies largely in ignorance of what went on elsewhere.

The situation changed with the increase in advertising expenditure by multi-national companies, placed by international advertising agencies, which brought a more global perspective to the committees managing the various syndicated media measurement services. Also, the media researchers from various countries began to communicate with each other. A major catalyst for this has been this series of international symposia held about every two years. Since the first symposium, held in New Orleans in 1981, the International Symposia for Readership Research have become the must attendance for all senior media researchers, and the chosen venue for presenting professional papers.

This document will not attempt to summarize all of the topic areas that have been discussed over the eight symposia held to date. However, it will look at what has been learned, and what is generally accepted, in areas relevant to choice of method, particularly when some change is being considered.

Readership Models

At the time of the first Symposium there were really only two readership models, recognition and recency, although each had its variants. There was considerable debate as to which method represented the truth. However, the debate lacked sophistication: the main strategy for each advocate was to present evidence as to why the other system was wrong.

Later Symposia showed more sophistication. There was a recognition that validity was a desirable but difficult goal, and that for the immediate future the objective should be reliability and comparability. The latter meant that, if there was some over or under estimation, it should be fairly constant over all publications.

One immediate question was whether the differences that might exist between recognition and recency were constant over the various groups of magazines or tended to favour one group over another, thus showing a lack of between model comparability. In the United States this was a real issue because of changes that had recently been made by one of the major syndicated services. This led the ARF to sponsor a major comparability study in the form of a large side by side controlled experiment.

The results were dramatic. On average Recent Reading results compared to TTB results were:

27% higher for weeklies,
80% higher for large monthlies,
96% higher for small monthlies.

The data were also examined by the two components of readership: the percentage of people who screen in as having looked at a copy in the past six months (Screen in rates), and the percentage of those people who qualify as readers (Read to screen). In the study the Recent Reading Screen in rates were about 10% higher, and the Read to Screen rates were some 50% higher. Thus the main cause of the higher Recent Reading results is at the stage when the respondent has to indicate whether they have read the publication within the defined periods (last 7 days, last month, etc.).

These results led to a concentration as to why Recent Reading gave higher results. The concepts of replication and parallel reading were explored. Defenders of Recent Reading could not deny their existence, but settled behind the rampart of them canceling each other out. However, more and more evidence emerged that replication was the bigger factor with most publications, and the issue of TV program guides (high replication and practically no parallel readership) was difficult for them to defend.

The other issue examined was the very difficult task that Recent Reading presents. The psychology of memory, evidence from studies of more verifiable events (such as visits to the doctor), and readership experiments, all pointed to telescoping being a serious issue among non-regular readers of publications. As Val Appel wrote in his paper: *Telescoping: The Skeleton In The Recent Reading Closet*, "As a result, not only does the Recent Reading method produce spurious audience estimates, it does so in a way as to seriously disadvantage weekly publications relative to monthlies".

For very pragmatic reasons this did not deter syndicated studies continuing to use Recent Reading. The marketplace demands coverage of some 150 magazines plus extensive market data. This can not be achieved practically by recognition methods. Therefore, although the comparability across types of publication was not met, Recent Reading was thought to be the only viable option.

By the early 1990's the possibility of using frequency as a direct measure of readership, rather than simply a method of computing turnover for media scheduling models, began to be discussed. Initially it was disparaged. Papers were written to show that people misinterpreted the question, or interpreted it in varying ways, or that it was difficult to find a form that dealt adequately with all publication frequencies. However, by the 1995 Berlin Symposium sentiment was more positive. Among several papers addressing Frequency, that by Douglas and Jones "Frequency of Reading for AIA: under researched?" showed how mail surveys using frequency had become well established within the United States for special markets. They also demonstrated that there were many variations of the Frequency question. By the last Symposium, held in Vancouver in 1997 the situation had further changed. Speakers from Scandinavia reported on successful applications within major media surveys, and the issue of irregular publications began to be aired.

It had long been recognized that Recent Reading is an inappropriate method for measuring irregular publications. However, within the greater scheme of things (a.k.a. as the needs of major consumer magazines) it has not been a pressing issue, and it has been largely ignored. But this was not the case with more specialized readership audiences such as farmers, doctors, financial executives, etc.. In these areas irregular publications are more prevalent, and often major players. Therefore, the readership model used had to be relevant to both regular and irregular publications. As Recognition by face-to-face interview was not a contender on financial grounds, and Recent Reading is an inadequate model for irregular publications, frequency looked an attractive option.

It is possible that irregular publications may become more prevalent, particularly if publishers know that the measurement system will not penalize their publication simply because it is irregular, or the Internet reduces available advertising revenue. Also, there is no question that a Recent Reading survey will penalize irregular publications. A more detailed discussion of this is given in Appendix 1.

