Marion Appeal
Burke-Inter/View, The Netherlands
Bouke Walstra
De Weekbladpers BV, The Netherlands

4.1 ON VALIDATING THE FIRST TIME READ YESTERDAY METHOD – A CONTINUING STORY

This paper was presented by Bouke Walstra, who prefaced it with the following statements.

I am sorry to inform you that my co-speaker, Marion Appel, will not be able to present her part of our paper. The SUMMO board decided a few days ago that the average issue readership results that go with the new Target Group research due to be published in a couple of weeks should be based on the results of the improved questionnaire resulting from the experiments described in this paper. Since Marion Appel is responsible for the production of the results, and delays in publishing them would not be acceptable, she cannot be here.

I have a second opening statement to make. The results of our experiments are currently being studied by an Advisory Board consisting of university professors. Since they have not yet given their verdict, our conclusions have not yet become law: they are at the moment our own conclusions.

INTRODUCTION

The SummoScanner is currently the leading Dutch national media survey which establishes the average and total readership of over 120 print media and several other media types. In this introduction we will summarise the SummoScanner method, as has been done previously by de Hond & Huzen (1983) and by Bouke Walstra (1985). The SummoScanner began in 1985 as the MediaScanner, a multi-client survey developed by Inter/View, and was adopted by SUMMO in 1983. SUMMO is a Dutch organisation whose members – publishing companies, advertising agencies and manufacturers – participate in the national media survey. Prior to the SUMMO decision to adopt the MediaScanner, Maurice de Hond and later Bouke Walstra conducted several experiments in order to improve the methodology. In this paper we will resume the continuing story of the SummoScanner, and bring you up to date on the lastest findings.

The SummoScanner has the following characteristics:

– a total of 125 telephone interviews are conducted daily from one central CATI Unit in Amsterdam, six days a week, which add up to 39,000 interviews per annum.

– the sample is automatically selected by random digit dialling.

– automatic recall of non-answering numbers on different days and times, during a maximum of four weeks.

– random selection of the person to be interviewed, chosen from the household members reported.

– random presentation of titles, within groups.

– answers entered immediately, the titles of the journals mentioned being inserted in the following questions.

– the questionnaire routing is automatically presented to the interviewer by the computer.

In comparison with face-to-face-research, this method has several advantages, an important one being that the whole population of 13 year and over has an equal chance of being selected. This advantage is slightly tempered by the fact that 6% of Dutch households have no telephone.

Another major advantage is the opportunity of checking on the interviewers during their work and improving on the quality of the interview by continuous training and correction. Furthermore, interviewer mistakes are reduced to a minimum by the automatic selection of the next question. As a result the interview is highly personalised in relation to the reading behaviour of the respondent.

Disadvantages of course occur in any method. The SummoScanner has some: the length of time an interview can take is limited, no use can be made of masthead cards in order to avoid misinterpretation of a title; also a larger percentage of refusals seems to occur in telephone interviews than in face-to-face.

Because of the chosen method of establishing the readership figures, via the first-time-read yesterday method, a large sample is needed in order to obtain an adequate statistical basis for making probability calculations. This point will be touched upon by Paul van Niekerk later in this Symposium.

The data that are gathered in the SummoScanner

(1) Introduction and inventory of household situation.

(2) Readership questions concerning dailies, weeklies, monthlies, other journals (see next paragraph).

(3) Additional information on media: copy source, passing on to neighbours, amount of time spent on reading yesterday, reading of the latest issue (and/or older ones) (yesterday).

(4) Viewing, listening etc, of television, radio, cinema and yellow pages (yesterday).

(5) Possession of motor cars in the household.

(6) Visiting behaviour at grocery stores.

(7) Demographics and yesterday's 'outdoor' behaviour.

The total interview time is now about 25 minutes; 20% of all interviews take more than 30 minutes. Expansion of the number of titles in the SummoScanner seems inadvisable, as the percentage of refusals appear to increase if the average time needed exceeds 25 minutes.

