Miscellanea

Under the heading Miscellanea, essays will be published dealing with topics considered to be of general interest to the readers. All contributions will be refereed for their compatibility with this criterion.

Questionnaire Measurement of Drinking Behaviour in Sample Surveys

John C. Duffy[1]

0. Introduction

In the past decade the use of sample surveys in investigations related to alcohol consumption and problems in the UK has grown remarkably (Dight (1976), Wilson (1980), Plant and Pirie (1979), Cartwright et al. (1978), Kendell et al. (1983)). A result has been renewed interest in survey methodology applied to this field. Wilson (1981) in discussing methodological aspects of a large survey in England and Wales addresses issues similar to those considered by Mäkelä (1971) and reviewed by Pernanen (1974).

This paper is concerned with questionnaire measurement of alcohol consumption in sample surveys having this as a main aim. Apart from question wording effects and bias in the method of measurement such surveys may suffer from deliberate concealment, sample frame defects and problems in contacting heavy drinkers. In general it is estimated that survey estimates of total population consumption constitute from 40% – 60% of consumption known to take place from official sources such as taxation records etc.

[1] MRC Unit for Epidemiological Studies in Psychiatry, Royal Edinburgh Hospital and Department of Statistics, University of Edinburgh, UK.

1. Individual Alcohol Consumption

In any population survey the desired information is elicited from individual respondents. In the case of alcohol studies the data of interest concern the drinking behaviour and history of the sample members, and it is the complexity of these phenomena that gives rise to problems of measurement. Individuals differ in the types of beverage which they drink, the amounts of each beverage they drink on occasions of drinking, and their frequency of drinking. Further complications arise from within-individual considerations. The same person drinking the same beverage on different occasions may drink different amounts, while in consecutive weeks an individual may undertake different numbers of drinking occasions which may involve the use of different beverages. An attempt has been made by Alanko (1982) to provide a statistical framework for these sources of variation, but available data do not permit a detailed assessment of the degree of fit of the elements of his model.

While these considerations can be used as a guide to evaluation of strategies of measurement and question construction, they do not constitute an externally validated set of criteria. Hence the methods to be described will be compared on a heuristic basis.

2. Aims of Survey Measurement

It will be useful to assess each method in terms of suitability as a means of achieving the possible aims of survey measurement in alcohol research. For this purpose the following aims might be considered.

i) Disaggregation of consumption and aspects of consumption by subpopulations, defined by criteria other than consumption level. Examples of this aim include ecological studies comparing different geographical areas, studies aimed at estimating the consumption of different demographic groups etc.

ii) Estimation of the proportion of consumption by beverage type. This aim is usually subordinate in alcohol studies, and might be more relevant to market research investigations.

iii) Classification of respondents by individual levels of consumption. This would be important in investigating the relationship between individual experience of problems and consumption, and in the past has formed part of studies on the distribution of alcohol consumption.

iv) Study of the occasions and contexts of alcohol consumption.

v) Modelling individual behaviour and associated parameter estimation.

3. Quantity-Frequency Measurement

The Quantity-Frequency (Q-F) Index introduced by Straus and Bacon (1953) represents an early attempt to obtain information about drinking behaviour from survey respondents. Typically respondents are asked about their drinking of each of three types of alcoholic beverage, namely beers, wines and spirits. For each beverage type respondents are asked to check the usual frequency of drinking that beverage on a list ranging from "three or more times a day" to "never." They are also asked about the usual quantity of each beverage type

which they drink on each occasion of drinking that beverage.

In order to take account of variability in respondents' drinking habits, the index was modified to the Quantity-Frequency-Variability Index described by Cahalan and Cisin (1968). This involved altering the questions relating to the quantities of each beverage as follows:

For each beverage type which the respondent claimed to drink once a month or more frequently, the proportions of occasions on which five or more, three or four and one or two units were drunk were obtained by questioning, the available replies being "nearly every time," "more than half the time," "once in a while" or "never."

