SEARCHING FOR BEST CASI PRACTICES: EXPERIMENTS FOR A NEW DATA COLLECTION METHOD

Alke Bassler & Costa Tchaoussoglou, SUMMO

Synopsis

This paper provides the follow up to the paper reporting on an experiment with computer assisted self interviewing (CASI) by Bart Soels and Costa Tchaoussoglou in Florence (Soels/Tchaoussoglou, 1999). In the past three years, CASI has been tested as a data collection method for the Dutch National Readership Survey, with positive results.

The majority of respondents is very well able to fill in the questionnaire by themselves, even if they lack prior experience with computers. In the few cases where a respondents is not able to fill in the questionnaire and the interviewer has to take over (CAPI), this does not influence the readership figures. Despite of this absence of interviewer bias – which is a major advantage in comparison to CATI – a contractor's effect (differences in results between two or more research companies) is still be found, which is surprising, since the contractor's effect is usually attributed to differences in the work of interviewers.

CASI can be used in face to face interviews as well as in a computer panel (multi channel distribution). Although there are differences in the results between the two, this is mainly due to differences of the sample composition. If the same kind of respondents are compared, the differences are very small or disappear. A representative sample is therefore as important as ever. When recruiting respondents for the face to face interviews, telephone screening proves to be superiour to a face to face approach.

Besides the absences of interviewer bias, the quality of data collected by CASI is improved by the greater diligence and accuracy of the respondents when filling in the questionnaire themselves and without time pressure. This effect is diminished when the most likely answers are filled in by default (to gain time), so that respondents need only change the answer if it is different. It is advisable to have respondents fill in all the answers.

The results of the tests have shown that CASI is a valuable data collection method, which can be applied in various situations where respondents are behind a (computer-)screen. It is therefore not only an appriopriate method for the present, but will be useful for years to come.

1. Background and objectives

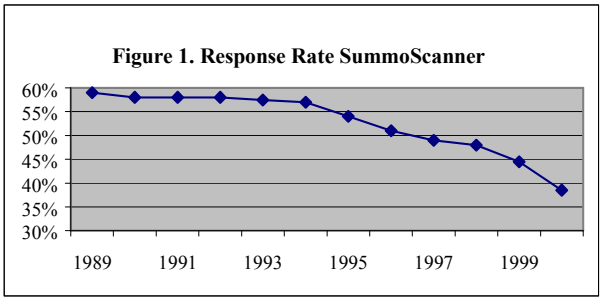
In 1999, Bart Soels and Costa Tchaoussoglou reported on an experiment of SUMMO, which tested the viability of computer assisted face-to-face interviews for the Dutch National Readership Survey (the SummoScanner). Until then – and onward till July 2001 – the SummoScanner has been carried out using CATI. For a number of reasons, a change from telephone interviewing to an alternative data collection method was deemed desirable, if not necessary:

a. *changes in the media landscape*
   In the 16 years of its existence, the SummoScanner has had to cope with a rapidly changing media environment. The development can be captured in one word: more! More magazines, more newspaper sections, more television stations, more media (e.g. the internet). Especially the increase of the number of titles – which were to be measured in the SummoScanner – put a strain on the study, because on the telephone, the time acceptable for respondents is limited. Attempts to limit the number of titles were made impossible by anti trust legislature.

   Also, more magazines cater to the same (niche-)audience. These magazines often not only have the same content, but the lay-out and names are very similar too. This increases the problem of title confusion, a problem which has always been greater for telephone interviewing than for data collection methods where visual cues can be used.

b. changes in society and research

And of course society has changed as well. Although people in the Netherlands are on average only called twice a year by pollsters (Veldkamp, 1998), most people consider it a major nuisance, probably because the public does not differentiate between market research and telemarketing (whether this is understandable and how to remedy this might make an interesting paper on its own). The willingness to participate in research is diminishing, in the SummoScanner as well as in research in general. This means that a major advantage of telephone interviewing – large, representative samples for relatively low costs – is losing ground compared to other data collection methods. That it still has its advantages will be shown later. But even if respondents can be found, who is to interview them? On the telephone, an interviewer is necessary and will keep being so in the near future. Using an interviewer presents a methodological and a logistic problem. Interviewers can cause bias; twice already SUMMO has reported on a drop of readership figures caused by interviewers who "rush" through the interview (Van der Noort/Tchaoussoglou, 1995; Tchaoussoglou/Van der Noort, 1999). The fact that good interviewers are very hard to find adds to the problem – indeed, any kind of interviewer is hard to find with a continuing booming economy in the Netherlands. Even if there would be a recession, the labour force will diminish over the next decades due to smaller generations.



