

# CAN COVER RECOGNITION REPLACE RECENT READING ?

**Stef Peeters, Ph.D., Jean-Matthieu Schmerber, Luc Eeckhout, CIM  
Bernard Cools – Space**

---

## **Synopsis:**

The CIM is the Joint Industry Committee that organizes the media currency studies in Belgium. Since 40 years, the press audience survey involves a face-to-face recruited sample answering Recent Reading questions (RR). Only the presentation and data collection method has evolved from paper, over CAPI, double screen CAPI to a combined CAPI-CASI approach.

This paper reports the results of a test Specific Issue Readership (SIR) questioning, based on cover recognition, integrated in the currency study itself (n = 10.500). Respondents answered the traditional RR questions and were shown afterwards the last 6 covers of titles they read during the last months (similar to the approach presented by NOM in the Netherlands). The objective was to see:

1. whether the SIR approach could become a credible, attractive alternative for RR questioning in CAWI,
2. whether the SIR approach generates sound accumulation data,
3. what kind of disruption the SIR approach would cause in the market compared to RR questioning.

To maximize our understanding of results, the new cover data were compared to other Belgian data:

1. RR data collected in the same survey for the print currency by the joint industry committee CIM, in 2009-2011,
2. cover recognition data collected in 1999 by the sales house Medialogue (*“Magazines Need Time”*, Peeters, WRSS Florence, 1999),

and the questioning procedure was analyzed further in a qualitative study.

The topic of accumulation and issue specific readership has a long history and is treated very differently around the globe. The comparison of these Belgian data allowed to end a perennial debate in Belgium in a conclusive way, both from a conceptual and an operational point of view.

## **1. Introduction**

Belgium is one of many countries where traditional recent reading questions are the basis for the average issue readership (AIR) currency. During 40 years of face-to-face interviewing, the prompting of press titles changed from written name cards, over colored mastheads on paper to today's logo's on screen. Data collection evolved from interviewer controlled paper to CAPI, over double screen CAPI to today's approach where the interviewee chooses himself between CAPI or self-paced CASI. We have known for decades that recent reading results are very sensitive to subtle changes in questioning or study design. We are also very much aware of the biases that can lead to both overestimation or underestimation of average issue readership (telescoping, replication, parallel readership...). For a long time, none of this changed the perception of traditional recent reading questioning as best practice in press audience research.

In recent years however, the openness to alternatives increased for several reasons:

1. Dropping response rates and rising costs of face-to-face research, combined with a difficult market for press, increased the pressure to look for cost effective alternatives, like mixed recruitment or online surveys.
2. Since the revival of readership accumulation research (e.g. Baim, 1999; Agresti, 1999, Peeters, 1999, Baim, 2001), time planning never made it into day-to-day work of media planners in Belgium. Results and methodology of the Belgian WAR project (Hermie, 2005) were, for example, not fully understood nor adopted by the majority of market players.
3. Thanks to the work of Intomart and NOM in the Netherlands (Faasse and van Meerem; 2003; van Meerem, 2005; Petric and Appel, 2007; Petric and Appel, 2009) and MRI amongst others in the States (Baim, 2001, Baim, 2007; Klein; 2007, Baron, 2009, McDonald, 2009), specific issue readership (SIR) based on cover recognition has reached the status of credible contender for recent reading (RR). Technology has solved the logistic problems associated with through-the-book research. At last, there is an alternative.

Publishers in Belgium generally acknowledged the potential of the new approach. They already envisioned/ imagined web based interviewing with dynamically presented magazine covers instead of boring repetitive recent reading questions... However, they also wanted to assess the kind of disruption an AIR based on cover recognition could cause in a market used to an AIR based on recent reading claims. Within this context, publishers and media agencies in Belgium asked the joint industry committee CIM to start a test that would allow to assess the merits of AIR based on cover recognition compared to RR. At the same time, CIM was urged to design a 'real life' test, not some artificial experiment with limited relevance. Like the Dutch, we were not interested as such in individual issue results, only in their average.

Admittedly, most members of the press joint industry committee were not aware that they were mounting a modern version of the historic battle between recent reading and through-the-book adepts. On past readership symposia, comparisons of specific issue readership and recent reading claims led to divergent conclusions. Three examples:

- In a carefully designed experiment (n=501, screen in: 6 months), Joyce (1985) found that full through-the-book results were equal to recent reading results for weeklies but lower for monthlies.
- Petric and Appel (2007) concluded from their validation study (n=2.989, screen in: 12 months) that AIR estimated from cover recognition was slightly superior to recent reading AIR for monthlies and clearly superior for weeklies.
- Peeters et alii (1999, n=2.528, screen in: last months) noted that AIR estimates based on cover recognition data were superior to recent reading estimates for monthlies and even more so for weeklies.

Without any doubt, differences in research design have contributed to differences in results:

- Street recruitment (Joyce, 1985; Peeters, 1999) or access panel recruitment (Petric and Appel, 2007),
- Face-to-face (Joyce, 1985; Peeters, 1999) or online interviewing (Petric and Appel, 2007),
- Full through-the-book (Joyce, 1985; Peeters, 1999) or cover recognition (Petric and Appel, 2007),
- Screen in 12 months (Joyce, 1985), 6 months (Petric and Appel, 2007) or last months (Peeters, 1999),
- RR first (Peeters, 1999), SIR first (Petric and Appel, 2007) or randomized order (Joyce 1985),
- Comparing untreated recognition data (Joyce, 1985; Petric and Appel, 2007) or modeled recognition data (Peeters, 1999, applied the so called 'First Issue Readership' modeling).

The only robust finding seems to be that the difference between recent reading and specific issue recognition is greater for monthlies than for weeklies.

## 2. Study design

The press technical committee decided that the most economic, pragmatic and efficient way to organize a conclusive test would be to integrate cover recognition within the framework of the ongoing readership currency study. This means:

- 10.500 interviews from June 2009 till May 2010, 10.500 interviews from June 2010 till May 2011.
- Face-to-face recruitment of random sample from population database (address x gender x age).
- Free choice between interviewer assisted interviewing (CAPI) or self-completion (CASI, with interviewer present).

