ACCOUNTING FOR WHAT WE ARE DOING: RECENT READING REVISITED

Andraž Zorko, CATI

There have been numerous papers and discussions on recent reading (RR) and its accuracy during the past years. This comes as no surprise. It is a method used in over 90% of NRS's throughout the world. At the same time it is the method with the greatest number of critics and possibly the greatest number of known model biases. I will not discuss those in my paper, as there is even a song about it. We know of telescoping, replicated readership, parallel readership and RQP ratio, all of them producing over- or under-estimations of AIR. A lot of effort was placed in the various ways of enhancing the method, yet it is obvious that none of them can eliminate the majority of the bias that occurs while estimating AIR with RR. Many analyses have shown that the RR method overestimates the AIR, although in some cases underestimation can also occur. Ron Carpenter (1999) has shown that there is an average overestimation of 33% to 50% for women's weeklies and women's monthlies respectively according to FRY and circulation methods (primary household copy). Therefore there is no doubt that RR overestimates the AIR, however at the same time we know it remains the method that offers us the best compromise between feasibility, cost and usability. TTB is not really feasible, FRY is too expensive and lacks usability (duplication data), FRIPI is just a modification of RR and Frequency of reading also suffers from overestimating. The question that arises is whether there is really nothing we can do about it?

I think we can all agree upon the statement that FRY – at least on a theoretical level - is the method with the most accurate estimations of AIR. However, as it does not provide duplication data it is unacceptable for media planners. Ideally, we should have a RR based survey with FRIPI questions to eliminate the replicated readership and an additional set of questions on the number of different issues read within the last publishing interval in order to eliminate the parallel readership effect, whereas at the same time FRY should be measured and used as a correction factor to deal with recall problems, i.e. telescoping. Is such a model feasible?

The Slovenian NRS model

In the year 2000 a JIC for print media was finally established in Slovenia. RR was prescribed as the currency method, but a sort of correction method was also demanded from the contractor. As we were conducting a media survey with the FRY method at the time we proposed a combined survey model that won the NRS. At this point it should be mentioned that at the time being we still do not have an ABC in Slovenia.

The NRS model in Slovenia now consists of two surveys. First there is a telephone CATI survey (sample size n=90 per day) in which the FRY method is used based on a spontaneous recall of yesterday's reading. It starts with the general question on whether they have read any newspaper or magazine the day before the interview and which one was it if the answer is yes. This is followed by the same question for five categories (dailies, supplements, other newspapers and magazines, business and computer press, free press). At the end there is another question on whether the interviewee remembers any other titles he/she has read yesterday (the recall has improved by nearly 20% after introducing this question) while the interviewer can jump backwards through the sections if necessary (for example when the interviewee suddenly remembers another daily he/she read yesterday while thinking of other newspapers). The first reading occasion is then measured for each title that was read on the previous day ("was this the first reading of this particular issue..."). We have conducted FRY in this way since 1999 (back than on a daily sample of n=150) and the approach has been modified since then according to the analysis and experiences so we now believe we have the optimum approach for this kind of conducting FRY by phone. However, there is the presence of an anomaly, which is discussed further on. In 2002 the response-rate for the telephone part of NRS was 57.5%.

The interviewee is then asked on whether he/she would accept our interviewer at their home for a longer face-to-face CAPI interview. In 2002 the daily sample of the F2F survey is n=22, while the response-rate was 49.3% in comparison to the telephone sample. There are of course some deviations in the demographics and reading habits when both samples are observed, but as we know exactly who says "no" to the F2F interview we can and we do weight the sample accordingly (the F2F sample is therefore representative by sex, age, education and region according to the entire population and by reading habits according to the telephone sample; day of week is also included). In this survey RR with the following characteristics is used:

- Title recognition is prompted with mastheads (six of them on one screen, dailies are put together on one screen, others are mixed randomly, except for the ones that look the same or have similar names, which are also put together on one screen to avoid title confusion, the screens are shown in random order);
- the questions are open-ended in order to avoid the RQP effect;
- the event history calendar is used in order to help the interviewee while recalling the last reading event (we use a three month calendar which includes major political, international, sports etc. events, which is then filled with the interviewee's personal events of greater importance), the number of days since the last reading occasion is the answer;
- a question on the first reading of the issue last read follows in order to obtain the FRIPI estimate of AIR;
- a question on the number of issues read on the day of the last reading event is also included (this is necessary in order to calculate the PEX score).