In many earlier papers the term 'frequency' was used as if there was only one format or type of frequency. However, at the Vancouver Symposium Brink and Napier made explicit the two different variants. In their terminology, the frequency questions that have been typically used to provide empirical probabilities for media models are "rate of occurrence" estimates, where the respondent is asked to provide average or usual estimates (e.g. "Out of every 4 issues, how many do you usually read"). However, when the intention is to measure readership the "episodic counting" estimates have usually been used, where the respondent is asked directly to say how many of the last 4 issues have been read.

A Specific Example of Change (the MMMB Survey of Doctors)

So much for theory. Now let's look at a particular example of change. The one I have chosen is the Canadian MMMB survey of physicians.

For a number of years the syndicated readership study of doctors in Canada (MMMB) used the cover recognition method conducted by face-to-face interviewing. By 1995 the increasing number of titles to be measured, the wish to cover doctors in various medical specialties, the desire for larger samples, and the high cost of face-to-face interviewing for both rural and urban coverage, led to the decision to design a new survey that would commence in 1997.

A review was commissioned from an experienced media consultant to suggest ways of improving the then current MMMB. After reviewing that report MMMB decided that the entire study should be rethought and the consultant was asked to produce a more thorough and extensive review that would set out all the options for a new study.

That report was presented in March 1996. This comprehensive review included an assessment of Medical Surveys around the world, with a summary and recommendations, a discussion of critical issues, and a draft specification for a new MMMB mail survey. When commenting on readership models it was reported that frequency is used in a number of niche studies in North America; that at the ARF Print Readership Workshop there had been general agreement that the benefits of frequency were that it was inexpensive, easy to administer, and most applicable to niche studies; and that in terms of accuracy (comparability?):

TTB advantages weeklies,
RR advantages monthlies,
Frequency advantages weeklies less than TTB and monthlies less than RR.

He finished his review of models by concluding that frequency is now regarded as an acceptable technique, with some advantages for mail surveys and niche markets.

There was one serious omission in the consultant's otherwise very competent review. That is the issue of irregular publications and the impact that it might have on the evaluation of suitable or appropriate media models. Apart from one passing reference the matter is ignored. The research literature has many instances of recent reading being unsuitable for the measurement of irregular publications, and some media services that use RR, such as MRI in the U.S., employ frequency for very irregular or seasonal magazines. Of significance is that the publications to be covered by MMMB have a greater proportion of irregular magazines with significant circulation than is the case for most other media surveys. Therefore irregularity has even greater importance.

A second issue that was ignored is the suitability of Recent Reading to a survey being conducted by mail. When using personal or telephone interviewing considerable care is taken to achieve day of week, and week to week, balance. This is much more difficult to achieve with a mail survey. One advantage of frequency is that it is not as time bound as recent reading and therefore is less affected by any lack of balance.

After extensive discussions, based on a review of the options produced by the consultant, the MMMB Board nevertheless opted to use Recent Reading. A specification was produced and sent to several research companies for proposals and bids. The accepted format for the new survey was:

- a) Readership model: Standard Recent Reading.
- b) Interviewing Method: Mail survey.
- c) Publication representation: Cover shells in colour.

The sequence of questions (asked vertically for each publication) initially asked a receipt question as Q.1, but since 1998 the sequence has been:

- Frequency (six items, verbal with some numerical clarifications)
- Recency (two items: qualifying period and longer ago)
- Amount read - last issue read.
- Time spent.
- Amount usually read.
- Rating of publication (four items)
- Then:
- Demographics and some prescribing habits.

There are circa 40 publications, represented within a 16 sided (eight pages) questionnaire. Some 10 of the publications have irregular publishing schedules.

When the 1997 study was released the audience data for many titles showed considerable change from the previous data obtained from the 1996 Recognition survey. In relative terms general monthly and bimonthly magazines improved and weekly and biweekly publications deteriorated. Also, irregular publications, some of which were major publications, showed sharp declines. Not surprisingly this situation led to criticism of the new survey by some publishers. In order of importance the main complaints were:

1. An inappropriate readership model had been used, particularly as some of the major publications had irregular publishing schedules.
2. The cover shell was an inadequate representation method.
3. The response rates of well below 30% were inadequate.

While I believe that there is substance in the last two concerns, in this paper I will address only the first one - the model chosen for the new study and the special problem of irregular publications.

The Appropriate Model Issue

In terms of assessing the validity of readership models at least two philosophies have been articulated. The first, supported by MMMB, essentially says that valid readership is what the particular survey produces. For example, in defense of their new data the MMMB Technical Appendix states that they consider readership as an operational definition dictated by the exact form of questioning used for a particular study. Thus, to MMMB, in an operational sense all surveys are correct if the execution of the survey is good.