This design avoids a number of important problems since:

- Samples for both measures were identical.
- Recruitment standards were identical to current currency standards.
- The research institute performing the test has also established the currency study for years (TNS Media).
- The comparison was untainted by simultaneous methodological changes, e.g. from CAPI to CAWI.
- All magazines great and small participating in the currency study were measured both ways (if they provided digital covers which 95 did).

The CIM Press Study consists of three parts :

- Readership questions (on average 16 minutes of question time).
- Questions on the use of other media (another 16 minutes).
- Questions on socio demographics (about 3 minutes).

The press part starts with a horizontal filter question (“Did you read or leaf through magazine X during the last months?”) for all of the approximately 200 press titles grouped by periodicity. Recent reading for screened in titles is the second horizontal question. Subsequently, three questions are grouped vertically by periodicity (daily, weekly, other) and title: frequency of reading, proportion read and acquisition mode. The order of presentation of periodicities is randomly varied. Within a periodicity, titles are shown at random in alphabetical order A-Z or in reversed order Z-A, the begin letter being also chosen at random.

The cover recognition questions were added as vertical questions for magazines only (a specific ‘day of the week question’ alternative question for newspapers will not be reported here). For the wording of the cover recognition question, we looked at the NOM questionnaire: “Which of these issues of [title] have you read or glanced through?”. Only the last 6 covers were shown for all frequencies (cfr. Petric and Appel, 2007), the oldest at the left (cfr. van Meerem, 2005). For covers identified as being read there was an additional question: “Did you read or glance through this issue during the last week [or month for monthlies]”. The questionnaire was updated 6 times a week allowing a new cover to be presented two days after its arrival at the newsstand, and interviewers were required to import updates daily. Compared to weekly questionnaire updates (Petric and Appel, 2007), this should reduce the overestimation of new issues, and more importantly, eliminate the unequal treatment of new issues from magazines with different days of publication.

Combining both RR and SIR in one study in a fixed order (first RR, then SIR) does make this test vulnerable to two potential flaws:

- It increases the duration of press questioning.  
As the SIR questions came within the first 20 minutes of the interview, the impact on fatigue or demotivation was considered marginal.
- An order effect may influence the comparison.  
From a methodological point of view, a split run study (cfr. Joyce, 1985) or randomizing the order of RR and SIR would have been the better alternative. However, this was not an option within the framework of a currency study. The risk of contamination of the RR currency, if SIR questioning preceded, was considered too important. Therefore SIR questions always came after the classical recent reading questions.

This article reports on the results of the basic cover recognition question “Which of these issues of [title] have you read or glanced through?”.

### 3. Results

#### 3.1. Observed (SIR) versus declared (RR) frequency.

A simple correlation between observed frequency and declared frequency (on a combined semantic/numeric scale) should give an idea of the consistency between the two measures at the level of the individual respondent. The correlation was calculated for each title separately:

- The average correlation, all magazines confounded, was .71 (range .46 to .84).
- The average correlation for weeklies was .76 (range .63 to .84).
- The average correlation for monthlies was .67 for monthlies (range .46 to .80).

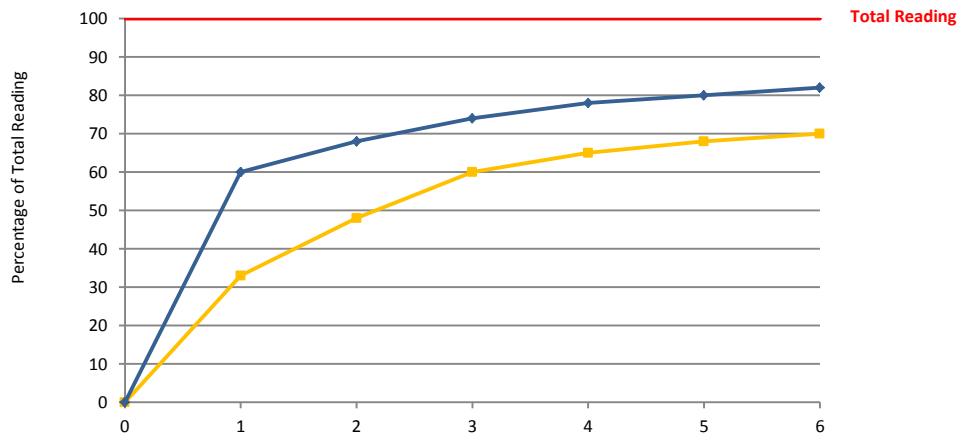
This pattern of results indicates a strong but far from perfect relationship. This is probably what can be expected from a declared ‘habitual frequency estimate’ and cover recognition for specific issues.

One might argue that the correlation is artificially inflated in our design since both questions are answered within the same 1 or 2 minutes. We do believe that chances that a response on a combined semantic/numeric frequency question is transferred to yes/no cover recognition are very small.

### 3.2. Observed versus expected accumulation curves.

This is a no brainer since the number of people reading an issue can only increase. There are plenty of variables that determine differences in pace and maximum level of accumulation (e.g. Baim, 1999; Peeters, 1999) but in general terms, raw accumulation data are expected to look like those in Graph 1.

Graph 1. Examples of expected accumulation curves



Unfortunately, this does not mean that observed issue readership results always increase continuously over time. A lot depends on the data collection strategy:

- A diary approach always produces continuously growing accumulation curves (e.g. Baim, 1999; Hermie, 2005, Siegers, 2008). It minimizes forgetting, measures all issues, including very old ones. If only it were not such a cumbersome procedure...
- Through-the-book and cover recognition methods do not produce continuously growing audiences over time. At some point in time, forgetting seems to affect results. According to Joyce (1985), on average after 6 weeks for weeklies and after 4 months for monthlies. van Meerem (2005) showed accumulation graphs for individual magazines, some of them with minor inconsistencies: at some point in time the audience seemed to decrease before increasing again. Petric and Appel (2009) also observed some irregular accumulation curves but nothing that a straightforward smoothing procedure couldn't repair.