The frequency of reading as well as numerous questions on the quality and circumstances of reading on the last reading occasion are also included. Before proceeding to the RR analysis of the AIR estimates allow me to put forward some other interesting findings from our NRS model:

- we found out that the FRIPI estimate can be up to 7% lower than RR on average and that there can be a difference of 2 points in the FRIPI estimate if different wording is used (WAS this the first... vs. WHEN was the first..., the latter producing lower AIR scores);
- according to the research similar to NRS which preceded our NRS model we found out that the event history calendar has reduced the AIR estimations by 5%;
- last but not least, the CAPI approach has produced a 12% higher readership measured in general if compared to the paper and pen method (the first quarter was performed by the paper and pen method).

Therefore we managed to establish a model within which the currency is based on the RR estimate, however the FRIPI question is also used in order to control the replicated readership effect (although RR estimate remains the currency at the time being). Furthermore we have managed to get the FRY estimate on a yearly sample four times bigger than the F2F sample. Still, two limitations remain: first, there is a slight problem with FRY when spontaneous recall is used, and second, after one year we do not have a reliable FRY estimate for every title included in the research. If we could eliminate these two problems FRY could be used as the correction method.

The Spontaneous Recall Problem.

Back in 1999, when the first results were obtained using the FRY method, we could clearly observe certain anomalies. After improving the spontaneous recall approach there were less anomalies but some of them seem to be impossible to eliminate. The problem was that we did not have a tool to identify particular "strange" result as an anomaly. However, with the NRS model such anomalies are now easily detected. As we have questions as regards the first reading within the publishing interval in a F2F survey and as the answers to RR questions are stated in terms of days it is possible to calculate a sort of FRY estimate also in F2F surveys. However, as it is based on a RR question we cannot treat it as a true FRY estimate. Therefore, we were not surprised to find out that these estimates are also a subject of some kind of short-term telescoping. In case of weeklies the shares of answers from 1 to 7 days ago should be – theoretically - distributed equally, but as we can observe in Table 1 this is not the case. The share of "yesterday" is almost twice as high as it should be. The table also clearly shows that the recall obviously works accurately only as far as up to three days in the past. Last but not least we found out that the deviation from theoretical figures correlates to the education of interviewees (Spearman's Rho = –0.66, $p < 0.01$) – the lower the education the higher the deviation.

TABLE 1 – Distribution of answers for the last reading of weeklies

answers for last reading	**yesterday**	**2 days ago**	**3 days ago**	**4 days ago**	**5 days ago**	**6 days ago**	**7 days ago**
Weeklies	10,0	5,0	5,3	4,7	4,2	3,8	4,4
weeklies without TV guides*	8,2	4,9	5,2	4,7	4,2	3,8	4,4
shares (1 to 7 days = 100)							
Weeklies	26,7	13,4	14,3	12,6	11,1	10,2	11,7
weeklies without TV guides	23,0	13,9	14,7	13,3	11,7	10,8	12,5
THEORETICALLY	14,3	14,3	14,3	14,3	14,3	14,3	14,3

* TV guides are used daily and the difference between weeklies with and weeklies without TV guides shows the effect of this fact.

If we compare FRY estimates from the RR question in a F2F survey with the FRY estimates from a telephone survey we can observe some great differences. Some of the titles have an index of 400 or more, which indicates an anomaly. All such titles can easily be placed in four groups: enigmatic, titles for children under 7 years, free papers which are not delivered at home and titles the title of which is a version of its vehicle and/or which are an integral part of its vehicle. All of them have a clear common characteristic – they are easily forgotten when spontaneous recall takes place. Obviously, we cannot use FRY estimates for such titles.