On a yearly basis a mail survey is held among 7,200 respondents who agree to participate in this follow up-survey. In this way a great deal of information on product usage, interests and activities of those people is gathered and linked to the already known facts about their media behaviour. This target group survey is now the most intensively used for media planning in the Netherlands. VNU has carried out a data fusion of their psychographic research PSYCHE with the SUMMO target group survey.

READERSHIP QUESTIONS IN THE SUMMOSCANNER

The choice in favour of yesterday's reading was influenced by the fact that reading in the past publication interval (the standard in the National Readership Survey until 1982) could not be validly assessed for weeklies and monthlies because of too great a demand on respondents' memories.

The FRY-based calculation model for reading probabilities offered a method in which, for the first time ever, the memory gap was equal for all media types, be it monthlies, weeklies or

dailies. (The calculation procedure is explained by Paul van Niekerk in his Symposium paper).

Also the problems of parallel reading and replicated reading could be solved by whether the respondent read this particular issue, yesterday for the first time.

An outline of the questions to establish average issue readership of print media

Four 'questions' are needed to establish readership:

(1) To begin with the respondent is confronted with every title in the survey, and asked whether a journal is *ever read or looked into*. This question determines the total readership of the media. The definition of total readership is a very broad one in order not to miss any reading event, whether the respondent is a chance reader or a regular reader. It is imperative that all fish are caught in this net, since the effect of the addition of chance readers to the first-time-read-yesterday question can cause great distortions in the probabilities calculation, when they are not all included in the total readership figure.

To maximise the chance of discovery of every reading event, all titles are enumerated to the respondent. The titles are grouped together according to their subject or target group. Within each group the titles are rotated.

This first question, the 'ever read' question, acts as a filter for the next question.

(2) What titles are *read or looked into yesterday*.

(3) The set of questions to establish whether an issue was *read yesterday for the first time* is subject to experiments we will report in more detail

later in the paper. No FRY questions are asked concerning the dailies, acting on the hypothesis that the number of replicated readers of dailies is negligible.

(4) Finally, after once more recalling all titles the respondent ever reads, the reading *frequency question* is posed. The frequency question is worded as follows "How many of the last six issues of ..... did you read"? Consequently some seasonal effect is to be expected, especially for dailies and weeklies. On the other hand many respondents answer this question as if you asked for their normal behaviour. As is generally known, the answers to the frequency question are subject to memory gaps that have different effects on different media. That is why this criterion must be used in combination with FRY, as was also suggested by Douglas and Lysaker (1983).

Combination of frequency and FRY should solve several problems:

(a) The same proportion of over and underclaims do not occur in the different frequency-classes. Occasional readers may underclaim their reading frequency, but regular readers may be more inclined to overclaim.

(b) Over and underclaims may be different for different media, as it is much harder to remember the reading of six monthlies, than of six dailies. Correction of the frequency answer by FRY eliminates this unjustice.

(c) Subscribers to a reading circle portfolio particularly, underestimate their reading frequency, with the result that the actual reading probabilities are much higher than the theoretical value.

CALCULATING READING PROBABILITIES: THE RELATION BETWEEN THE RECENCY QUESTION AND THE FREQUENCY QUESTION

Theoretical considerations

It has always been hypothesised that the reading frequency question itself cannot be used for calculating reading probabilities. Of all the newspapers and the magazines they ever look into, respondents cannot reproduce exactly how many of the last six (or 12) issues they have read or looked into.

It is assumed that the frequent readers overestimate their reading frequency and infrequent readers underestimate the frequency. A frequent reader tends to forget that he sometimes (because of holidays or other reasons) has not the time nor the opportunity to read an issue of a title he otherwise always reads. In that case the reported out of the last six issues should, for instance, be five out of six issues.

With journals that are read out of habit (especially daily newspapers, radio/TV guides) overestimation will be more prominent than with other journals. Another factor is that overestimation through just forgetting will be related to the publication interval of a journal (the timespan for six issues of a monthly is much longer than for a daily).

Generally we would say that there is a relationship between the publication interval and the accuracy with which the reading frequency is reported. Underestimating by infrequent readers is not necessarily underestimation when it concerns the nought out of six frequency category. The theoretical probability should be a little above zero, assuming that nought of six really means less than one of six.