Much to our surprise, our test results revealed a large number of irregular accumulation curves. Curves of individual magazines are detailed in attachment (separately for Dutch and French weeklies only and monthlies only). Petric and Appel (2009) made a distinction between the large majority of normal curves, showing continuously increasing accumulation, and irregular curves produced lower results than one of the previous intervals either in between or after 6 issues. Our test showed not a single normal curve, neither for weeklies nor for monthlies. In fact, less than 1 out of 10 magazines showed a peak audience after 6 weeks for weeklies or 6 months for monthlies, as shown in Table 1.

Table 1. Distribution of type of accumulation curve CIM 2010.

Type of accumulation curve	All	Weeklies	Monthlies
Continuously growing curve	0%	0%	0%
Curves with lower intermediate week	9%	12%	7%
Curves with (at least) lower sixth week	91%	88%	93%

While this phenomenon was dramatically present in the data of the monthlies (93% lower after 6 intervals), it was also prominent in weeklies (88% lower after 6 intervals). Clearly, our cover recognition results don't fit the theoretically expected. Another way of looking at the data is by identifying after how many weeks the curve shows its peak (Table 2).

Table 2. Distribution of peaks in observed accumulation curves CIM 2010.

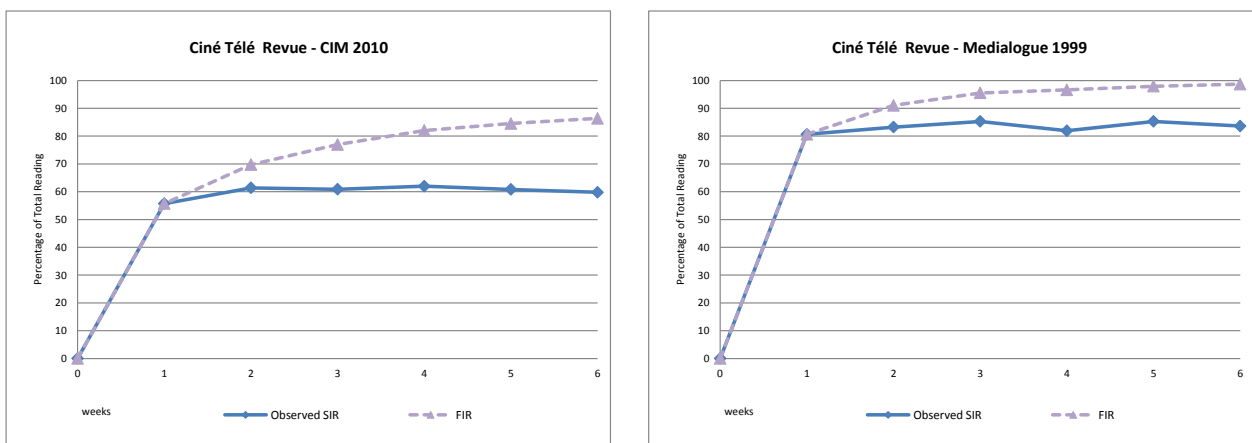
Peak	All	Weeklies	Monthlies
After 1 week or month	0%	0%	0%
After 2 weeks or months	35%	39%	30%
After 3 weeks or months	34%	24%	41%
After 4 weeks or months	16%	15%	17%
After 5 weeks or months	6%	7%	6%
After 6 weeks or months	9%	12%	7%

In our test, the peak in cover recognition for weeklies seems to be around 2 weeks, and around 3 months for monthlies.

It looks as if there is something terribly wrong with our data. We just couldn't figure out what it was. That's when we decided to compare our data with the raw data of Peeters (1999) graciously made available by Sanoma Media Belgium. The First Issue Readership (FIR) modeling applied in this study was invented by a Belgian magazine sales house that wanted to stress the importance of magazine accumulation and identify factors that influence pace and accumulation levels. In attempting to identify the determinants of accumulation, they chose a model that in fact maximizes the accumulation. They used the sum of all readers who read any issue of a magazine for the first time within the reference period as an estimate of the accumulated audience, thus creating by definition a continuously growing function.

For our purposes, we were interested in a direct comparison of the raw data. While the recruitment procedure, sample size, time frame and number of magazines involved was different from our test, the actual sequencing and formulation of questions was very similar. Graph 2. shows the raw data (and, in the dotted line, the FIR modeling) for the largest television weekly in Belgium. On the left the CIM 2010 data, on the right the 1999 data.

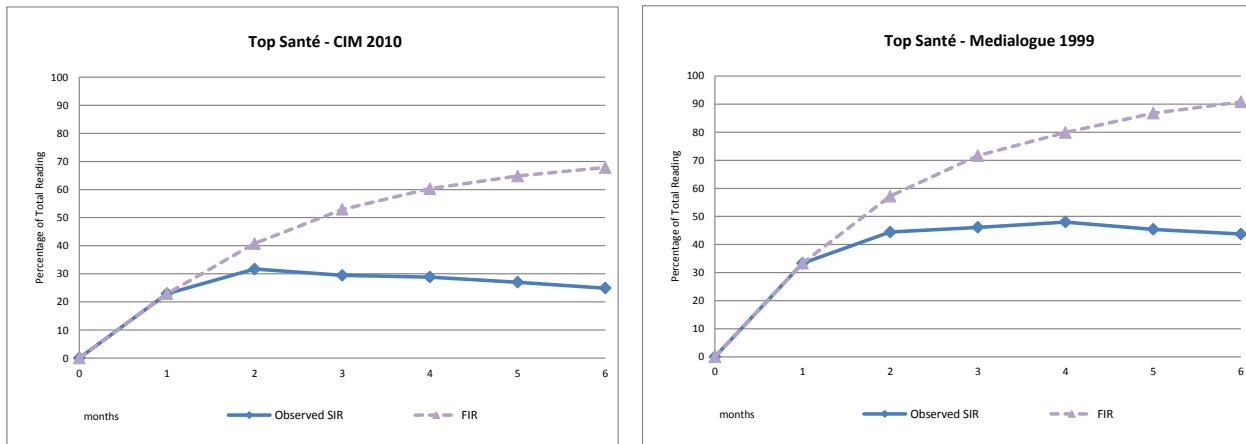
Graph 2. The observed accumulation of the largest television weekly in 2010 and 1999.



