FUTURE PERFECT: UNIFORM FIXED PROBABILITIES?

Costa Tchaoussoglou, SUMMO, The Netherlands

Introduction

Recent Reading (RR) is the way of measuring readership most widely practised. Until the early '80s the Netherlands made no exception to that rule. After the first Readership Research Symposium in New Orleans, however, the First Reading Yesterday (FRY) approach was introduced by Maurice de Hond (Inter/View). His aim was to tackle two major readership measurement problems encountered when using RR.

First, the measurement of FRY intended to minimise the dependence on memory, the accuracy of which represents a problem for the RR method that becomes the more serious, the less frequently a title is published. Second, the RR model bias due to parallel and replicated readership would be reduced.

For well-known practical reasons the measurement of FRY has been linked to Computer Assisted Telephone Interviewing (CATI). In 1986 the joint-industry media research organisation SUMMO adopted the CATI FRY system of Inter/View. After ten years of FRY experience time has come to draw up the balance.

The Current FRY Measurement

Since 1982 the original concept of Maurice de Hond has been improved several times and the more so after the adoption by SUMMO. In the current question-centred system all titles are read out in order to establish which magazines were 'ever' read and which ones had been read or looked 'yesterday'. For 'yesterday' reading of a magazine there is first the question whether it took place at home, at home and somewhere else or only out-of-home.

For yesterday reading at home, the interview then established whether the magazine had arrived that day or not. If it had arrived earlier, whether yesterday was the first time it had been looked at and if that was claimed to be the case, whether there had been contact with the magazine when it had first arrived.
For TV-guides a specially adapted version of the questionnaire had to be developed, in order to solve the problem of two consecutive issues (current week and next week) being read interchangeably by subscribers.

For every title claimed to be 'ever' read the source of copy (subscription, single copy, reading circle, other) and frequency of reading (how many out of the last six issues) are established.

The Calculation of Average Issue Readership

In theory, AIR-figures for a weekly magazine can be calculated as the number of FRY times the number of interviewing days, i.c. 6, during a week. The current practice in the Netherlands, however, is rather different. In order to achieve more reliable figures AIR is not derived directly from the number of FRY for each title, but indirectly from the calculated reading probabilities for different groups of titles. Not only does this grouping of titles lead to less validity, but the grouping itself is open to question, i.e. what (objective) criteria should be used to group the magazines?

For each group of weeklies separately, reading probabilities are calculated as: 6 times the number of FRY, among readers with equal frequency claims, divided by the number of readers with that particular frequency claim.

When it occurs that a reading probability is calculated above 1.0 , that probability is arbitrarily fixed at 0.97 and the other probabilities of the group of titles in question are down-weighted accordingly:
if for example a 'probability' 1.05 is calculated, all probabilities are down-weighted by a factor 0.92 (= 0.97/1.05).

In spite of all the efforts that have been made to improve the method, it is still suffering from this obstinate disease.

Is FRY to be sacrosanct forever?

Although the method has not changed since the end of 1988, that was by no means our intention at that time. On the contrary plans were made to calibrate, at least for larger titles, the 'group' probabilities by the individual total number of FRY. For the smaller titles calibration was not feasible and therefore it was decided to maintain the existing procedure as it were.

The same fate was granted to proposals intended to improve the current classification into groups, because neither of them was really based on relevant objective criteria and thus justify the resulting changes in readership levels.

Meanwhile we carried out a Recent Reading experiment in a CATI setting, but that came to an untimely end (Marion Appel and Costa Tchaoussoglou "Will the future find Utopia in the Netherlands?", Hong Kong 1991).

Gradually we had to accept the fact that the final result of our modus operandi is AIR figures that lean heavily on the 'ever read' and frequency questions, instead of the individual FRY scores. And in the case of a calculated probability above 1.0, even the FRY scores of a group of titles are no longer prevailing.

Uniform Fixed Probabilities

This conclusion led us to consider the concept of readership figures for all titles based directly on the 'ever read' and frequency questions. The general idea of using uniform fixed probabilities was accepted, but before putting such a system into operation several questions had to be answered:

- has the present frequency question the same meaning to all respondents, and is it independent of the title in question?

- are other 'ever read' filters and/or frequency questions more reliable?

- what probabilities should be attributed to the different frequency claims?

- is there any evidence for exceptions to the general rules?

- in what way can the experience with FRY be applied?

Frequency Question: "Usual" behaviour or past period?

We all know that the answers to numerical reading frequency questions if taken at face value are wrong. Regardless whether the questions are expressed in terms of 'usual' behaviour or of reading in a specific past period, regular readers tend to over-state their frequency, while the opposite is true for irregular readers. With this rule of thumb in mind we tested the question 'how many issues out of six do you read usually?' against 'how many issues out of the last six?', in order to find out whether 'usual behaviour' would produce less probabilities above 1.0.

The results of this test were rather confusing. Indeed, there was a tendency of decreasing probabilities among 'the six out of six'. This was caused by a substantial increase in the number of regular readers. Such an increase was not only regarded as highly implausible, but also we were aware of the fact, that changing over to 'usual behaviour' would bring about a revolution in readership levels. Therefore it was decided to stick to the 'familiar' question, because otherwise we would loose ten years of experience and have to start all over again in building a new system.

Inherent in this decision was to maintain the use of a single scale applying to all publications and to accept the fact that respondents may find it easier to give an estimate of the number of issues read out of the last six published for a daily newspaper than for a weekly or monthly magazine. This was not felt as an insurmountable problem, because we had no intention to estimate AIR directly from the answers to the question on reading frequency.