Another USP of CATI – the fact that it is computer assisted – is no longer its prerogative. Face to face CAPI or CASI as well as Webinterviewing offer the same advantages of automatic routing, rotation, etc.

Therefore, SUMMO decided in 1997 to investigate which data collection method would best be fit to reduce the problems of the SummoScanner and prepare the Dutch NRS for the near future.

A series of tests has been carried out:

- a qualitative pilot (carried out in 1999 and reported on by Soels/Tchaoussoglou, 1999)
- a quantitative pilot (carried out in spring 2000)
- a sampling pilot (carried out in autumn 2000)

The qualitative pilot showed that CASI would be a promising data collection method, now and in the future. An important advantage of CASI is that it is not bound to one particular way of recruiting respondents, but can be used in face to face interviews as well as in Webinterviews, in fact in all sorts of situations where people are behind a (computer)screen. Multi channel distribution of the questionnaire is considered as a way to improve the representativity of the sample, because respondents can answer the question how, when and where they feel most comfortable. The questionnaire itself should be the same for all respondents, regardless of the way in which they are recruited.

The aim of the tests following the qualitative pilot - and the subject of this paper - was to determine the best practices for a CASI-NRS. In particular, the objectives were to:

a. investigate the possibility of combining data collection methods
b. test the speeding up of the filter question
c. determine the effect of the changes on the readership figures
d. investigate the phenomenon of 'contractor's effect'
e. determine the best sampling procedure

First, an outline of the design of the quantitative and the sampling pilot will be given. Second, the results concerning each objective are being presented. The paper concludes with some remarks on the future of readership research in the Netherlands.

2. Design

a. Quantitative pilot

The fieldwork was done by two research companies (Interview-NSS and NIPO), each conducting approximately 350 face-to-face interviews. Additionally, a good 300 interviews were conducted using a computer access pool (named Capi@home, by NIPO). An access pool is a database of respondents – in this and most cases with access to an (online) computer – who are willing to participate regularly (once a month) in research. The face to face interviews as well as the access pool interviews were followed by a (paper) self completion questionnaire.

Split run design

Two split runs were executed in the face to face interviews.

CASI vs. CAPI (Interview-NSS)
To eliminate interviewer bias, respondents ideally fill in the questionnaire themselves (CASI). The qualitative pilot has shown that this is not a problem for the majority of the respondents, even if they have never used a computer before. In the NRS it might nevertheless be necessary for a small group of respondents to have the interviewer fill in the questionnaire for them (CAPI). To determine whether this will result in different readership figures, the face to face interviews conducted by Interview-NSS were split in half: in one half, respondents filled out the questionnaire themselves (CASI), in the other half this was done by the interviewer (CAPI).

Speeding up the filter by default coding (NIPO)
The qualitative pilot showed that – compared to asking the ever read question per title – an EML-filter leads to substantial time gain. Especially if titles with low reach are added, it is very efficient. Respondents first identify groups of titles in which at least one title has been read in the past 12 months; subsequently, the frequency question is posed for the titles on the selected EML-screens:

Usually I read … of the issues of …
None - almost none - 1 out of 4 - 2 out of 4 - 3 out of 4 - almost all - all

Because some respondents in the qualitative pilot felt that they had already discarded some of those titles (which they did in their mind or orally, but could not fill in in the EML-question) and because the answer for most titles would be 'none', it was tested whether it would be possible to have the answer 'none' filled in in advance (default), so respondents would only have to change the answer for those titles they actually read. Of course this should not lead to a loss of accuracy.

Including the access pool interviews, schematically the quantitative pilot is pictured below.