Independently of the differences in overall level, there is a striking resemblance in the raw data. After two weeks, there is no significant increase. On the contrary, at different points in time, the graph shows a (small) loss of audience.

Maybe the velocity of a television magazine is exceptionally high and atypical ? Graph 3. shows the accumulation of the largest monthly in Belgium. The observed cover recognition data are far from continuously growing. The FIR modeling in all likelihood overestimates real accumulation.

Graph 3. The observed accumulation of the largest monthly in 2010 and 1999.



To allow a more comprehensive comparison, we calculated the distribution of the type of curves and the distribution of peaks on the original Medialogue 1999 data. In Table 3, results are presented for all magazines confounded, and separately for weeklies and monthlies.

Table 3. Comparison of distribution of type of accumulation curve 1999 - 2010.

Type of accumulation curve	CIM 2010			Medialogue 1999		
	All	Weeklies	Monthlies	All	Weeklies	Monthlies
Continuously growing curve	0%	0%	0%	0%	0%	0%
Curves with lower intermediate week	9%	12%	7%	3%	0%	7%
Curves with (at least) lower sixth week	91%	88%	93%	97%	100%	93%

This historical comparison strongly suggests that the large proportion of irregular accumulation curves was not an accident. It's a phenomenon that can be replicated and that needs explaining.

Table 4. Comparison of distribution of peaks in observed accumulation curves 1999 - 2010.

Peak	CIM 2010			Medialogue 1999		
	All	Weeklies	Monthlies	All	Weeklies	Monthlies
After 1 week or month	0%	0%	0%	1%	0%	0%
After 2 weeks or months	35%	39%	30%	34%	33%	7%
After 3 weeks or months	34%	24%	41%	34%	27%	29%
After 4 weeks or months	16%	15%	17%	16%	27%	43%
After 5 weeks or months	6%	7%	6%	6%	13%	14%
After 6 weeks or months	9%	12%	7%	9%	0%	7%

There are some differences when we compare the peaks in 1999 and 2010 (Table 4), especially for monthlies where the observed peaks are noticeably earlier in the new CIM data. This may be due to:

- the differences in selection of monthlies,
- the bias to more regular readers in 1999 study.
- or real differences in reading behavior over 10 years, i.e. smaller audiences and less frequent reading.

These differences do not change the fact that the peak analysis illustrates the non-continuous growth of observed cover recognition results.

The impact for cover recognition as a methodology to establish AIR is potentially devastating. If we accept these data as reliable, the only way to make them useable is through elaborate statistical modeling.

### 3.3. Audience estimates based on cover recognition versus recent reading.

From a research point of view, size doesn't matter. It seems a silly idea to compare specific issue readership and recent reading. Of course there will be differences : SIR might be lower (monthlies in Joyce, 1985), equal (weeklies in Joyce, 1985) or higher (Petric and Appel, 2007). So what ? You don't compare apples with Brussels sprouts !

Very wise consultants have proposed to filter cover recognition results by recent reading claims, and subsequently establish for each magazine the continuously growing function with the best fit to the observed data. If we were to apply these parameters to existing recent reading AIR, wouldn't we have the best of two worlds? This may be an honorable and pragmatic way to avoid a head on confrontation of SIR and RR. However, the Technical Committee estimated that this approach would be tentative at best, if not violating the observed data. It therefore decided to explore the potential consequences of a full paradigm shift on audience levels and competitive relationships, knowing that size does matter for publishers...

Results were clear. When expressed as a percentage of declared recent readership, on a total of 95 magazines, 82 of the maximum observed accumulations stayed well below the declared recent readership. There was no meaningful difference between weeklies and monthlies.

Is this what could be expected? The literature is inconclusive. At a conceptual level, any hypothesis requires so many assumptions, that we don't even want to start a theoretical argument. But it clearly was not what the market hoped for. That is why we went a step further.

### 3.4. Additional verifications

To better understand what happened, we asked another research institute (Research Solution) to analyze our interviewing procedure in a qualitative study. In the qualitative study, twenty four people were invited to answer the press questions, including cover recognition, exactly as in the currency study. Afterwards, they were interviewed on the procedure and the way they answered different questions. At that moment, they had all paper magazines at their disposal. The most interesting conclusions:

- The cover recognition question in itself is easily understood.
- People who read a lot of magazines sometimes showed attention lapses where they click the same cover positions for two or more magazines on a row, before spontaneously recovering (but without correcting the previous screens).
- The most important memory aid consists of the image and the cover titles, especially for people who acquired the magazine themselves.
- The date of publication is a secondary aid that in case of doubt leads to reconstructive reasoning based on habits or very specific events. E.g. "My mother is a subscriber. I go there every Wednesday and leaf through it. I must have read that issue.", "I don't usually read it. But I was travelling that day and someone on the bus brought it with him. So I did read it then", "My father was in hospital then. I read it there. That's why I remember".
- Only a small minority expressed the need to leaf through the full paper issue. E.g. "If it's a magazine I received from someone else, the cover doesn't help", "Those covers look alike; I recognize some of them from lying around at home but it would help to leaf through them". On the other hand, when given the opportunity afterwards, not a single respondent that went through the full magazine, felt the need to change his answer recorded earlier.
- In case of doubt, almost all respondents report their answer was "no".
- A detailed comparison of recent reading and cover recognition for all 181 magazines read by the 24 respondents, allowed to check the consistency between the two measures:

- In 58% of the cases, responses were consistent (this does not mean necessarily “correct”, it means that recent reading and cover recognition responses were compatible),
- In 20% of the cases, the comparison was inconclusive (reading last months or last period without cover recognition is perfectly possible if it concerns issues older than the last 6),
- In 22% of the cases, responses were inconsistent (no reading in last period, but last issue identified as being read).
- The percentage of consistent responses was the same for weeklies and monthlies (59% and 57%).
- The percentage of inconsistencies was higher for monthlies (29%) than for weeklies (17%). These averages remind the report of Faasse and van Meerem (2003, p.536) of “relatively large numbers of false ‘past reading’ claims during the first measurement week, ranging from 18% to 34% of total issue readership”.
- Inconsistencies concerned in 3 out of 4 cases tertiary readership: the magazine was not acquired by the respondent or his family but passed on by someone else or read somewhere else.
- However, tertiary readership in itself doesn’t seem to be the reason for inconsistencies: it is the combination of irregular reading with tertiary readership that seems fatal for reliable recall or recognition.