Another factor that influences so-called underreporting of the reading frequency has to do with the phenomenon of the reading circle portfolio, a portfolio with ten to fifteen magazines in it, most of them weeklies. The issues in a portfolio can be up to ten weeks old, depending on the subscription price one is prepared to pay.

Somebody who subscribes to a portfolio that is, say, four weeks old and reads *all* issues of magazines in the portfolio, correctly answers the reading frequency question by reporting that he has read two of the six last issues of the magazines.

Most of the portfolio readers will grasp the intention of the frequency question. They will then report how many they read of the last six issues of the magazine contained in the portfolio. Yet some will be as strict as in the example. The conclusion is that the frequency figures certainly contain estimation errors and distortions of different kinds, due to forgetting as well as distribution factors.

The assumption of recency methods (including the method based on yesterday's reading) is that the reading probabilities based on the recency question are nearer the truth than the theoretical reading probabilities. In other words, the recency question does not contain systematic errors as is the case with the frequency question. In several Symposia papers it was shown that this assumption is certainly not valid for the question that tries to establish reading in the last publication interval. Phenomena such as telescoping, replicated reading and parallel reading distort the results.

Asking for yesterday's reading behaviour should solve the problem.

– by only counting the first contact with a magazine, replicated reading is removed.

– parallel reading is easily established and accounted for.

– yesterday is so close that telescoping should not occur.

From experiments whose results were shown in Montreal, the assumption of validity of answers to the yesterday question was supported.

With that telescoping was also removed as an unacceptable source of errors. So far, so good; the big problem however appeared to be the FRY question. Only when the first contact with an issue can be validly established, can these figures be used for calculation of the reading probabilities and subsequently for the establishment of the AIR.

In the next chapter our efforts to improve the validity of the FRY concept will be summarised. For the moment let us assume that our FRY results are not biased. The question remains as to how we arrive at acceptable and stable AIR-figures. In principle, on an individual title basis, we should calculate reading probabilities for each publication period of the research (for the SummoScanner that is every quarter).

Using FRY as the basis for the calculation this is not possible. Even for the dailies the number of yesterday readers per frequency category is not enough to ensure stable results on a quarterly basis. Therefore, reading probabilities are calculated every quarter on the basis of the last four quarters (ie a sample size of 39,000). In this way possible seasonality is flattened in the published AIR results.

For the majority of the titles in the Scanner, the number of FRY cases is not enough to establish the reading probabilities on the title level, even when calculated on a yearly basis. Similar titles (similarity of distribution patterns and/or reading patterns) are grouped together to provide an adequate statistical basis.

The logic of this grouping together of different magazines is the assumption that the estimation errors of respondents on the frequency question are not title-specific. It is, however, to be expected, that the difference in timespan of the frequency question for monthlies, weeklies and dailies influences are estimation errors. It is therefore necessary to do the grouping with titles that have the same publication interval.

The results of this procedure and the statistical stability of the results will be shown by Paul van Niekerk in his paper.

We think we have now given the necessary background for understanding the relevance of the efforts to validate the FRY question, for it is the very backbone of the method.

EXPERIMENTS IN 1984/1985

Summary of the Salzburg paper

The earlier experiments on the MediaScanner (now SummoScanner) had a wider scope than those we are dealing with in this paper.

Due to experiments both sample and weighting procedures were completely overhauled in 1985.

Besides that, intensive interviews were held to check all the major reading questions, focusing in particular on the yesterday and the FRY questions. It appeared that it was the FRY question that really created a problem in terms of overreporting as well as underreporting.

Experiments on a larger scale were carried out using a simple type of intensive interviewing, designed to test alternative question wordings for the FRY questions. As a result of these experiments two question wordings were recommended for use in the scanner interview on a split-run basis.

However, the overall conclusion was that these alternatives did not solve the problem. Both over and underestimation remained, although underreporting was less prominent than over-reporting. It was suggested that a person who has read a magazine yesterday will answer the FRY question with 'yes' if yesterday was his first serious contact with the magazine, forgetting that days before he had already looked into the same issue superficially. Were he questioned the day after his first contact with the issue he also would have said 'yes' to the FRY question.

This being the hypothesis for the overreporting phenomenon, we suggested following a different route.