Figure 2. Design quantitative pilot

Face to face Interview-NSS	**Face-to-face NIPO**	**Access pool NIPO**
Split 1: CASI (no default) (n=199)	*Split 3:* (CASI) default (n=166)	*Split 5:* Computer access pool (no default) (n=311)
Split 2: CAPI (no default) (n=174)	*Split 4:* (CASI) no default (n=170)	

Sampling
The sampling procedure was a restricted random walk. Interviewers are supplied with start addresses. Around these addresses (left, right and across the street), they are to attempt to realize three interviews. For each interview, they can approach a cluster of five adjecent (private) addresses. If they are refused at the first door or if there is no one at home, they approach the next door and so on. After five doors they must stop. If not successful, they can revisit the same addresses (where there was no refusal of course) twice again. Once an interview takes place in a cluster, or if after three visits there is no complete, that cluster is not used again. Each attempt to contact an address is noted in a contact form.

Interviewers were instructed to try to contact most addresses in the evening and to spread their interviews over all days of the week. Also, they offered incentives to the respondents (10 guilders for the interview, 15 guilders for the self-completion questionnaire following the interview).

b. Sampling pilot
After the test in the quantitative pilot had shown that a face to face approach of respondents did not result in an acceptable sample, a screening by telephone was tested. Ideally, respondents should

1. receive a letter,
2. after which there would be a CATI-interview,
3. at the end of which an appointment would be made with the selected respondent to accomplish the readership survey.

To do this, a match had to be made between addresses and telephone numbers. Two approaches were tested:

a. drawing a sample of addresses and matching them with phone numbers (approach A)
b. drawing a sample of telephone numbers and matching them with addresses (approach T).

Obviously, the match could not be made for all sample units, e.g. unlisted telephone numbers could not be found. Therefore, for each approach there were two conditions: 1. match succesful (i.e. the ideal procedure could be followed) and 2. no match (i.e. a different procedure had to be followed).

This resulted in four conditions:

Figure 3. Design sampling pilot

A1 - address known - successful match with telephone number/ name - personalized introductory letter - telephone screening	**T1** - telephone number known - successful match with address/name - personalized introductory letter - telephone screening
A2 - address known - no match with telephone number/name - unpersonalized introductory letter - face to face approach	**T2** - telephone number known - no match with address/name - no letter - telephone screening/'cold calling'

Respondents were approached for the screening and, if possible, for the readership survey. The approaches A and T were evaluated concerning the response rate and especially concerning the composition of the achieved samples.

3. Combining data collection methods

In the pilot (and in the subsequent Dutch NRS) two combinations of data collection methods have to be made:

a. CAPI and CASI
b. Face to face and access pool

In the quantitative pilot, for both combination it was analysed whether the difference in data collection method resulted in different readership figures. Because of the limited sample sizes, readership figures (number of filter screens selected, read in past 12 months and AIR) were not compared per title, but for all titles, all magazines and all newspapers.

a. CASI vs. CAPI

Even though there are differences, these differences are not significant. The greatest differences occur with newsweeklies, which have a higher reach (though the difference is not significant) when the interviewer is asking the questions. Obviously, socially desirable answering is causing this. This effect is a further affirmation for our preference not to use interviewers. Anyway, it is not so great as to pose a threat to a design, where CAPI will be used if the respondent is not able to fill in the questionnaire him/herself.

Table 1. Comparison readership figures CASI/CAPI

	CASI	CAPI
All titles		
Read in past 12 months	1231%	1252%
AIR	513%	505%
All magazines		
Read in past 12 months	898%	902%
AIR	350%	339%
All newspapers		
Read in past 12 months	218%	234%
AIR	79%	90%

b. Face to face vs. access pool

Readership figures in the face to face interviews (split 4) and the access pool (split 5) differ significantly for most title groups (only the interviews from the same research company and with no default coding are compared). This is mainly due to differences in the two samples: there are more male respondents in the access pool, they are more often married and on average 4 years younger than their face to face counterparts. More people have a job and of course computer ownership and computer skills are higher.

Table 2. Comparison face to face/access pool

	Face to face		Access pool	
	n		n	
Selected filter screens	170	10,85	311	13,05
Read in past 12 months				
all titles	170	1700%	311	2166%
all magazines	170	1289%	311	1663%
all newspapers	170	267%	311	350%

However, if the same kind of people are selected in each sample – in this case people who own a computer and have a job – there is no significant difference in the number of filter screens selected (the first step of determining readership). Also, there is no significant difference in reading of magazines in the past 12 months. There is still a significant difference for reading of newspapers in the past 12 months, but this is caused by the higher percentage of men in the access pool sample. If gender is kept constant, the differences become smaller, in fact, they almost completely disappear for women (table 3).