A further question concerning the frequency of drinking of any alcoholic beverage was also asked, and categorization proceeded on the basis of the reply to this question, together with the modal and maximum quantities of the most frequently consumed beverages.

The Volume-Variability Index was introduced by Cahalan and Cisin (1968) and attempted to classify each respondent by his average daily volume of consumption and, within several groups at particular daily volumes, to subclassify by the day-to-day variability of the respondents. The average daily volume was estimated by multiplying the frequency of consumption (expressed as drinking occasions per 30 days) by the modal amount of each beverage.

The response categories for Q-F based questions may be further elaborated, for example more recent practice in respect of determining the proportion of occasions on which various amounts of each beverage are consumed would use a wider scale of proportions, and an increased number of amount categories.

Although this method is widely used there are some difficulties with its interpretation in

terms of rate of consumption. Firstly, the respondents may understand the questions as attempting to elicit modal rather than average frequencies and quantities. If the within-individual (temporal) distributions of these quantities are positively skew then the averages will often be greater than the modes. Some indirect evidence on this question may be obtained from surveys using a different measurement approach. When respondents are asked about the typicality of their reported consumption in the week before interview the majority reporting untypical consumption assert that drinking in that week was more than usual.

A second difficulty concerns possible question wording and response category effects. Practical experimentation could lead to improved question wording and a fuller understanding of respondents' interpretation of the questions. Response category effects in this area are well known and discussed by Bradburn and Sudman (1982). Their conclusion is that these effects are best allowed for by providing very high response categories.

Finally a technical problem arises. If the averaging procedures implicitly used by the respondent in answering the questions relate to different time periods for establishing frequencies and amounts then some degree of bias will result in converting these responses to a rate of consumption.

It follows that in so far as Q-F measurement can overcome the above problems it could be used to achieve aims i) to iii) of Section 2. To address the study of occasions and contexts it would require to be supplemented by additional questions, since it is unclear to what extent the usual frequencies and quantities may be related in terms of occasions. Modelling the process of individual consumption would require information concerning the temporal variability of the quantities elicited by Q-F measurement, and is not currently feasible.

4. Description of Occasions

It is possible to distinguish several approaches to measurement, all based on the description of drinking occasions by the respondent. Two of these will be considered here, an approach based on last week's consumption, and a related method which aims to cover a time period which will include around at least four occasions of consumption per respondent.

a) Last week's consumption
For several years now standard practice in UK surveys has been to ask the respondents to provide detailed information about consumption on each drinking occasion in the last week. Respondents are first questioned about drinking on the day immediately before the interview, and then about each of the preceding six days in reverse temporal order. Interviewers are normally instructed to provide memory prompts to respondents who appear uncertain as to their activities on any particular day. Often supplementary questions about a longer period of past time are asked of respondents who report no consumption in the week before interview, but it is notable that most published reports exclude these respondents from the analysis. An alternative is to exclude only those who report a less than weekly usual frequency of drinking to an additional question.

This methodology is generally thought to decrease possible memory problems, and to allow less scope for question wording effects by minimizing individual interpretation of questions. Nevertheless the method is not without drawbacks and anomalies.

Between 20% and 35% of respondents report no consumption in the week before interview depending on the type of population being surveyed. For all practical purposes this represents a waste of resources, since their returns will rarely figure in any analysis.

At the same time a recent survey in Edinburgh (Waterton and Duffy (1984)) showed

that a very simple questionnaire concerning last week's drinking by day rather than occasion, and not using memory prompts, produced estimates of consumption as great as those elicited by more detailed reporting by occasions. If interviewers on standard surveys really are providing memory prompts and proceeding as requested then it appears that they may be spending time (and money) unnecessarily. Alternatively it is possible that in practice interviewers do not use extensive prompting and simply report no consumption on days about which the respondent is unsure.