HOW TO CONTINUE ON THE BASIS OF EARLIER EXPERIMENTS

Up to 1986 much time and money was spent on finding alternative question wordings for the FRY questions in order to eliminate over-reporting.

In Salzburg we suggested that it probably would be necessary to leave this route and concentrate on unravelling the FRY question in a step-by-step questioning procedure. The reason for probing further into simple alternative question wording, even after Salzburg, had to do with questionnaire length and, consequently, money.

It was to expected that new media would be intergrated into the scanner. This happened after the deal with SUMMO. The result was a lengthening of the average interview time by up to 25 minutes.

Because experiments with the FRY question have to be on a grand scale (first time yesterday readers of monthlies especially are seldom found, so big samples are needed) and are therefore expensive, we could not just go on trying to validate alternative question wording in the hope of succeeding some time.

One thing seemed clear. Since we had to try getting more attention from the respondent when asking the FRY question, we suggested using an introduction to the question in order to clarify the FRY concept to the respondent and give him time to consider.

In order to do so we had to rebuild the questionnaire and make it 'concept centred' instead of 'title centred'.

Until then in the questionnaire per (yesterday read) title, questions were subsequently asked about the number of minutes read yesterday, which issue was read yesterday: the latest or an older number (or a combination of the latest issue and older ones) and subsequently the FRY question.

We now decided to leave the experiments on question wording and follow a more fundamental route. The first step was to suggest using an introduction to the FRY question and to put the question for all the titles one after another, instead of posing other questions in between.

RESULTS OF SUMMOSCANNER EXPERIMENTS IN 1987 AND 1988

Concept centred interviewing

In order to facilitate the understanding by the respondents of the concept of first-time-read

ON VALIDATING THE FIRST TIME READ
YESTERDAY METHOD – A CONTINUING STORY

yesterday, it was thought necessary to use an introduction to the FRY question. In 1987 an experiment was set up in a shadow media-scanner, lasting four weeks. In this experiment the structure of the interview was changed so that the questions concerning one concept (eg. first time read yesterday) were posed for all titles one after another, before continuing with the next concept (reading frequency).

In the experiment four different combinations of introductory (two alternatives) and question texts (two alternatives) were used:

(A) A magazine can be looked into for the very first time in different ways, immediately when one comes across the issue, or a day later.

(B) If you have read an issue of a magazine, it is possible to have looked in it before, for example the day before yesterday or even before that.

(I) The issue of ..... that you read yesterday, when did you look into that issue of ..... for the very first time: was that yesterday or before?

(II) When yesterday you read ..... was it the very first time you read that issue of ..... or did you look into that issue of ..... before?

By means of these different wordings we tried to find out which element would appeal more directly to the respondent:

"Think of the moment when you got the journal, did you look into it?" (A II); "Do you usually read an issue on different days, and was this the case with this particular issue?" (B I) (A I) and (B II) being cross-overs).

To test the improved understanding of the FRY concept several check questions were added, immediately after the normal questionnaire was ended. Overclaims and underclaims were

checked by using these questions as a simple form of intensive interviewing: a series of questions was once again asked concerning all titles that were read yesterday.

At first it was established if the respondent had been reading at home or elsewhere. If it was read at home, the respondent was asked to remember when the issue arrived in the home. After that the FRY question was asked once again.

If it was read somewhere else, it was established whether this was chance reading, or often-recurring behaviour. Depending on the answer the possibility that the issue had been read before was suggested, and the FRY question was asked once again.

Results

Table 1 shows the results of the check questions and the percentage of the initial responses that proved to be correct/incorrect.

The validation of the FRY question is not only carried out by intensive interviewing but must also be carried out by means of the internal consistency of the FRY levels within the frequency categories.

In view of the problems of validating the frequency question it is almost impossible to find out the truth about reading probabilities within each frequency segment. However, we can be sure that the probability score within six of six readers cannot be higher than 1.00 In fact it should be somewhat lower. As a result of the check questions, the average reading probabilities were reduced (Table 2).