While demonstrating certain limits of our questioning method, the qualitative study did not reveal any major argument to disqualify our cover recognition questions or the recent reading/cover recognition sequence.

## 4. Discussion

The joint industry committee in Belgium included cover recognition questions for magazines in the press currency study. New covers were introduced two days after their appearance at the newsstand. This long term test, led to the following observations.

- The logistics went unexpectedly smoothly.
- The number of recognized covers from the last 6 issues correlates strongly with previously declared frequency.
- None of the magazines showed continuously growing accumulation curves based on cover recognition.
- Almost all of the observed accumulation curves stayed below the declared recent reading level.

The irregular nature of the observed accumulation curves seems to conflict with earlier results from the Netherlands (Petric and Appel, 2007; Petric and Appel 2009). They are however consistent with the raw data from earlier Belgian research (Peeters, 1999). Can we make any sense of these differences in results produced by two equally trustworthy Joint Industry Committees?

### 4.1. No accumulation data without modeling ?

Are the differences really problematic? Observed audience accumulation data are always to some degree irregular. Before using them in media planning, accumulation curves have always been re-estimated (Peeters, 1999), smoothed (Petric and Appel, 2009) or otherwise modeled with the continuously growing function that best fit the data (e.g. Baim, 1999). It’s a fact of life: also cover recognition data are subject to biases and forgetting. Therefore, decent modeling of accumulation data is not a sin, it’s a virtue.

Unfortunately, our data seem so far off the mark, that they raise the question whether accumulation modeling doesn’t become indecent at one point?

Peeters (1999) used FIR modeling to ‘overcome’ the limitations of the observed data. While his analysis was very insightful of determinants of accumulation, the method used by this magazines sale house is unfit to estimate accumulation levels within the context of a currency study.

A light weight modeling procedure is of no avail either. In our case, the Dutch smoothing procedure, or for that matter any ‘best fit’ modeling, simply leads to flat curves from the second or third week after publication onwards. Forcing continuously growing functions would necessarily imply new theoretical assumptions which in our opinion significantly reduce the appeal of specific issue cover recognition as a sound basis for average readership.



An imposed marriage of accumulation data, filtered on recent reading, with the recent reading currency, was not accepted in Belgium. It was considered a very pragmatic way out of any cover recognition and recent reading inconsistencies, but at the same time it seems a blatant denial of the significant differences between average issue readership based on RR and on SIR.

#### 4.2. Is sampling size critical?

An obvious criticism on our test is the limited size of our face-to-face sample (n = 10.500, covering 52 weeks) for a bilingual universe (Dutch and French speakers). This sample size proved to be sufficient in the past to produce more or less stable recent reading data. On the other hand, weekly samples of unequal size that are not necessarily representative of the Belgian population, may simply be insufficient for decent cover recognition research.

Baim's (2007) presented a study design that is not affected by this type of flaws: MRI used weekly samples of 2.500 online surveys that were calibrated each week for comparability. Their objective was the analysis of specific issue audiences. Since we are not interested in specific issue audiences but only in the average specific issue audience, we assumed that, over 52 weeks and 10.500 respondents, issue volatility and weekly sample variability are randomized and should not affect the observed averages.

While this assumption is theoretically correct, the noise in the data might still obscure the real accumulation signal. One way to verify, would be to check whether or not the irregularities in accumulation are correlated with the number of observations for a magazine. Unfortunately, defining irregularities supposes we know what's 'regular'. Which we don't, as this was precisely the object of the test...

If sample size really is key, it raises an important issue. Small (and multilingual) markets like Belgium, that do not have very deep pockets, are faced with an almost impossible challenge to marry budget and quality in cover recognition research.

#### 4.3. SIR after RR means trouble?

In both the current and the past Belgian study (Peeters, 1999) recent reading questions preceded cover recognition. The Dutch validation study (Petric and Appel, 2007) did exactly the opposite. We know how sensitive recent reading questioning is to wording, order effects etc. Does the same hold for cover recognition ?

Different hypotheses have been put forward:

- The preceding recency question may act as a primer resulting in more attention to more recent covers.
- Respondents may feel they already have responded to a reading question and consider the cover recognition as an annoying repetition in other words.
- Respondents are simply fed up with these repetitive questions and the easiest way to save time is to limit the effort to identify covers, especially more difficult, older ones.

While there is no conclusive way to validate or invalidate these hypotheses, we do have an opinion on all three of them.

- It seems to us very unlikely that the name of a magazine, as used in a recency question, primes or activates cover images of that magazine that will therefore be more easily recognized a few minutes later. This would be a process that cannot be explained by known perceptual, conceptual, repetition or semantic priming effects as described in cognitive psychology. It would be a very peculiar variant of associative priming between title and an individual cover that dissipates over weeks or months? We certainly hope that our memory is inefficiently organized to that extent...
- Cover recognition may be considered as an annoying repetition of previous audience questions. This is not what we saw in the qualitative study but we cannot exclude this to be a factor in a somewhat more impersonal setting as the currency study. In itself this doesn't explain why the two or three youngest covers perform relatively well compared to the oldest covers. The third hypothesis would have to be invoked.
- Clearly we cannot exclude that respondents are fed up, limit their effort and therefore do not recognize older covers to the extent they should. However, if we assume this to be true in a face-to-face setting where there is still some pressure by the mere presence of an interviewer, how can we expect this not to happen in web based questioning ? It would mean that active effort on the part of a well-motivated respondent is a prerequisite for reliable cover recognition results.