This method is suitable for all the aims of Section 2, with the qualification that restricting attention to last week's consumption produces a low yield of information concerning those individuals who drink infrequently, and about the occasions and contexts of their drinking. In many surveys the major focus of concern is the association between alcohol-related problems and the rate of alcohol consumption, and so lack of information concerning infrequent drinkers (who will generally have a low rate of consumption) may not be a disadvantage.

b) Variable length time period

In populations with a widespread pattern of frequent drinking last week's consumption may be taken as a reliable means of estimating the drinking behaviour of the bulk of the population, assuming truthful responses. If the pattern of consumption is more sporadic this will not be so, and an alternative approach based on actual drinking over longer periods of time may be adopted.

A typical implementation of this method is to ascertain the usual frequency of drinking by the respondent, and use this information to determine the length of time period over which drinking behaviour will be recorded. In a Finnish version the time period is chosen so as to provide roughly four drinking occasions per respondent. This enables a rate of con-

sumption to be calculated simply and directly for infrequent drinkers. The method also facilitates construction of measures such as frequency of intoxication for all respondents directly from reported drinking behaviour.

While it is relatively simple and speedy to enquire about respondents' behaviour over the past week day by day, and to provide memory prompts, this obviously represents a formidable task over a longer time period. Accordingly, the interview proceeds in terms of dates of occasions volunteered by respondents, with an associated increased probability of forgetting and telescoping, that is, misperceiving past behaviour as taking place more recently than it did.

5. A Time-Series Approach

Prior to the 1968/69 Finnish ALKO study other studies in Nordic countries had used an approach based on the last or the last two occasions of drinking. A similar method has recently been revived in the literature as a result of a large study (approximately 10 000 respondents) in New Zealand.

Gregson and Stacey (1981) write:

"the series of drinking events in a life history may be regarded, statistically as activity from a renewal process with partially censored information .... If the problem of estimating drinking levels is taken as being a problem of inference from the recent history of a time series, then three separate data are necessary and sufficient (there are strong mathematical assumptions about the stationarity and aperiodicity of the time series).

1. The quantity drunk at the last drinking occasion before being interviewed ..., call this q_1 mls ethanol.
2. When last drinking occurred; t_1
3. When next to last drinking occurred; t_2."

They then identify

t_1-t_2 as an estimate of the mean inter-drinking interval,

q_1 as an estimate of intake/occasion, and the rate of drinking is thus $q_1/(t_1$-$t_2)$ mls/day.

Employing this method Gregson and Stacey's estimate of total population consumption achieves 75% coverage of known consumption.

A similar method was criticized on the grounds of bias by Ekholm (1969), and Skog (1981) makes the same criticism of Gregson and Stacey's approach to measurement. Essentially the ground of criticism is that even if q_1 and t_1-t_2 are unbiased estimates of quantity and inter-occasion time, $1/(t_1$-$t_2)$ is not an unbiased estimate of frequency.

A further point not treated explicitly by Skog is that bias is introduced by any association between q_1 and $1/(t_1$-$t_2)$. The degree of bias in this measurement will therefore vary from individual to individual, depending on the distribution and type of periodicity in an individual's pattern of drinking occasions, and the association between quantity and time period distributions.

6. Discussion

Q-F type measures and their derivatives are intended to provide a categorization of drinkers, in order to explore associations between the categories and other attributes and problems of the individual respondents. More direct attempts at eliciting individual consumption form part of research projects aiming to model the relationships between such things as experience of health or other problems and alcohol consumption at the individual level. In many cases the questions relating to problems cover a much longer period of past time than the seven days preceding the interview, and this makes clear that it is not the last week's consumption of the individual that is of interest. Rather, last week's consumption is used as indicating some sort of parameter of the individual's pattern of consumption over a relatively long period of time. That this is not without pitfalls may be instanced by taking the Alcohol Dependence Syndrome (Edwards et al. (1977)) as a case in point. If one finds that last week's consumption is positively associated with the presence of symptoms of the syndrome is one to conclude that such consumption is related as a cause of the syndrome or simply as an effect?