The Sample Size Problem

Another group of titles that we cannot use the FRY estimate for correction are the ones with an insufficient annual sample. In the year 2002 there were 86 titles for which we had a sample of n=20 or higher as a base for the FRY estimate. So we need a way to produce a FRY estimate for the remaining 63 titles if we want to use FRY as a correction method. How can we do that? The answer is: we could do that by means of analysing the difference between RR and FRY estimates on the base of the titles in which the analysis can be performed in order to obtain a sort of weight with which we can estimate the FRY result based on the RR result. We found regression analysis to be the appropriate approach to use for such a purpose.

The Analysis

The sample

The sample used is a total sample of both parts of the surveys between January 2002 and June 2003. The size of the telephone and F2F sample is n=49.477 and n=11.630 respectively. However, the sample for the analysis can be found in the titles and their characteristics.

The Titles

There are 149 titles included in the NRS database. 80 titles have met both criteria that were set in order for the title to enter the analysis:

1. The sample of its RR and FRY figure had to be at least n=20 (in this event CV is around 0.2, which is reliable enough for use in such an analysis).
2. The title should not suffer from the described anomaly at its FRY figure.

The Variables

After some preliminary analyses the following variables were included in the final analysis (the names of the used analysis can be found in brackets). Each title included is described by:

1. readership
   - RR estimate of AIR from a F2F survey (RR)
   - FRY estimate of AIR from a telephone survey (FRY)
   - the difference between both calculated as a ratio FRY/RR (RATIO)
   - this difference ratio is then normalised with the natural logarithm in order to achieve comparativeness between over- and underestimations; value below 0 therefore indicates overestimation of RR and vice versa (RATIO ln)
   -
2. demographics
   - the share of women in the RR audience as an indicator for gender (SEX)
   - the average age of the RR audience (AGE)
   - the average education of the RR audience, calculated from a 6 point scale ranging from 1 (not completed elementary school) to 6 (university degree), where those who are still attending school are excluded (EDU)
   -
3. type of readers
   - the share of secondary audience in the RR audience consisting of all who obtained the last read issue by mere coincidence, that is in waiting rooms, visiting friends, etc. (SECOND)
   - the share of regular readers in the RR audience (REGULAR)
   -
4. quality of reading
   - the average number of reading days as reported by interviewees (DAYS)
   - the average time spent on reading the issue from the first to the last occasion of reading, in minutes (TIME)
   -
5. the issue period (PERIOD)

I suppose one could ask what are the reasons behind selecting a particular variable. Education has proven to be indicative for the accuracy of answering (as shown above) and age does influence the memory and ability to report past events. Why sex? Well, why not? Furthermore, some of our past analyses have shown that women tend to answer the questions more accurately than man. The influence of the secondary audience was also shown above, while regular readers could be the ones who exaggerate when answering the question on their last reading (answering what they usually read instead of what they actually read). The number of reading days and the time spend on reading the issue could also be interesting as we assume that it is more likely to remember the day of the last reading more accurately if the reading time was longer. The issue period is a natural choice I suppose as there were many arguments that RR mainly overestimates in the event of monthlies combined with several other characteristics such as content and even its robustness (Shepherd-Smith, 1999).

The Foreplay

Before proceeding to the final analysis let's take a look at some interesting figures on RR and FRY estimates as well as on the index of RR readership over FRY readership. We can observe that RR overestimates the total readership by 33% and that the average overestimation per title is approximately 50%. You will also notice that there is approx. one quarter of titles where RR underestimates the AIR if compared to the FRY estimate.