Conclusions

We were very disappointed by the results of this experiment. Neither of the four alternative

ON VALIDATING THE FIRST TIME READ
YESTERDAY METHOD – A CONTINUING STORY

Table 1

Results of the check questions

Weeklies: Claimed first time read yesterday

	A I	A II	B I	B II
Total (gross)	690	664	658	645
	%	%	%	%
Total	100	100	100	100
Correct	74	76	73	74
Incorrect	26	24	27	26

Weeklies: Claimed non-first time read yesterday

	A I	A II	B I	B II
Total (gross)	722	750	777	590
	%	%	%	%
Total	100	100	100	100
Correct	94	93	95	86
Incorrect	6	7	5	14

Monthlies: Claimed first time read yesterday

	A I	A II	B I	B II
Total (gross)	147	157	187	141
	%	%	%	%
Total	100	100	100	100
Correct	65	76	64	81
Incorrect	35	24	36	19

Monthlies: Claimed non-first time read yesterday

	A I	A II	B I	B II
Total (gross)	199	148	221	183
	%	%	%	%
Total	100	100	100	100
Correct	89	90	89	90
Incorrect	11	10	11	10

Table 2

Average reading probability before and after correction

	Before	After
Weeklies	.64	.51
Monthlies	.66	.55

Reading probability in 6 of 6 class

	Before	After
Weeklies	1.22	1.02
Monthlies	1.72	1.36

wordings had been able to eliminate the mistakes that are made by respondents in trying to answer the FRY question: check questions gave about the same amount of corrections on FRY for each alternative. However, the possibility of achieving a reasonable reduction in the number of mistakes seemed to lie in the wording of the FRY question itself, analogous to the check questions used in the experiment.

Furthermore, it was decided that the concept-centred structure of the interview is preferable to the title-centred structure. This conclusion was reached on the basis of interviewer remarks that proved the interview ran more smoothly.

Step - by - step approach to yesterday's reading

Following the concept of the check questions that were devised earlier, a new experiment was set up in order to test these questions: now as an integrated part of the questionnaire and no longer recognisable as check questions. In this four-week experiment in a shadow mediascanner during February 1988 the following routing was tested:

ON VALIDATING THE FIRST TIME READ
YESTERDAY METHOD – A CONTINUING STORY

In this experiment we concentrated on the in-home reading, since this accounts for over 80% of all cases of yesterday reading.

We tried to make manifest to the respondent what he usually does with a specific journal: is it left unopened and read on a specific day (weekend), or is it glanced through immediately on receipt, and what happened with this specific issue? As a result, in most cases it would no longer be necessary to use the FRY question itself, because we would be able to establish reading before yesterday by means of the elimination of possibilities (Figure 1).

Figure 1

Routing of FRY questions
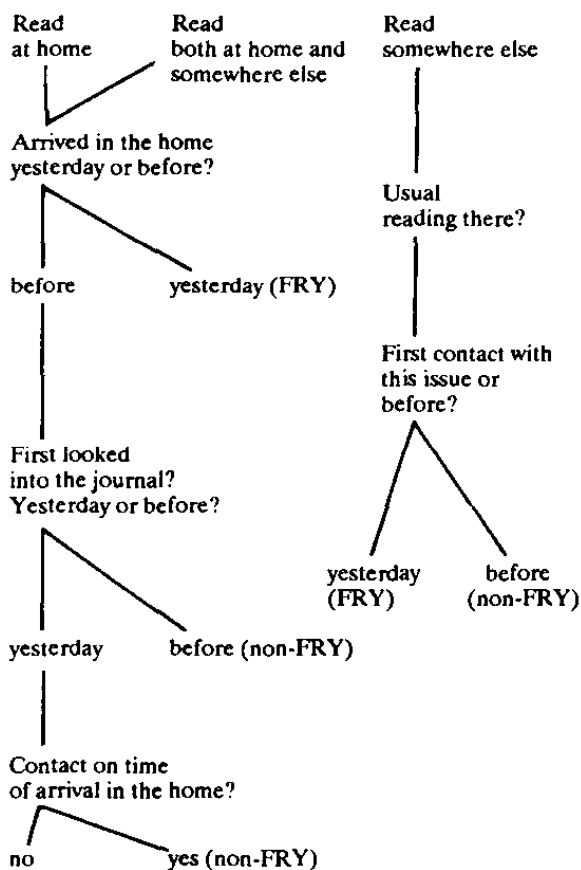
Results

86% of yesterday's reading events took place in the home.