Recent attempts to provide a framework for modelling individual consumption over time indicate the complexity of the problem. It is possible to build on mathematical modelling of other types of consumer behaviour as Alanko (1982) has done. Such models as have been proposed relate to estimation of the distributional form of consumption between individuals (Gregson and Stacey (1981)) and have no immediate practical application to the measurement problem. The problem of obtaining a tractable mathematical model of the natural history of individual consumption remains for the future.

Further developments in this field should include the comparison of methods of measurement both on the basis of existing data sets and also by experiment. Such comparisons could also take into account the relative costs of different measurement strategies. Studies of the suitability of different data acquisition strategies for the various methods of measurement would be of great practical value.

7. References

Alanko, T. (1982): A Statistical Model for Describing Drinking Behaviour and Distribution of Alcohol Consumption in Sample Surveys. Licentiate Thesis, University of Helsinki, Finland. (Unpublished).

Bradburn, N.M. and Sudman, S. (1982): Asking Questions. San Francisco: Jossey-Bass.

Cahalan, D. and Cisin, I.H. (1968): American Drinking Practices: Summary of Findings from a National Probability Sample. I. Extent of Drinking by Population Subgroups.

Quarterly Journal of Studies on Alcohol, 29, pp. 130–151.

Cartwright, A.K.J., Shaw, S.J., and Spratley, T.A. (1978): The Relationship Between Per Capita Consumption, Drinking Patterns and Alcohol-Related Problems in a Population Sample, 1964–1974, Part 1, Increased Consumption Changes and Drinking Patterns. British Journal of Addiction, 73, pp. 237–246.

Dight, S. (1976): Scottish Drinking Habits. London: HMSO.

Edwards, G., Gross, M.M., Keller, J., Moser, J., and Room, R. (1977): Alcohol Related Disabilities. Geneva: WHO.

Ekholm, A. (1969): Bias in a Certain Estimate of the Number of Drinking Occasions Per Year. In: Jonsson, E. and Nilsson, T.: A Joint Nordic Study of the Drinking Habits of Adult Males. Mimeographed.

Gregson, R.A.M. and Stacey, B.G. (1981): Distribution of Self-Reported Alcohol Consumption in New Zealand, 1978–1979. Psychological Reports, 47, pp. 159–170.

Kendell, R.E., de Roumanie, M., and Ritson, E.B. (1983): Effect of Economic Changes on Scottish Drinking Habits 1978–82. British Journal of Addiction, 78, pp. 365–379.

Mäkelä, K. (1971): Measuring the Consumption of Alcohol in the 1968–1969 Alcohol Consumption Study. Social Research Institute of Alcohol Studies, Helsinki.

Pernanen, K. (1974): Validity of Survey Data on Alcohol Use. In: Gibbins, R. et al., ed.: Alcohol. and Drug Problems, Vol. I. New York: John Wiley.

Plant, M.A. and Pirie, R. (1979): Self-Reported Alcohol Consumption and Alcohol-Related Problems: A Study in Four Scottish Towns. Social Psychiatry, 14, pp. 65–73.

Skog, O-J., (1981): Distribution of Self-Reported Alcohol Consumption: Comments on Gregson and Stacey. Psychological Reports, 49, pp. 771–777.

Straus, R. and Bacon, S.D. (1953): Drinking in College. New Haven: Yale University Press.

Waterton, J.J. and Duffy, J.C. (1984): A Comparison of Computer Interview Techniques and Traditional Methods in the Collection of Self-Reported Alcohol Consumption Data in a Field Survey. International Statistical Review, 52, pp. 173–182.

Wilson, P. (1980): Drinking in England and Wales. London: HMSO.

Wilson, P. (1981): Improving the Methodology of Drinking Surveys. The Statistician, 30, pp. 159–167.