#### 4.4. Is cover recognition superior to recent reading recall ?

It has been suggested that recent reading questioning is as silly as asking when you last sneezed... Cognitive psychology would sufficiently have demonstrated that recognition is the superior method for tapping memory, compared to recall. While this is certainly true in a lot of lab experiments, it is overly simplistic to claim that cover recognition is, almost by definition, superior to recent reading questioning. It's like saying that multiple choice exams are easier than open questions exams. Ask any student or teaching assistant: it depends! It depends on the subject, the type of question and the trickiness of the answer alternatives. Cover recognition is not fundamentally different:

- False positive cover recognitions were identified both in our qualitative study and by the observations of Faasse and van Meerem (2003). Memory retrieval cues offered by magazine covers clearly provoke comparative processes and decisions that are not only based on actual memories but also on other familiarity judgments.
- The observation in our qualitative study that the date of publication serves as a powerful secondary aid in case of doubt, is a clear indication that cover recognition is not always an effortless process. The type of reconstructive reasoning based on habits or very specific events, demonstrates that cover recognition is in fact not always based on cover recognition process at all. In these cases, it is bound to lead to correct as well as incorrect identifications.
- The observation in the qualitative study that lapses in attention make respondents slip into a kind of repetitive mode before they spontaneously recover in an attentive mode, suggests that cover recognition might be at least a fragile on this point as recent reading questioning.

It is very likely that the accuracy of cover recognition is strongly influenced by the combination of the type of cover strategy, source of copy and reader engagement. This type of analyses was beyond the scope of this test. However, we do expect differences in cover recognition between buyers and readers in waiting rooms, between news magazines and women's weeklies, between home decoration monthlies and ICT monthlies etc.

From the point of view of a Joint Industry Committee, this means that a switch from recent reading to cover recognition not only would affect overall reading levels but probably also, and more importantly, the rankings between titles. This is the kind of paradigm shift a market only wants to consider if the superiority of the new results can be demonstrated beyond reasonable doubt. Which our test did not.

### 5. Conclusion

Average Issue Readership based on cover recognition is not the white knight we hoped for.

The most optimistic interpretation of the Belgian data is that SIR only works with very large samples. That shouldn't be a problem for bigger and richer markets, or those that are willing to compromise on sampling quality. It is a major problem for Belgium.

There is a more gloomy interpretation. Specific Issue Readership based on cover recognition may not be beset by telescoping, parallel reading and other recent reading biases, but it suffers from its own biases.

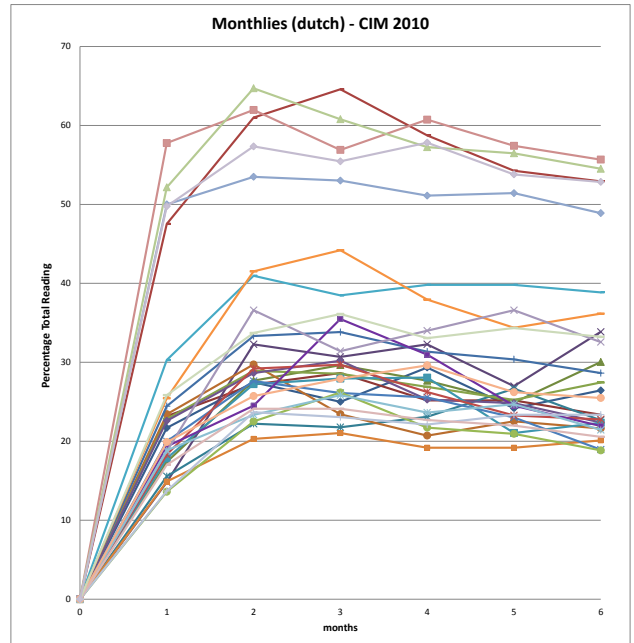
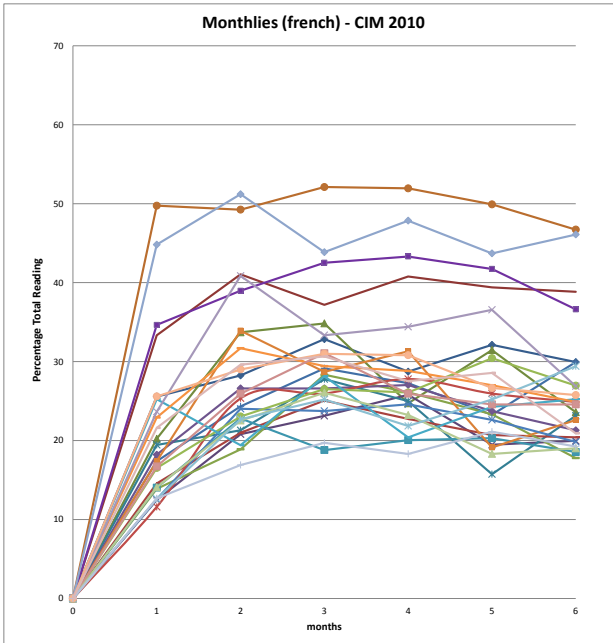
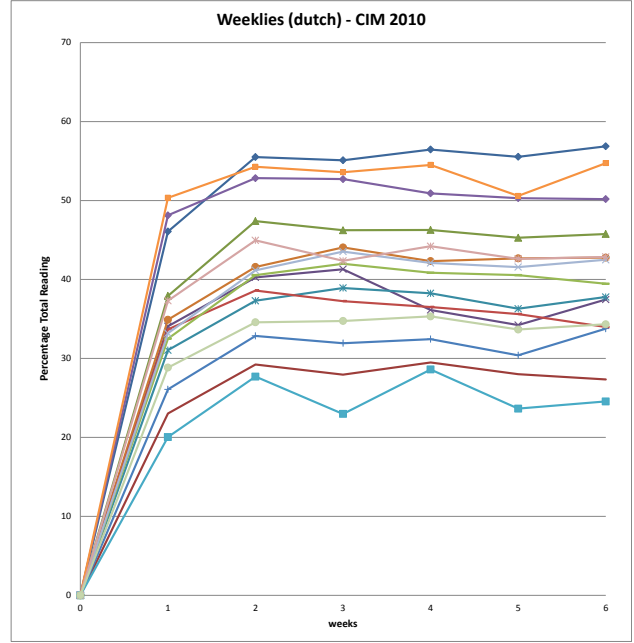
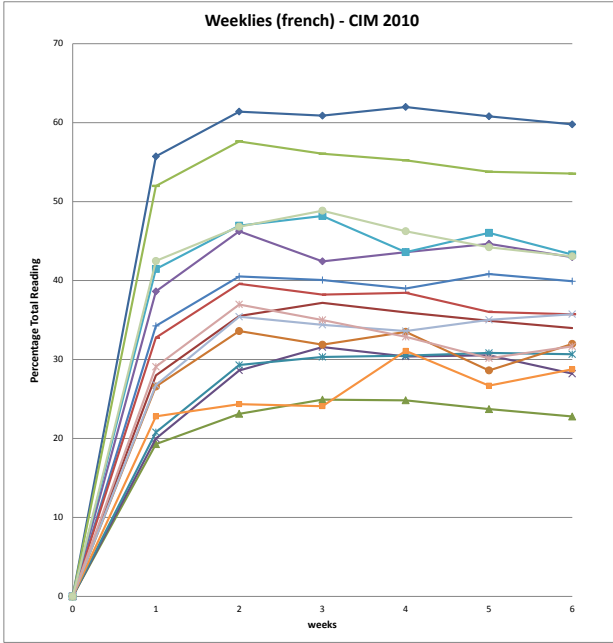
- It does require attention, focus and effort.
- Cover recognition is sensitive to forgetting, especially for monthlies. The interactions of forgetting with cover strategy and reader engagement have only been scratched upon in readership research to date.
- Cover confusion between issues from the same title or between different magazines, for instance, does exist and probably won't affect all magazines to the same degree. At this stage, we do not fully understand its dynamics.