TABLE 2 – Basic statistics of RR and FRY readership

	RR	FRY	INDEX (RR/FRY*100)
Sum	505,6	380,6	133*
Minimum	0,9	0,3	52
Maximum	28,0	18,4	335
Mean	6,3	4,8	153
Median	4,1	3,0	148
Std. Deviation	6,0	4,5	65
Quartile 25	2,4	1,5	99
Quartile 50	4,1	3,0	148
Quartile 75	7,5	6,8	199

* this index refers to the Sum readership of RR and FRY, all other statistics refer to INDEX

Of course, if the majority of criticism of the RR method is aimed at the over-estimation of the AIR – some of them even talk about double counting – one is interested in the relatively large proportion of titles where RR underestimates the AIR. The hypothesis here is that this phenomena occurs with titles with a larger proportion of secondary audience, which could underlie the parallel readership phenomena (for example: reading in the waiting rooms at doctors, hairdressers etc.). The assumption which underlies the hypothesis is that if somebody has read a particular title by coincidence he would be more likely to forget it when thinking of the last readership in a RR interview (especially if this took place more than a few days ago) while on the other hand one would remember it when asked about the yesterday's reading in a FRY interview. Furthermore, if he/she reads paricular magazine at the waiting room for example, it is more likely he will read older and more than one issue. The data confirmed this hypothesis. There is a significantly higher proportion of secondary audience within the titles with underestimated AIR by RR. While all observed titles have an average of 28.5% those with an underestimated AIR by RR have an average of 33% ($p < 0.1$) and the other titles have an average of 26.9%. We found another significant difference related to the proportion of the secondary audience – titles with underestimated AIR by RR have a younger audience ($p < 0.01$). The explanation could lie in the fact that younger people have a greater probability to run on titles they do not read on a regular basis as they spend more time out of home than elderly people.

In the tables below we present the basic statistics for the variables included in the analysis.

TABLE 3a – Basic statistics of the included variables: Descriptives

	RATIO ln	AGE	SEX	EDU	SECOND	REGULAR	DAYS	TIME
Mean	-0,33	37,2	51,2	3,9	28,5	40,3	1,8	26,8
Median	-0,39	38,5	51,3	3,8	29,2	35,2	1,7	27,1
Std. Deviation	0,45	7,5	17,8	0,4	14,5	16,4	0,6	5,2
Minimum	-1,21	19,3	9,3	2,8	2,1	11,2	1,0	11,5
Maximum	0,64	52,7	87,9	5,0	61,3	76,3	3,6	39,9
Quartile 25	-0,69	32,6	42,2	3,6	17,8	26,5	1,4	23,3
Quartile 50	-0,39	38,5	51,3	3,8	29,2	35,2	1,7	27,1
Quartile 75	0,01	42,9	62,7	4,1	39,3	55,1	2,1	30,0

TABLE 3b – Basic statistics of the included variables: Frequency for issue period

	Frequency	Percent
daily	6	7,5
twice weekly	2	2,5
weekly	29	36,3
fortnightly	6	7,5
monthly	36	45,0
bi-monthly	1	1,3

We might have a slight problem here as the distribution of this variable is clearly bimodal and moreover the sample for non-weeklies and non-monthlies are not large enough. Above all we have discovered that the difference between FRY and RR varies significantly ($p < 0.1$) by category – but not as it would be expected. Although dailies have the smallest difference (and it would

be even smaller if it were not for two specific specialised titles) monthlies for ex. have a smaller difference than weeklies. We can therefore assume that it is not the issuing period that is decisive when RR is over-estimating the AIR as some authors have implied in the past...

TABLE 3c – The FRY/RR difference by issuing period

	Frequency	RATIO ln
daily	6	-0,18
twice weekly	2	-0,87
weekly	29	-0,49
fortnightly	6	-0,23
monthly	36	-0,21
bi-monthly	1	-0,44

TABLE 3d - Basic statistics of the included variables: Correlations (Pearson)

	RATIO ln	AGE	SEX	EDU	SECOND	REGULAR	DAYS	TIME	PERIOD
RATIO ln	1,00	-0,42	0,07	0,37	0,04	-0,22	0,00	-0,08	0,21
AGE	-0,42	1,00	0,07	-0,26	-0,59	0,49	-0,28	0,03	-0,27
SEX	0,07	0,07	1,00	-0,15	0,10	-0,02	-0,03	-0,12	0,18
EDU	0,37	-0,26	-0,15	1,00	-0,01	-0,18	-0,02	-0,11	0,30
SECOND	0,04	-0,59	0,10	-0,01	1,00	-0,74	0,36	0,39	0,37
REGULAR	-0,22	0,49	-0,02	-0,18	-0,74	1,00	-0,02	-0,18	-0,06
DAYS	0,00	-0,28	-0,03	-0,02	0,36	-0,02	1,00	0,52	0,50
TIME	-0,08	0,03	-0,12	-0,11	0,39	-0,18	0,52	1,00	0,22
PERIOD	0,21	-0,27	0,18	0,30	0,37	-0,06	0,50	0,22	1,00