The other view, towards which I lean, is that while agreeing that any readership score is an operational result and that no model provides some absolute truth, the choice of model should place emphasis on comparability across different types of publication, and that a severe penalty for any one type should mean the exclusion of that model, no matter how well the survey using it is conducted.

Some data can be constructed that looks at the issue of comparability across types of publication. The 1996 MMMB, which used Recognition in a personal interview, also asked a frequency question of the episodic counting type. The 1998 MMMB survey used Recent Reading in a mail survey, and also asked a frequency question. The frequency question was of the rate of occurrence type, and had an additional "Occasionally" category.

Below are shown, for 20 publications common to both surveys, and based on General Practitioners, the readerships from the two surveys, either direct (recency or recognition) or derived by applying theoretical probabilities of 1, 0.75, 0.5, 0.25 or 0.0. For the 1998 Occasionally category 0.1 has been used. The data are shown indexed on the 1996 Recognition data = 100.

	Weekly	Bi-weekly	Monthly	8 - 10/yr	6 - 5/yr	4/yr
96 Recognition	100	100	100	100	100	100
96 Frequency	107	105	106	105	103	113
98 Recency	77	81	106	57	55	134
98 Frequency	73	73	92	51	43	89

The 96 Recognition data have been used as the base index (100) for two reasons. Firstly, while not practical in today's market, the Recognition method, when asked of a limited number of titles within a face to face interview, has been generally recognized as closer to reality than other methods, except for full TTB. Secondly, they also represent what was the media reality prior to the change to Recency. The data show that the 1996 episodic counting frequency question provides readership estimates some 5 per cent higher than the Recognition readership, and that margin is reasonably constant over publication groups. On the other hand, the 1998 Recency and 1998 Frequency based on "rate of occurrence" frequency provide a more erratic set of measures.

Another thing that appears surprising is that only among monthly and quarterly publications does Recency give higher readerships than Recognition. However, they are the only pure regular groups. The weekly and bi-weekly groups both have two publications with the larger of the two being irregular. Is this yet a further indication that Recency is an inappropriate model for irregular publications?

Logistical Problems

One potential way of avoiding the problem of irregular publications is to restrict the data collection periods to those times when all publications are in a regular publishing mode. However, this can create its own problems.

Most major readership surveys cover at least 50% of the year. The previous MMMB had been almost continuous throughout the year. And the MMMB consultant had recommended continuous data collection so as to avoid hyping. Yet the 1998 MMMB was restricted to 11 weeks - five weeks in Spring/Summer and six weeks in the Fall. Even this very limited period did not fully meet the objective, as it covered a non publishing period for one of the major irregular weeklies. Theoretically this can reduce the published readership of that publication by 9%. The impact of a single missed publication can be that large because the overall interviewing period is so severely curtailed. There are, of course, other reasons why such a constricted data collection period is inappropriate for a major survey that purports to represent total readership of physicians.

In further recognition that irregular publications can cause difficulties for Recent Reading MMMB has stated that when necessary they could either collect data with extended categories (i.e. past week, past two weeks, etc.) and make the necessary adjustments based on completion date, or use the recommendations of Dick Lysaker as to when frequency should be used as a substitute measure.

The first method is logistically complicated. It has to deal with vagaries of publication delivery and the return of questionnaires, and also is essentially a future intention rather than a potential current remedy, in that the necessary data have to be collected in a future survey before it can be applied.

The Lysaker solution is not an established procedure. When the Canadian PMB was contemplating a change to Recent reading, but was concerned with the impact on some irregular publications, Dick Lysaker was asked to provide a solution. His response essentially set out the criteria for when a publication should have frequency substituted for recency to derive its readership.

Once the survey data became available the MMMB decided that it was not necessary to apply the Lysaker Method to even those irregular publications that met the Lysaker criteria, because, in their view, there was no evidence that readership had been influenced by not publishing. The evidence supplied for this view consisted of some data in index format. However, I believe that their conclusion may be incorrect. An explanation for this view is given in Appendix 2.

The suggestion that Recent Reading was the wrong choice for MMMB is not intended to be a general indictment of Recent Reading. The model has served media research well in many environments. However, Recent Reading does have problems in some circumstances, and in particular it is much less suitable when there are significant irregular publications and the data are to be collected by mail.

It is suggested that the Canadian MMMB change is an example of an inappropriate change being made in spite of the evidence that was available to them, both in the general literature and from their own consultant's report. How can such a situation occur?

One possible explanation is that media research is a somewhat arcane discipline that is not fully understood by publisher, agency, and advertiser representatives that make up the Boards of industry media research surveys. If such people are not fully comfortable with methodological issues they may find it easier to support what they feel is the most acceptable and common choice of their peers. Recent Reading is the most commonly used model throughout the world. This fact alone may have provided Recent Reading with considerable credibility. The differing contexts in which the model was applied would not be considered. An example of this is found in the Canadian Advertising Research Foundation's appraisal of the new MMMB survey. When commenting on the validity of the study the sole comment was "The Recent Reading method is widely accepted in North America and Europe as a valid technique for establishing readership". No reference was made to the data collection method or to irregular publications. For the appraisal authors (who were not print media research specialists) popularity seemed to be sufficient for uncritical acceptance.