Telescoping: The Skeleton in the Recent Reading Closet

Valentine Appel[1]

1. Research Methods

In the United States, magazine audiences are generally measured using one of two different measurement techniques. The first, which is known as the *through-the-book* method, requires that respondents express certainty that they had read or looked into a suitably aged (five weeks for weeklies and 11 weeks for monthlies) test issue after having been shown the cover and taken through the editorial content.

The second technique, called the Recent Reading method, requires that respondents answer with certainty that they had read or looked into any issue of the magazine in the previous publication interval – past month for monthlies, past week for weeklies, etc.

2. Results of These Methods

Until recently, largely because of the historic similarity of the American Target Group Index (TGI) recent reading audience estimates to the Simmons through-the-book estimates, most American researchers were of the belief that the two methods produced roughly equivalent results (see Simmons Market Research Bureau (1978)). So much so, that in 1978 the Simmons company announced that,

starting with the 1979 Study, Simmons would use both methods in order to expand the number of titles that were being measured: the through-the-book method would continue to be used to measure all magazines with other than a monthly publishing frequency as well as all monthlies with a rating of 3 % or greater, and the recent reading method would be used for the smaller monthlies.

The historic similarity of the audience levels produced by the two methods notwithstanding, there was sufficient concern among magazines, agencies, and advertisers about what came to be known as the *mixed method* that the Advertising Research Foundation (ARF) was successful in raising nearly $500,000 to conduct a methodological study to assess the comparability of the two methods. The reason for the enormous price tag was the then generally held belief that whatever differences would be produced by the two methods would be small and the sponsors wanted assurance that an average difference of as little as 10 % would be statistically significant.

Five months before the release of the ARF Top-Line Findings, when the results of the 1979 Simmons Study were first announced, the industry was stunned to discover that the Simmons recent reading estimates were nearly twice as large as what they had been accustomed to seeing.

That conclusion was subsequently confirmed by the ARF which found that for monthly magazines the recent reading method genera-

ted audience estimates which were 86 % higher than those produced using the through-the-book method, and that for weekly magazines the overage was 27 % (see Advertising Research Foundation (1980)).

This one finding has generated a storm of controversy in the U.S. such as has not been seen in the advertising research community for some time. The controversy concerns two central issues. The first issue has to do with the procedure and propriety of adjusting recent reading levels to conform to those achieved using the more traditional through-the-book procedure. The second issue has to do with the question of which of the two methods is closer to providing the correct audience estimates.

This paper will confine itself only to the second issue: the validity of the two methods.

3.   Validity of These Methods

When the Simmons company first announced that the recent reading estimates they were producing were nearly double those which either Simmons or TGI previously had reported, the recent reading audience estimates were immediately labelled as implausible first by Simmons, which offered an adjustment procedure to bring the estimates in line with through-the-book levels, and then by the industry. Later, the newly-founded Mediamark Research Inc. (MRI) produced their own recent reading magazine estimates which were to compete with the Simmons estimates. The Simmons questionnaire had asked simply whether or not the publication had been read in the last month, while MRI had developed what they described as a "perfected system" which went on to specify the length of the publishing interval in great detail, even to informing the respondent of the specific date when it began (see Joyce (1979)).

When the MRI data became available, it was clear to all that their technique had pro-

duced results which were virtually identical to the Simmons recent reading estimates.

The results implied that about as many people read two magazines a day as read the daily newspaper (see Mediamark Research Inc. (1980)). And despite the fact that the U.S. Department of Health, Education and Welfare found 21.7% of the U.S. adult population to be functionally illiterate (see U.S. Department of Health, Education and Welfare (1975)), the results indicated that 94% of all adults read an average of 11.6 magazine issues in the average month (see Magazine Publishers Association (1980)).

The root cause of this controversy is the fact that no one has ever been able to establish an objective standard of truth – a criterion if you will – against which the several magazine audience measurement techniques can be evaluated. The Advertising Research Foundation has been busily studying this problem for a number of years with little success, even having gone so far as to have conducted an unsuccessful study in which a former New York City Police Department fingerprint expert was engaged as a consultant to try to identify particular readers of particular magazine copies (see Greene and Maloney (1976)).