This means that respondents fill in the questionnaire in the same way, regardless of its way of distribution. The data from the different collection methods can therefore be combined without bias. This finding is in line with a study presented during the Symposium in Berlin, where Candon and Fagot (1995) concluded that face to face interviews could well be combined with self completion interviews (although small but significant differences where found in their study).

Table 3. Comparison face to face/access pool among computer owners with a job

		Face to face		Access pool	
		n		n	
Male	**Selected filter screens**	27	12,00	106	13,41
	Read in past 12 months				
	all titles	27	1885%	106	2253%
	all magazines	27	1437%	106	1660%
	all newspapers	27	348%	106	432%
Female	**Selected filter screens**	40	13,35	70	13,69
	Read in past 12 months				
	all titles	40	2245%	70	2387%
	all magazines	40	1835%	70	1891%
	all newspapers	40	290%	70	341%

Obviously, the total sample of face to face interviews and access pool interviews has to be carefully weighted to avoid bias due to unrepresentativity in the overall sample. Also, the sample from the access pool has to be stratified, so the weighting need not be have to be too great. Over time, it is expected that the access pool will have a more representative composition, so that samples will be more like the face to face sample.
If precautions regarding the sample are taken, CASI is very well suited for multi channel distribution of the questionnaire.

4. How to speed things up

To determine whether the time required to fill in the questionnaire could be further reduced, the frequency question was adapted. The frequency question is posed, if a respondent selects an EML-filter screen (i.e. the respondent has read at least 1 of the magazines on the screen in the past 12 months, see figure 4).

The respondent then has to fill in how many issues s/he usually reads (see figure 5).

Figure 4.
Filter question (EML): Have you read or looked into one or more of these titles (at least once) in the past 12 months?

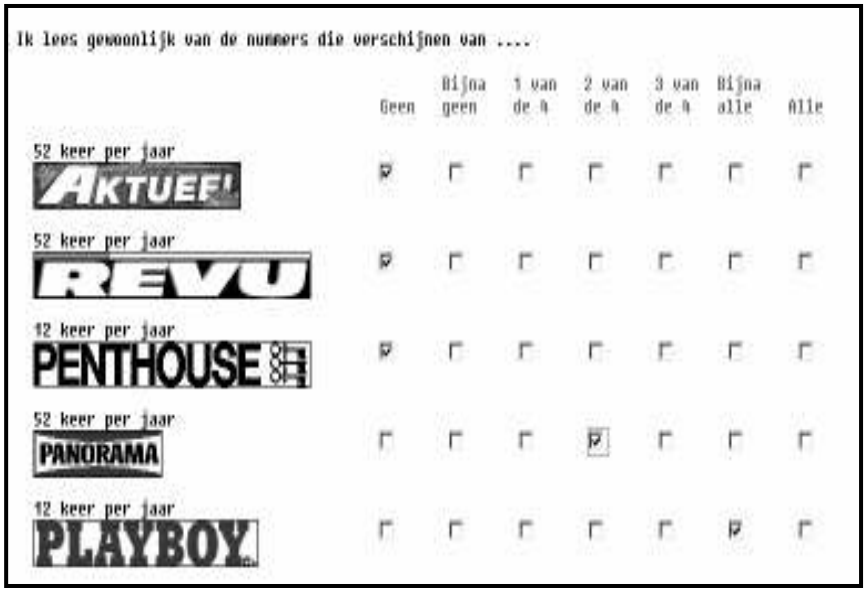
There are 3 to 5 titles on the filter screen and the same 3 to 5 titles appear in the frequency question. On average, a respondent will have read only 1 or 2 of those titles, in other words, for most titles the answer to the frequency question is 'none'. If this answer is already filled in by default and needs only be changed in those cases that a title has been read, this should lead to time gain. Of course, this poses the threat that respondents are less accurate and overlook titles.

To test whether time gain can be accomplished and whether this has an effect on readership figures, a split-run was set-up: of the face to face interviews conducted by NIPO, half had the answer 'none' filled in the frequency question by default (split 3), while half had not (split 4).