AIR based in cover recognition is almost always lower than AIR based on recent reading. More importantly, differences between the cover recognition based and the recent reading based AIR estimates vary significantly between titles. Rankings based on the new measure could therefore significantly differ from rankings based on recent reading.

Admittedly, our test has also its limitations. We will not claim that a decent audience currency based on cover recognition is not possible. However, in our opinion, cover recognition based average issue readership would not necessarily lead to credible accumulation curves. It would, on the contrary, almost certainly create an earthquake in our market. More fundamentally, we are not prepared to put another 40 years of painstaking research in identifying all determinants of cover recognition. In Belgium, the chapter of a press currency based on cover recognition is closed for the foreseeable future. We will focus our attention now on the challenging measurement of combined paper and digital readership.

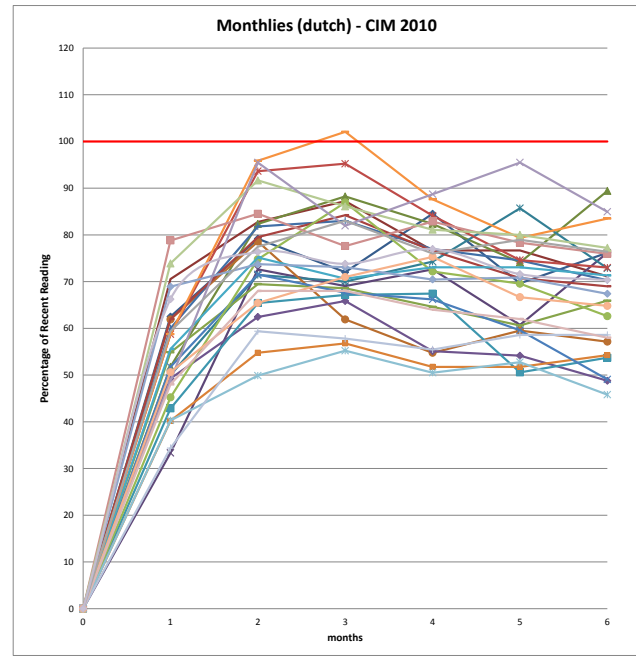
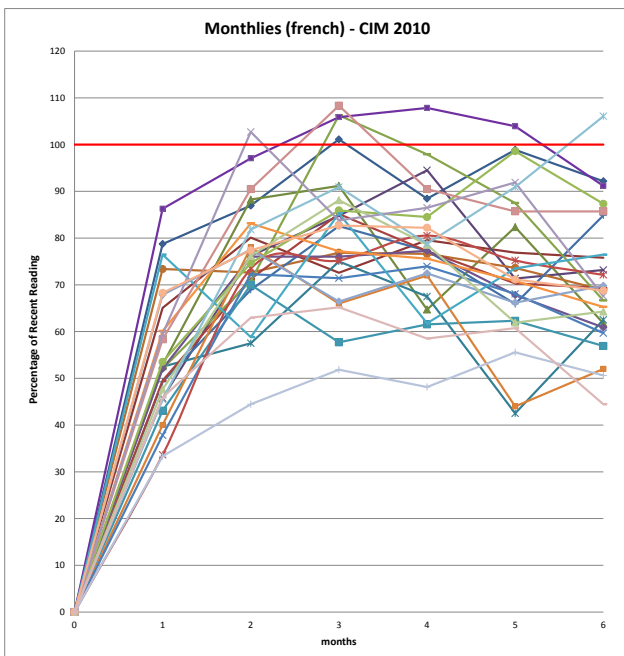
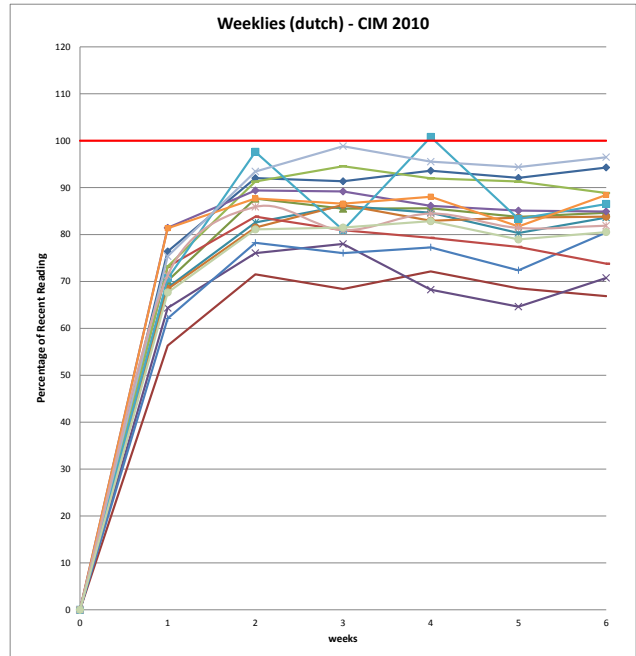
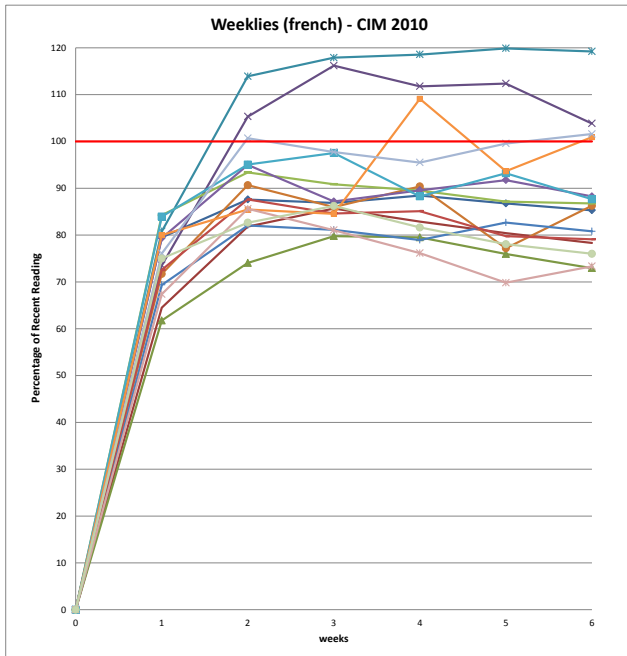
**Attachment 1. Observed accumulation based on cover recognition as a percentage of total readership.**