By following the routing described above, in 16% of all cases of yesterday reading the journal was received in the home yesterday. Table 3 shows the FRY results in this four-week-experiment.

In total 43% of yesterday's reading events were claimed to have been the first contact with this specific issue of the journal in question.

76% of FRY was found with respondents who read their journal at home, so this group is by far the most important group for intensive search for mistakes.

The calculation of reading probabilities, which took place for all groups of titles, showed that indeed mistakes were still made. Table 4 shows

Table 3

FRY results in concept centred experiment

	%		%
Yesterday's reading events:	100	FRY:	43
Read at home	86		
Received yesterday (FRY)	16		16
Received before yesterday:			
Read before yesterday	52		
First time yesterday	17		17
Read elsewhere	14		
Usually there:			
Read before	4		
FRY	9		9
Not usually there:			
Read before	0		
FRY	1		1

ON VALIDATING THE FIRST TIME READ
YESTERDAY METHOD – A CONTINUING STORY

Table 4

Reading probabilities

	Frequency	
	6 of 6	Average
Women's weeklies	1.25	.56
Monthlies	1.84	.68

the reading probabilities of two groups of journals that still showed probabilities of over 1.00 in the six of six frequency class, the women's weeklies and the monthlies.

Possible remaining mistakes are:

(1) Yesterday's reading itself.

(2) Yesterday received in the home (especially when someone is not at home during the day).

(3) If received before yesterday, first contact yesterday.

(4) If incidentally read elsewhere, read before yesterday.

We were certain that these mistakes were not completely eliminated.

Conclusions

The February experiment shows that respondents are still strongly biased towards 'yesterday' because of the alternatives offered by the interviewer:

Did you receive the journal YESTERDAY or BEFORE?

(If received before:) When did you look into this issue of ..... ? Was it YESTERDAY or BEFORE YESTERDAY?

By trying to keep the interviewing time as short as possible we had chosen to minimise the number of alternatives for the respondent. The results were quite unsatisfactory. Respondents who are not immediately sure of their answer might be biased towards answering 'yesterday'. To avoid the positive yesterday effect, we had no alternative but to throw open the answers to these crucial questions, thereby stimulating the respondent to remember previous contacts with this issue.

Does one glance through or read the issue immediately upon receipt of the new issue? (This is also subject to more or less habitual behaviour. Of course, a relatively important group of occasional readers are indeed called upon to remember exactly when this first contact occurred.) We believed that – within the concept centred interview – the idea of first contact with an issue is made much clearer to the respondent.

Undirected step-by-step approach to yesterday's reading

The beginning of the second quarter of 1988 was designated as the moment when all improvements to the SummoScanner were to be implemented – ie sampling, weighting, expansion of the population, several improvements in the questionnaire, concept centred-interviewing, new grouping of titles for the calculation of reading probabilities, as well as the 'definitive solution' for the FRY problem. In a few weeks preceding the first of April we tested the new wording and indeed found that a smaller percentage of yesterday's reading events was claimed to be the first time: 28% of all cases in the home reading (was 33%).

These results led to the implementation of open questions, that avoid as much as possible biasing respondents towards yesterday (Figure 2).

As a result of the concept-centred interviewing and the undirected questions concerning FRY, we found correct levels of reading probabilities for most groups of titles, with the exception of monthlies and Radio and TV guides (see Table 5). For remaining over-reporting agreed editing rules are being used in the SummoScanner.

The level of overclaiming the first time read yesterday seems to have become much more acceptable than before. At the end of this year we will have more evidence; these are the first results based on nine weeks interviewing.

On the whole we are satisfied with the progress made in the field of recording FRY as a means of correcting the frequency question, even though some unsolved problems still remain.