Summary

One theme of this paper is that there is often a discrepancy between what meetings such as this discuss as best, or most appropriate, research practice, and what collections of people from advertisers, agencies, and publishers serving as Board members of media research operations actually decide to do. This has been illustrated by a particular instance in Canada. However, I suspect that there have been similar situations in other countries.

The second theme has focussed on the evaluation of frequency as a viable readership model. There seems to be an increasing body of evidence that suggests that frequency may be a reasonable compromise model for the measurement of average issue readership, particularly in situations when the use of Recent Reading seems inappropriate. The "episodic counting" frequency measure appears to be relatively neutral across the various publishing groups. Frequency also fits in with the demands of current media survey specifications: namely, the coverage of a large number of publications, allowing sufficient interviewing time to collect a significant amount of non-readership data, and using data collection methods other than the personal interview.

For several studies focussed on specialized universes, such as farmers, doctors, and financial executives, this has already been recognized and incorporated into on-going surveys. Partly this is because those special markets have had significant irregular publications.

Looking into the future it is possible that, even for major studies of the total population, irregular publications will become more significant, and personal interviewing will become less viable. If so, the appropriate frequency measurement applied in a mail survey may be worthy of consideration. Perhaps it is time for ARF, or some other organization, to sponsor another major comparability study.

Appendix 1. Irregular Publications

The likely impact of applying Recent Reading methodology to irregular publications can be examined theoretically, using various assumptions. One is to assume that replication and parallel reading cancel each other out, which is the same as assuming that people read the magazine on one occasion on the day of receipt.

To use an actual example, the MMMB study assumes that some publications with other than 12 issues a year should be treated as monthlies. Below is shown the actual number of issues in a year, the assumed number of issues, and the impact on readership if the data collection was evenly spread over the year.

Actual/ year	Assumed/ year	Impact
8	12	- 33%
9	12	- 25%
10	12	- 17%
11	12	- 8%

The impact will be reduced to the extent that the data collection period is chosen to minimize the impact of irregularity. However, this strategy means that the data collection period becomes constrained, and even then it is often logistically impossible to have a survey period that avoids all non publishing periods of the irregular publications.

Appendix 2. Readership and Irregularity

The MMMB Technical Appendix has stated that there is no need to consider adjustments for irregular publications (whether Lysaker or some other method) because analysis has shown that there has been no impact on readership. In support of this view they have released some data in indexed format and without identification of the eight publications.

The data made available consisted of all aggregate readerships represented as 100, and the maximum and minimum readership in any of the 7 weeks as an index of the 100. The data are:

Publication	Average	Min.	Max	Spread
A	100	68	121	53
B	100	90	125	35
C	100	50	181	131
D	100	93	109	16
E	100	92	108	16
F	100	81	119	38
G	100	91	131	40
H	100	0	135	135

The publications will vary in terms of irregularity during the 7 week period, and will vary in terms of size, which has an impact on the Standard Error of the percentage readership. Also, the average weekly sample is 169 doctors and I have had to assume that the interviews were evenly spread. On this basis I have calculated the S.E.'s at various readership levels and the equivalent index based spread between plus and minus 2 S.E.'s. This is shown below.

Readership level	2 Stan. Errors	Index Low	Index High	Index Spread
50%	7.7%	85	115	30
40%	7.5%	81	119	38
30%	7.1%	76	123	47
20%	6.2%	69	131	62

For any publication over a seven observation (weeks) period the probability of having one result at + 2 S.E.'s and another at - 2 S.E.'s is about 0.01 or 1%. Therefore with eight publications one would not expect any instances by chance.

As can be seen, if all the publications were at the 50% level 6 of the 8 would be outside the + or - 2 S.E.s spread. At the 20% level 2 of the 8 would be outside.

Although handicapped by not having all the data it would appear that, assuming that a number of the irregular publications have readerships above 20%, there is evidence of above expected variation, which may well be due to the irregular publishing of some publications during the seven week field period.

References

- Valentine Appel, "Telescoping: the skeleton in the Recent Reading closet",
Worldwide Readership Research Symposium, New Orleans, 1981
- Stephen Douglas & Richard Jones, "Frequency of reading for AIA: Under researched?"
Worldwide Readership Research Symposium, Berlin, 1995
- Cheryl Brink & David Napier, "Calibration of Frequency-of-Reading scales"
Worldwide Readership Research Symposium, Vancouver, 1997