4.   Telescoping Phenomenon

All of the ARF's efforts in this regard have been directed to attempting to validate the through-the-book procedure, believing that it is not possible to validate the non-issue specific recent reading method. Actually, however, it is a simple matter to demonstrate that respondents are incapable of judging accurately the recency with which past events have occurred. Psychologists have been studying this phenomenon for some time (see Hinricks and Buschke (1968)) and have generally concluded that:

(a) the longer the time interval between the event and the judgement of the recency of that event the less likely is the judgement to be accurate. From this we would expect that the judgement of whether a magazine had been read in the past month would be less accurate than the judgement of whether it had been read in the past week.

(b) the longer the time interval between the event and the judgement of the recency of that event, the more likely is it to be perceived to have occurred more recently than it actually did. From this principle one would expect that the recent reading method would spuriously favour magazines with longer publishing intervals where the method requires that recency judgements be made over longer periods of time.

The phenomenon has come popularly to be known as *telescoping*, and in our view completely explains the fact that the recent reading method produces inflated estimates in general and disproportionately higher estimates for monthlies than for weeklies.

5. The Television Test

However except for a few proprietary studies conducted by broadcasters, most of the research on the subject of telescoping has been conducted in the psychological laboratory using simple words or pictures as stimuli and judgements over very short time intervals. In preparation for this paper, therefore, we decided to perform a real life demonstration, using weekly television programmes, to show the inability of respondents to recall accurately whether or not an event had occurred even within as short a time period as seven days. We chose to perform the demonstration using weekly television programme viewing because, unlike magazine reading, the time of the viewing occasion is precisely known and there is no possibility of complications caused by replicated and parallel viewing.

The study was conducted by telephone using the Bergen County, New Jersey telephone directory as a sampling frame. The sample was limited to female household heads, and all interviewing was conducted after 18.00 in order to ensure a proper representation of working women. A total of 700 interviews were completed, 100 on each of seven consecutive days divided equally between two field periods; December 9–15, 1980 and January 11–17, 1981.

The interview proceeded as follows: respondents were read a list of 20 weekly television shows and asked whether each one had been watched in the past 30 days. Then for each programme watched, the interviewer asked whether the respondent happened to have watched that show in the past week, that is in the seven days since last (day of week) not including today. Those answering "yes" were classified as "recent viewers."

We reasoned that if the respondents' judgements of the recency of the telecast were accurate, we should observe the same ratings for these shows regardless of the day on which the recent viewing question was asked. However, to the extent that the recency judgements were distorted by the telescoping phenomenon, one would expect to find the ratings to be different depending upon the time interval between the telecast and the interview.

Fig. 1 shows the mean recent viewing rating of these shows aggregated according to the time interval between the day of the telecast and the day of the interview. The mean ratings are plotted on the vertical axis and on the horizontal axis are plotted the number of days between the day of the telecast and the date of the interview.

The mean rating observed as a function of the time interval from the day of the telecast to the day of the interview is represented by the seven dots, and the diagonal line represents the least squares best fit.
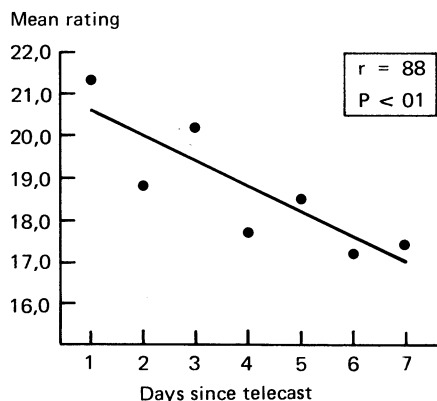
Mean rating

Fig. 1. Recent Viewing Ratings by Time Lapse of Interview

As can be seen, the longer the time interval between the day of the telecast and the day of the interview, the lower is the mean rating. The Pearson product moment correlation coefficient is .88 and is significant at the .01 level using five degrees of freedom. This relationship proves conclusively that respondents are incapable of judging accurately whether an event such as their most recent viewing of a weekly television programme occurred within the past seven days or not.