* Spearman's Rho is used instead of Pearson as the distribution of this variable is not normal

Regression Analysis

The goal of this analysis is to predict the FRY estimate of AIR solely on the RR audience variables. At the same time we can observe which variables influence the accuracy of the RR estimate of AIR the most and therefore explain especially the overestimation of the RR method.

Linear regression analysis is performed by a method in which RATIO ln is dependent variable while the rest are independent variables. With this model we can explain 42% of the total variance ($R = 0,65$; $R^2 = 0,42$). The table below presents the results.

TABLE 4 – Regression analysis

	B	Std. Error	Beta	t	Significance	Tolerance	VIF
(Constant)	1,10	0,71		1,55	0,13		
AGE	-0,04	0,01	-0,63	-4,66	0,00	0,45	2,21
SEX	0,01	0,00	0,21	2,12	0,04	0,87	1,15
EDU	0,13	0,11	0,13	1,25	0,22	0,73	1,37
SECOND	-0,02	0,01	-0,77	-4,01	0,00	0,22	4,48
REGULAR	-0,01	0,00	-0,42	-2,59	0,01	0,31	3,21
DAYS	-0,07	0,10	-0,09	-0,74	0,46	0,53	1,89
TIME	0,01	0,00	0,22	1,72	0,09	0,51	1,95
PERIOD	0,01	0,00	0,20	1,83	0,07	0,66	1,52

In a way the results are surprising. While they confirm the assumption for regular readers (higher share means higher overestimation), age (older audience means higher overestimation), sex (higher share of women means lower overestimation) and reading time (higher reading time means lower overestimation) they are surprising in the case of education and secondary audience (the issue period is not a surprise any more as we have seen above in table 3c). Education was expected to influence the difference as it has a relatively high correlation (0.37, $p < 0.01$) with the difference ratio and does influence the accuracy of answering as we have seen above. Apparently other variables "took its glory away". But that is why we are using regression analysis instead of simple bi-variate correlation analysis. For the secondary audience we have already previously shown that the

share is higher in the case of titles with underestimated AIR. Now the analysis is telling us that a larger proportion also means higher overestimation. So we took a deeper look and found out that there is a difference among the titles with overestimated AIR, namely the ones that have lower overestimation have the lowest proportion of secondary audience. This fact already indicates the next move necessary.

As the RR method both over- and underestimates the AIR we should do the regression analysis solely using the titles with overestimated AIR. In this case the sample is perhaps to small as there is a general rule of 10 units per one predictor. However, it seems to work. The explained variance remains the same while we have "lost" the significance at two predictors. The share of women and total reading time are no longer significant. We can therefore conclude that overestimations happen only in the case of titles with an older audience and with those with a larger proportion of secondary audience and regular readers while the influence of sex and total reading time is significant only when titles with underestimated AIR are also included in the analysis. This should indicate that sex and total reading time are important variables with such titles but the average for these two variables among such titles does not significantly differ from the overall average. The sample of the titles with an underestimated AIR is too small to perform a regression analysis.

Now let's take a look at how reliable is the prediction of the difference ratio (RATIO est.) if we use the coefficients above with the RR audience data. The regression formula is applied and 'delogarithemised':

RATIO est. = exp (1.101 - 0.038 * AGE + 0.005 * SEX + 0.134 * EDU - 0.024 * SECOND - 0.011 * REGULAR - 0.074 * DAYS + 0.007 * TIME + 0.007 * PERIOD)

In the table below the basic statistics are shown. The correlation between the actual difference (RATIO) and prediction (RATIO est.) is 0.58 ($p < 0.01$) with the average relative error of 27.3%. The result indicates that there might be a problem with the titles with underestimated AIR. In addition, if we look at particular titles there are some substantial differences. That is why it is necessary to introduce an additional independent variable into the analysis.