A comparison of the two splits shows that there is an average time gain of 3 minutes (total interview time: 34 minutes). But also, the default answer leads to a drop of the readership figures: even though respondents in both splits select an equal number of filter screens (10), the group with the default answer has a significant lower reach for most magazines than the group without the default answer. In other words, fewer titles pass the frequency filter. Newspapers are not effected.

The possible explanation is that having to fill in an answer per title urges respondents to think more diligently. In the qualitative pilot it was seen that respondents took their time and tried hard to remember the correct answer. It is plausible that the extra effort leads to a more valid estimation of the frequency. It is therefore not advisable to use default answers in questions crucial to the determination of readership.

Figure 5. Frequency question:
Usually I read … of the issues of …
None – almost none – 1 out of 4 – 2 out of 4 – 3 out of 4 – almost all - all



5. Readership figures and contractor's effect

What will happen to the readership figures if the data collection method is no longer CATI, but CASI? Appendix A gives a first impression, 'read in the past 12 months' from CASI is shown as an index of the CATI readership figures (only face to face interviews). There will be a drop of reach for most sorts of titles, with the exception of weekly magazines, which profit from the change of method. The decrease is due to a change of the filter question: while in the SummoScanner the 'ever read' question is used, the new filter has a time frame: 'read in the past 12 months' and subsequently: 'how many issues do you usually read?'. Whereas the old filter is very wide, the new one is based on habitual behaviour (even though the answer can be 'almost none'), within a certain period of time. The decrease is not uniform for all sorts of titles: especially those titles lose readership, which in the telephone interview might profit from title confusion, such as sportsmagazines, car magazines or home and garden magazines. This supports the idea that title confusion is reduced by visual cues.

Most striking is the difference between the results of the two research companies. Even though the design of the studies is identical, Interview-NSS reports much lower readership figures than NIPO. This phenomenon is known as the contractor's effect. Readership figures in the NIPO-sample are 30% higher than those of the Interview-NSS-sample. To determine whether this is caused by sample differences, a regression analyses was done using the research company, sex, age, work and computer ownership as explanatory variables. The difference – although less – is still mainly attributed to the research company: from 2.6 filter screens which were originally selected more by NIPO-respondents, there are still 1.8 left. There is no significant difference for newspapers.

Table 4. Differences research companies

	'company' as explanation	'company' and 'socios' as explanation
Number of filter screens	2.59	1.88
Number of magazines read	3.89	2.87
Number of newspapers read	0.388*	0.098*

n = 916, * not significant

Usually, the contractor's effect is ascribed to differences between interviewers of different companies. As this study uses CASI, this cannot be the (main) explanation. The only possible explanation might be the interview software; there were minor differences between the two companies (e.g. to proceed to the next question, the respondent had to click on a button, which was placed in the left upper corner in one case and in the right lower case in the other). The advise for future research is therefore that even minute details be done identically, if more than one company is contracted to do the research.

6. Response

In the quantitative pilot, a face to face approach was used (cluster of 5 addresses, which could be visited up to three times). The results proved unsatisfactory. Too many women were interviewed, also, young people were under represented and there was no good representation of region, work and computer ownership. In short, we interviewed the old and the housewives – the stay-at-homers. This was also shown by the response:

Table 5. Response face to face interviews

	Per cluster		Per address	
	n	%	n	%
Complete	709	67	709	22
Refusal	342	33	1684	51
Not at home			875	27
Total	1051	100	3268	100

The refusal rate, though it may seem high, was no surprise, as it is in line with experience from other research projects. The problem is caused by the substitution of people who are not at home and who are replaced by people who are at home – obviously, not the same kind of people. It was therefore decided to replace the face to face approach with a telephone screening. This was tested in the sampling pilot.

Two approaches were compared:

b) address as sampling unit
c) telephone number as sampling unit

Where possible, these were matched with name and telephone number or address respectively, so a personalized introductory letter could be sent before contacting the household. Because a match could not be made in all cases, the approaches had to be split:

A: A1 – address, name and telephone number known
A2 – only address known (respondents were approached at the door)
T: T1 – address, name and telephone number known
T2 – only telephone number known (cold calling)

The telephone screening could be held in A1, T1 and T2. The results of A2 (the face to face approach) are questionnable, probably interviewers did not keep track of all attempts to contact an address, therefore, no totals for A is given.