6. Forgetting or Telescoping

If memories were perfect, one would expect to find the same percentage claiming to view on each of the interviewing days regardless of whether the interview was conducted the day after the telecast or six days later. Obviously, memories are not perfect, or we could not observe the relationship seen here.

What causes it? If these were the only data available one could convincingly offer either of two equally plausible explanations. The first explanation would be that people simply forget with the passage of time, and that the true audience levels are actually higher than the recent viewing estimates would indicate.

The second explanation would be that we are looking at the result of telescoping caused by some people imagining that an event which actually occurred eight or more days ago happened within the past seven days.

Perhaps the telescoping concept requires more elaboration, and a concrete example will help. Were I to survey a sample of people on the day following the telecast of a weekly show, and were I to ask whether they had watched that show in the past seven days, virtually all of those who had watched the day before would answer that they had, as would some proportion of nonviewers who had actually watched eight days ago, but imagined it to be seven.

The next day, a smaller proportion of such nonviewers who had actually watched nine days ago would falsely answer "yes," and one would expect this proportion to drop with each successive day until the day of the next telecast.

Recognizing that the recent viewing estimates are necessarily in error, and wishing to resolve the question as to whether the declining audience levels were the result of telescoping or simple forgetting, we designed a questioning procedure to provide what we believe to be a more accurate estimate of viewing levels – more accurate because it shortened the recall period from seven days to one in order to minimize problems of memory distortion, and more accurate because it followed the ARF recommendation for obtaining measures of yesterday reading. Basically, it was the same method which is used both by Simmons and by MRI for measuring yesterday readership of daily newspapers.

Accordingly, everyone claiming to have viewed the show in the past seven days was asked for the last time she happened to watch it, not including today. Those answering "yesterday" on the day following the telecast were then classified as 'yesterday viewers.'

We reasoned that if the recent viewing rating was lower than the yesterday viewing estimate it would argue in favour of simple forgetting. If, on the other hand, the recent viewing rating was higher, it would argue in favour of telescoping caused by confusion of the recency of the last viewing occasion.

The mean yesterday rating was 14.4 %, the mean recent viewing rating, 18.7 %. Thus, the recent viewing rating was 30 % higher, and statistically significant at the .001 level using the method of sample replicates with nine degrees of freedom (see Frankel and Frankel (1977)). This highly significant difference supports the validity of the telescoping hypothesis.

7. The Magazine Test

Mindful, however, of the fact that our basic interest is with magazine audience measurement rather than television viewing, we performed the same exercise using nine weekly publications.[2]

National Enquirer	People
Newsweek	The Star
New York	Time
New Yorker	TV Guide
US News & World Report	

Then again we produced two estimates of weekly reading: a recent reading estimate based on the past seven day claim and the other based on the number of yesterday readers.

However, since we know that some consumers read some magazine issues on more than one day, it was necessary to take that fact into account in order to generate an average issue audience estimate. Accordingly two separate attempts were made to estimate the incidence of first time yesterday reading of the issue via direct questioning for each magazine. Without going into detail, suffice it to say that

[2] Actually ten publications were included. *Midnight Globe* was subsequently deleted after learning of a name change.

both attempts produced first time reading estimates which when converted to weekly ratings were less than half the recent reading estimates.

Being reluctant to conclude that the recent reading estimates for weekly magazines are more than twice as large as they should be, we also explored the possibility of correcting the yesterday reading level for each magazine by dividing the yesterday reading incidence by the mean number of reading days as published in the 1980 Simmons report (Table 1).