TABLE 5a – RATIO est. compared to RATIO

	RATIO	RATIO est.
average	0,79	0,73
min.	0,30	0,36
max.	1,90	1,20

If the RR method would produce merely overestimations we would have no problem at all. But since every fourth AIR estimation within the analysed database is underestimated according to FRY method we must introduce a variable to control the two-way RR deviation effect. Therefore we have introduced a dichotomised variable (RATIO d) where value 0 indicates the titles with overestimated AIR according to RR and value 1 indicates the titles with underestimated AIR. Such a variable splits our regression model into two parts instead of splitting the analysis into two populations of titles.

The formula for predicted difference now stands like this:

RATIO est. = exp (-0.843 - 0.015 * AGE + 0.002 * SEX + 0.156 * EDU - 0.012 * SECOND - 0.005 * REGULAR - 0.076 * DAYS + 0.001 * TIME + 0.005 * PERIOD + 0.702 * RATIO d).

The imposed variable has performed its task successfully. On average we have improved the reliability of the prediction by over 10%. The correlation between the actual difference (RATIO) and prediction (RATIO est.) is now 0.90 ($p < 0.01$) instead of 0.58 while the relative error is now below 20%, namely 16.8%. There are also fewer relevant differences at the title per title level. Table 5b shows the improvement.

TABLE 5b – RATIO est. with imposed variable compared to RATIO

	RATIO	RATIO est. with imposed variable
average	0,79	0,76
min.	0,30	0,40
max.	1,90	1,74

However, if we want to use the regression coefficients to estimate the FRY score based on the RR score and RR audience characteristics we do not have such a variable in the event of titles with an anomaly or a too small sample. We must therefore analyse the difference between the estimated and actual difference ratio before introducing the imposed variable.

The Analysis of the Estimated Difference Ratio

These are the basic statistics for the relative error of the estimated difference ratio. We can see that 25% of the titles have an acceptable error of 11% or less while 25% of the titles have a relative error of approximately 35% or more which can not be

accepted. We will therefore try to analyse what are the common characteristics of these titles. There is only one title with an error of over 100% (141%) that is why the Mean value differs substantially from the Median, which is therefore a better estimator for average in this case.

TABLE 6 – Basic statistics for relative error of RATIO est.

	relative error of RATIO est.
Mean	0,27
Median	0,21
Minimum	0,00
Maximum	1,41
Quartile 25	0,11
Quartile 50	0,21
Quartile 75	0,35

We have classified the titles into four groups according to the level of the relative error: titles with an error lower than 11% (20 titles), titles with an error of 11% or higher but lower than 25% (28 titles), titles with an error between 26% and 39% (16 titles) and titles with an error of 40% or over (16 titles). The goal of this approach is to find the common characteristics of titles with a similar relative error. The table below shows our findings.

TABLE 7 – Average AIR by RR and FRY and average difference ratio by groups of relative error for RATIO est.

relative error ->	up to 11%	11% - 25%	26% - 39%	40% or more	Total
AIR by RR	8,6	6,4	6,6	3,1	6,3
AIR by FRY	5,4	4,2	7,3	2,4	4,8
RATIO	0,6	0,7	1,1	0,8	0,8

This table clearly shows two things:

1. we do have a problem with the titles with underestimated RR, which are obviously more commonly represented in the group with a relative error ranging between 26% and 39%;
2. the group with the highest error has a significantly lower average AIR.

As there is nothing we can do about the first problem (if we want to estimate the AIR based on RR audience we can not predict – at least not empirically - if a particular title's AIR by RR is under- or overestimated according to FRY) we must focus on the second one. It is obvious that the relative error is higher if AIR is lower on average. Lower AIR means lower sample. Therefore we will repeat the regression analysis, but this time on a sample of n=64 titles instead of n=80. This time the sample limit criteria is raised to n=30 which is a very well known limit in statistics. This will also represent a test of the stability of our first solution - if our model is stable then the coefficients should not change substantially.