EVALUATION

Since the start of readership measurement by Burke Inter/View with the FRY method on the basis of a CATI procedure, efforts have been made to improve the instrument. Results were shown at the Montreal and Salzburg Readership Symposia and now again in Barcelona.We think it is time to consider where we stand. Have we succeeded in developing in The Netherlands a media research approach that can withstand the judgment of time? To put it simply: do we have research now that gives fair and stable readership figures for monthlies and weeklies as well as for dailies? According to our experiments with the FRY question it seems we have succeeded in eliminating respondent bias to a considerable extent. In this respect the SummoScanner has reached a standard never before attained in Dutch readership research. But can we say that we now have the best possible method of measuring readership?

It is not easy to answer that question with a simple yes or no. Besides the advantages and
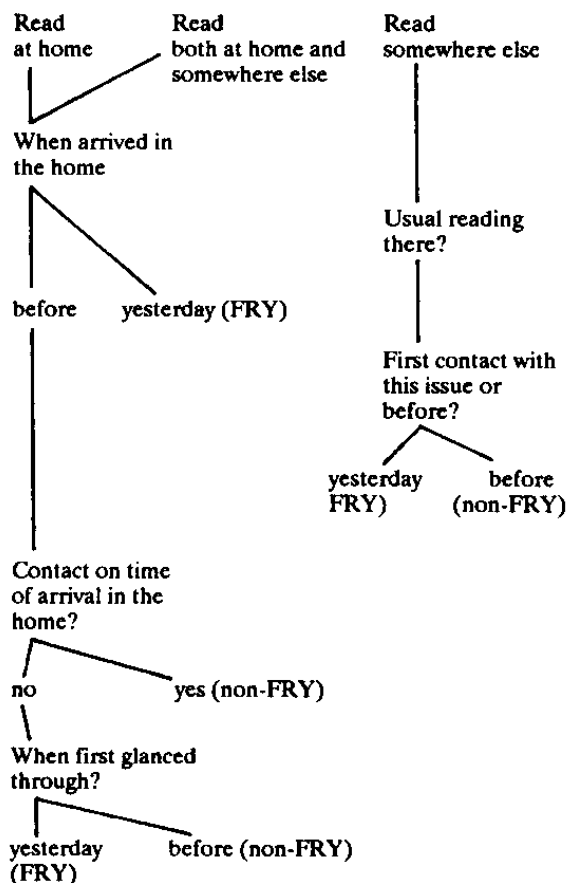
Figure 2

Routing of FRY questions (undirected approach)

Table 5

Reading probabilities

	Frequency	
	6 of 6	Average
Weeklies	.92	.42
Bi-monthlies	.83	.48
Monthlies	1.11	.44
R/TV guides	1.16	.80

disadvantages of the CATI system noted before, using the FRY method has its own pros and cons. Telescoping, replicated and parallel reading causing major distortions in recent reading methods in which reading in the latest publication interval is established, is under control within the FRY method.

The big disadvantage of the FRY method is the sample size that is needed. Even with the SummoScanner the yearly sample of nearly 40,000 is not enough to get stable reading probabilities on a title basis, especially for monthlies and smaller weeklies. Even with the *Kampioen* (the monthly magazine of the Dutch Tourist Organisation, with a circulation of more than 2 million copies), the number of the FRY cases is not enough to calculate reading probabilities on a title basis. As a result of the necessity of grouping titles together to obtain statistical robustness, title specificity in the AIR results is only effected by the results of the frequency question.

Only if the assumption is valid, that for similar titles the relation between reading frequency and FRY is the same (in other words, memory faults regarding the reading frequency have the same pattern for the titles that are grouped

together), the grouping is justified. The SummoScanner is founded on this plausible but yet unproven hypothesis.

With this conclusion we think we have made it clear that yet another Readership Symposium is needed (and much more than that). Who will take up the gauntlet?

References

De Hond, Maurice and Huzen, Walter (1983). New approach to readership surveys: the Mediascanner. Montreal Proceedings.

Douglas, Stephen A and Lysaker, Richard L (1983). The audience levels produced by the 'claimed first reading' method. Montreal Proceedings.

Van Niekerk, Paul (1988). Reliability of the first time read yesterday method. Barcelona Proceedings.

Walstra, Bouke (1985). Validating the yesterday first-time-read method. Salzburg Proceedings.