When we compared the mean rating thus obtained with the 1980 Simmons through-the-book ratings drawn from a roughly comparable sample (female homemakers with listed telephones, living in the New York ADI), we found that the means were very close and that the ratings on a magazine by magazine basis correlated +.96 with each other (SMRB TTB 11.4 %; yesterday recall 12.0 %). In other words, the adjusted yesterday recall estimates were quite comparable with those obtained through-the-book.

The next step was to compare these estimates with those obtained using the recent reading method. The recent reading estimates at 15.1 % were higher than the yesterday recall at 12.0 % by 26 % ($p<.001$), which is virtually identical to the +27 % difference that was reported for weekly magazines in the ARF Comparability Study, and very close to the +30 % difference which emerged for television viewing when no correction for multiple day viewing was required.

Table 1. Yesterday recall adjustment factors

Title	Reading days
National Enquirer	2.2
Newsweek	1.8
New York	1.8
New Yorker	2.4
People	1.8
The Star	2.3
Time	2.1
TV Guide	5.5
US News & World Report	2.1

8. Conclusion

(a) The recent reading method in theory is perfectly reasonable if one can accept the assumption that respondents can not only accurately remember that they have been exposed to a particular media vehicle, but also that they can accurately judge the recency of the last such occurrence.

However, we have proved conclusively that consumers are incapable of making accurate judgements of whether a media exposure – in this case the viewing of a weekly television show – occurred within the past seven days or not. If consumers were capable of making such judgements, there is no way that reported audience levels could show the pattern of decline we have seen as the time interval increases between the day of the telecast and the day of the interview.

(b) The recency method produces weekly television audience estimates which are about 30 % higher than the estimates which are produced on the basis of yesterday recall using the procedure recommended by the ARF to measure yesterday reading of newspapers.

(c) In the case of magazines, the same yesterday recall method, modified to accommodate the fact that magazines are frequently read on more than one day, produces audience estimates for weekly magazines which closely approximate those obtained using through-the-book procedures.

(d) When the recent reading magazine estimates were compared with those obtained on the basis of yesterday recall the recent reading estimates produced a 26 % overage relative to the yesterday recall estimates.

(e) The 26 % overage is roughly comparable to the 30 % overage which was reported for television viewing where no adjustment for multiple day exposure was necessary. Moreover, it is virtually identical to the 27 % overage which the ARF reported for the recent reading method relative to through-the-book for weekly magazines.

We interpret these facts to mean that the recent reading method significantly overstates magazine audiences and does so by a process called telescoping caused by the inability of the respondent to judge whether or not a particular event has occurred within the publication interval.

Although we did not directly address this issue for monthly publications as we did for weeklies, all of the information available both in the psychological literature and in the ARF Study suggests that as the publishing interval increases so does the severity of the telescoping problem.

As a result, not only does the recent reading method produce spurious audience estimates, it does so in such a way as to seriously disadvantage weekly publications relative to monthlies.

9. References

Advertising Research Foundation (1980): A Controlled Field Experiment Comparing Three Methods of Estimating Magazine Audiences. Comparability subcommittee of the Magazine Research Development Council, January.

Frankel, M.R. and Frankel L.R. (1977): Some Recent Developments in Sample Survey Design. Journal of Marketing Research, 14, pp. 280–293.

Greene, J.D. and Maloney, J.F. (1976): Fingerprints Can't Test for Validity. Journal of Advertising Research, 16, 3, pp. 49–50.

Hinricks, J.V. and Buschke, H. (1968): Judgement of Recency Under Steady State Conditions. Journal of Experimental Psychology, 78, 4, pp. 574–579.

Joyce, T. (1979): SMRB's Adjustment Plan: Methodological Issues. Mediamark Research Inc., September 10.

Magazine Publishers Association (1980): Taking the Measure of Magazines ... Facts, Trends & Perspectives. Magazine Newsletter of Research No. 33, August.

Mediamark Research Inc. (1980): The Level of Magazine Reading.

Simmons Market Research Bureau (1978): An Historical Analysis, Simmons vs TGI. September.

US Department of Health, Education and Welfare (1975): HEW News, October 29.