The next table clearly shows that our regression model is stable. The coefficients have not changed substantially, the explained variance is higher (over 50% - which is rarely achieved) and the relative error while predicting the difference ratio is also lower on average. However, if we use the new coefficients with the titles that were excluded, this solution seems to be worse. While on one hand we have improved the accuracy of predicting with the titles that were left in the sample we have actually worsened the accuracy with those that were excluded. Not dramatically, but still... It looks like using this approach is like whipping the sword of Damocles in a way...

TABLE 8 – Comparison of the results from two samples

sample of titles	80		64	
explained variance	42%		55%	
	B	Significance	B	Significance
(Constant)	1,101	0,13	1,086	0,14
AGE	-0,038	0,00	-0,038	0,00
SEX	0,005	0,04	0,007	0,00
EDU	0,134	0,22	0,133	0,21
SECOND	-0,024	0,00	-0,026	0,00
REGULAR	-0,011	0,01	-0,015	0,00
DAYS	-0,074	0,46	-0,068	0,49
TIME	0,007	0,09	0,009	0,04
PERIOD	0,007	0,07	0,011	0,00
Accuracy of RATIO est.				
Mean	0,27		0,23	
Median	0,21		0,19	
Correlation	0,58		0,68	
For total sample both				
Mean	0,27		0,30	
Median	0,21		0,21	
Correlation	0,58		0,59	

And here the analysis stops ... for now.

Conclusion

So what have we achieved?

1. Over 50% of the difference between RR and FRY estimates of AIR is explained. RR works in favour of the titles with audiences that:
    a. are older
    b. have a higher proportion of regular readers
    c. or a higher proportion of secondary readership, but not in the case of the titles with the highest proportion of secondary readership, where underestimation occurs.

While there was no doubt that age influences the ability of recollection, we have proved that this is significant for RR overestimation. The other two explain in which cases the recall of the usual instead of the last reading takes place more often.

2. The secret of RR underestimations – at least the part of it that is hidden beneath the parallel readership phenomena is also revealed. The secret lies in the even higher proportion of secondary readership and the higher proportion of younger readers also plays a significant role.

3. Higher education level has proven to be a decisive factor for the accuracy of the answers on the last reading occasion. If you have an educated audience it is more likely that their answers will be more accurate. However, education is less significant than the characteristics mentioned above.

4. We have shown that regression analysis could be used to generate a correction for its estimates of AIR. However, the relative error of the predictive model is still too high for now to use it as a correction for currency. But this problem will weaken with the growth of the sample size.

5. On the other hand we can already use this approach for the titles with a so-called anomaly of the FRY method when performed with a spontaneous recall.

By the end of 2004 the sample of our NRS will double in size. As the sample will grow the sample of titles with reliable FRY estimates will grow and moreover the RR samples will grow, which means that the predictions will be more accurate. As we have seen from the analysis above the relative error of the difference estimate will be reduced. Furthermore, we will try to find

other variables that could help us improve the explained variance, which is, to be honest, not bad at all. Greater effort will be invested in lowering the relative error of the difference estimate in order to put it within the frames acceptable by the market as a correction factor for currency. The largest proportion of the effort should be placed into getting over the problem of identifying the titles with underestimated AIR by the RR method. It would be so much simpler if RR would only overestimate the AIR...

As there is no new method on the horizon and as we now know more on RR over- and underestimations we believe that finding an accurate way to correct RR estimates of AIR based on in-depth analysis of its inaccuracy is the right way towards more accurate readership figures. We have done the first step. Was it only a small step for the researcher (that is me of course) or could it also be a giant leap for the readership research community? The community shall tell.

Bibliography:

Session papers of the 9th WRRS (Florence, 1999), esp. Brown, Carpenter and Shepherd-Smith.
M. Brown. Effective print media measurement, 1999.