Intermezzo: FRY versus Frequency

Wrong frequency estimates are certainly not the sole explanation for calculated probabilities above 1.0. For several titles a similar phenomenon is observed, when probabilities are calculated for subscribers, regardless of their frequency claim. One might object that these findings are subject to sampling errors in the small numbers of FRY scores. Therefore we made the same analyses for large groups of titles.

The following example will demonstrate quite clearly the inconsistency in our FRY-model:
From 17,200 male subscribers to monthly magazines 10,664 claimed to have a frequency of 'six out of six'
and 469 had read an issue yesterday for the first time. The calculated probability for this case equals   1. 0
(= 25 * 469 / 17,200). Supposed that this is caused by wrong frequency claims only, then at least 45 FRY
claims must have a lower frequency. That would imply probabilities of 1.0 for all frequencies ranging from '
one out of six' up to 'six out of six'; 691 FRY claims among 14,792 readers:

$$(469 + 122) * 25 / (10,664 + 4,128) = 1.0.$$

For other cases where we find probabilities above 1.0, the supposed mistake in the frequency claim must
range at least from 'four out of six' up to 'six out of six'.

The conclusion is straightforward: the inconsistency in the FRY-model will not be cured by a more accurate frequency scale.

The analyses made us also aware of the fact that readers with the same frequency claim, may differ
significantly in calculated probabilities, according to different answers to the source of copy question. In
general probabilities decrease from subscription to 'other':

weekly magazines
women reading probabilities (FRY)

frequency

read out of six	subscrip-tions	single copy	reading circle	other
0	.12	.06	.19	.04
1	.56	.25	.34	.25
2	.72	.34	.44	.25
3	.81	.59	.50	.43
4	.88	.72	.60	.50
5	.87	.72	.60	.83
6	1.00	1.12	.84	.74

These results gave rise to more questions:

.   Are these differences only to be found in a FRY-model or would they also occur if the probabilities were based on Recent Reading?

.   Are there other criteria than source of copy, which would show a similar effect?

Another experiment with Recent Reading

From the available data we cannot decide whether FRY or frequency is responsible for the observed
fluctuations. Therefore another RR-experiment was conducted in which the 'FRY-questions' were
substituted by a simple 'last week/month' question. The results of this experiment showed the same patterns
as in the FRY model:

weekly magazines
women reading probabilities (RR)

frequency	subscrip-tions	single copy	reading circle	other
0	.19	.03	.17	.04
1	.36	.18	.54	.22
2	.63	.32	.41	.25
3	.56	.41	.79	.46
4	.59	.58	.62	.60
5	.86	.75	.72	.57
6	.89	.78	.80	.69

Thus, we concluded that the frequency question does not have the same meaning in terms of probability of
reading to all readers.**Evidence gathered from AID and Regression analysis**

Using AID (Automatic Interaction Detector) and regression analysis we were able to identify three more characteristics of readers, with a significant contribution to the variance in the FRY-scores. The analyses were done separately for weeklies and monthlies, with the number of FRY-scores in the 'six out of six' group being the dependent variable. Sex, age and education proved to be significant independent variables. Sex was not significant for all groups of titles. The youngest age group turned out to produce incredibly high FRY-scores. To a lesser extent this applies also to the higher educated.

Approaching the final stage: Assessing the probabilities

There is still some analytical work in progress. To present a first draft of the model at this very moment may therefore be considered as rather speculative. Nevertheless the outlines of the model are already quite sharp.

First of all we group titles only by frequency of appearance i.e. daily newspapers, weekly and monthly magazines. Reading probabilities will uniformly be assigned to all titles within a group.

The analyses made of FRY and RR data up to now, suggest that we take care of the following general rules when we assign distinct probabilities:

> (all other things being equal)
> . the higher the frequency claim the higher the probability
> . the 'higher' the source of copy the higher the probability
> . women will get higher probabilities than men for woman's magazines and vice versa
> . young people will get higher probabilities than older people for youth magazines
> . subscriptions included in membership fees and non-paid subscriptions will have lower probabilities

Assigning probabilities according to these rules does not lead us yet to a unique solution. In other words the level of the resulting AIR figures can be 'chosen'.

In this crucial stage we have to keep in mind that, for almost seven years the SUMMO readership figures have been widely accepted. Furthermore we are not in search of the gold standard, but merely trying to improve the current system that has some arbitrary elements. What we want is to develop a system that is ' fair' to all titles, i.e. offers 'equal' treatment unless there is objective evidence to differentiate.

We find ourselves between Scylla and Charybdis. We have to deal with the reluctance to quit FRY, as it is the symbol of readership research in the Netherlands. On the other hand we feel the necessity to admit to ourselves openly that in that in practice FRY plays a minor role.

The solution we are thinking of is to assess the reading probabilities along the general rules as mentioned. Simultaneously the probabilities should be fixed at such a level that the average level of the current readership figures is preserved as much as possible.

A first trial on 10 magazines showed that such a procedure is technically feasible:

	current	trial
TV-guide	12.1	11.8
News weekly	4.8	6.0
TV-guide	24.9	24.2
Woman's weekly	26.5	26.0
Illustrated	18.1	18.0
Gossip	19.5	19.0
Glossy monthly	3.1	3.2
'Club' magazine	34.5	34.1
General monthly	8.7	8.6
Special interest	6.9	6.9

The question remains, however, whether these readership figures will be accepted. Even if it turns out that there is only a marginal difference with the current figures, the fact that they are no longer directly based on some 'outside' criterion like FRY or RR, may cause an emotional shock to some people.