These graphs are meant to give a general overview, not to allow analyses of individual magazines. The vertical scale goes from 0 to 70 % where 100% = total readership ('Ever read last months?'). The horizontal scale shows the 6 last covers. Data are not modeled.



**Attachment 2. Observed accumulation based on cover recognition as a percentage of recent readership.**

These graphs are meant to give a general overview, not to allow analyses of individual magazines. The vertical scale goes from 0 to 120 % where 100% = recent readership (the currency). The horizontal scale shows the 6 last covers. Data are not modeled.



## References

- Agresti Joseph, Baim Julian and Frankel Martin. “*Magazine audience accumulation: basic modeling issues.*” Worldwide Readership Research Symposium, Florence 1999.
- Baim Julian, Frankel Martin and Agresti Joseph. “*Magazine audience accumulation: development of a measurement system and initial results.*” Worldwide Readership Research Symposium, Florence 1999.
- Baim Julian, Frankel Martin, Agresti Joseph and Becker Risa. “*Magazine audience accumulation: discussion of research issues, modeling and application.*” Worldwide Readership Research Symposium, Venice 2001.
- Baim Julian, Frankel Martin, Galin Michal, Agresti Joseph and Zarnitz Kerry. “*Measuring Issue Specific audiences.*” Worldwide Readership Research Symposium, Vienna 2007.
- Baron Roger, Klein Caryn and Jacobs Lori. “*The Volatility Score: a new measure of the issue-to-issue stability of a magazine’s audience.*” Worldwide Readership Research Symposium, Valencia 2009.
- Faasse John and van Meerem Leendert. “*What the world needs now.*” Worldwide Readership Research Symposium, Cambridge, MA 2003.
- Hermie Patrick, Lanckriet Trui, Couvreur Véronique and Lansloot Koen. “*The WAR Project. A small step up the ARF media research ladder, a giant step forward in magazine planning.*” Worldwide Readership Research Symposium, Prague 2005.
- Joyce Timothy. “*A Comparison of Recent Reading and Full Through-The-Book.*” Worldwide Readership Research Symposium, Salzburg 1985.
- Klein Caryn, Jacobs Lori, Rovitzky Alan, Galin Michal, Baim Julian and Frankel Marty. “*Issue Specific audience: perspectives on application.*” Worldwide Readership Research Symposium, Vienna 2007.
- McDonald Scott, Baim Julian, Frankel Martin, Becker Risa, Galin Michal and Mejica Francis. “*Explaining variability in Issue-Specific magazine audience: an exploration examination.*” Worldwide Readership Research Symposium, Valencia 2009.
- Peeters Stef, Debeer Véronique and Lanckriet Trui. “*Magazines deserve time – The Build up of magazine audience over time.*” Worldwide Readership Research Symposium, Florence 1999.
- Petric Irena and Appel Marion. “*The readership currency: Dutch design.*” Worldwide Readership Research Symposium, Vienna 2007.
- Petric Irena and Appel Marion. “*New learning from Specific Issue Readership: accumulation of readership and calculation of Average Issue Readership.*” Worldwide Readership Research Symposium, Valencia 2009.
- Siegers Danielle, Hermie Patrick and Masson Peter. “*Bringing magazine measurement into the 21st century.*” Worldwide Multi Media Measurement: 2008.
- Van Meerem Leendert. “*Specific Issue Readership measurement, part II.*” Worldwide Readership Research Symposium, Prague 2005.