Table 6. Response

	A1	A2	A total	T1	T2	T total
Response screening	49%	-	49%	53%	36%	48%
Response face to face interview	21%	34%?	?	17%	11%	16%

There is little difference overall between A and T. Both have a response rate of approximately 50% for the screening and around 20% for the subsequent face to face interview. Cold calling (T2) results in a lower response rate, however, since it is only necessary for a limited number of sampling units, it does not reduce the overall response rate too much.

The composition of the samples is much more in line with the general population than it was in the sample for the quantitative pilot. No approach did markedly better than the other, however, the address-sample has a theoretical advantage over the telephone number sample. A Finnish study (Kuusela & Vikki, 1999) showed, that it is getting increasingly harder to have a proper sampling frame for telephone numbers. Mobile phones and ISDN with multiple numbers per household pose a problem, since it is not clear whether telephone numbers from all providers can be obtained and if so, an individual may be included several times through household number, mobile phone, while others are included only once. The use of mobile phones in particular seems to be related to certain phases in life; under- or over representation is therefore not random. In the Netherlands, this problem is expected only to increase in the near future. Therefore, the address sample, which is kept up to date and will be kept up to date as long as there still is conventional mail, is better fit for the future.

7. In conclusion

The pilots have supplied us with more insight into the do's and don'ts of a CASI readership survey:

- The use of visual cues reduces title confusion.
- Having the respondent fill in the questionnaire, improves accuracy, if no short-cuts like default coding of answers is used.
- If it should be necessary that occasionally the questionnaire is filled in by the interviewers, this will not bias the readership figures for most magazines.
- Distributing the questionnaire by various channels is also very well possible.
- The observed contractor's effect is a reminder to pay attention to even minute details, if the readership survey is conducted by more than one research company and of course also, if there is only one contractor.
- Even though the telephone is no longer acceptable as a data collection method, it is still superior to a face to face approach for finding and recruiting respondents.

After all these valuable lessons, we imagine readers want to know how readership research will look like in the Netherlands in the near future. While finishing this paper, we neither know when the new readership survey will start nor whether it will indeed make use of CASI. The reason is that a new JIC is being founded and no consensus has been reached yet whether daily newspapers and magazines are to be measured single source or separately. It may also be possible that the decision-making process will rather be slowed down than fastened by the findings in our experiments: *the more you know, the less you do*. This was exactly the reason for our Belgian colleagues, who made several major changes in their national readership survey without testing anything: "…, we also knew that if Belgian operators were confronted with figures resulting from tests, some of them would get scared for their own performances and that the risk of political blockage would be huge." (Haute, 1997).
Our best guess is, that – as usual - decisions will not be made, they will just happen. Nevertheless we are convinced that CASI will be the major data collection method in the future.

References

Candon, J.L. & Fagot, C. (1995). "Personal In-home Interviews Versus Self-administered Mail Panel: Does It Really Matter for Measuring Print Audiences?". Worldwide Readership Research SymposiumVII. pp. 53-72.

Haute, F. van den (1997). "Changing Almost Everything in the New CIM Press Survey … And Observing 'Another Belgium'". Worldwide Readership Research Symposium 8. pp. 23-33.

Kuusela, V. & Vikki, K. (1999). "Change of Telephone Coverage Due to Mobile Phones". www.jpsm.umd.edu/icsn/papers/kuuselavikki.htm.

Noort, W. van der & Tchaoussoglou, C. (1995). "The Importance of Being Constant. The Effects of Questionnaire Overload." Worldwide Readership Research Symposium VII. pp. 111-116.

Soels, B. & Tchaoussoglou, C. (1999). "What's New Pussycat? CATI, CAPI and CASI: Or What Else?" Worldwide Readership Research Symposium 9. pp. 231-237.

Tchaoussoglou, C. & Noort, W. van der (1999). "Divide and Unite – Splitting the SummoScanner and Data Fusion". Worldwide Readership Research Symposium 9 Session Papers. pp. 231-237.

Veldkamp (1998). Bellen & Gebeld Worden. [Calling & Being Called]. Study commissioned by DMSA.

Appendix A: Indices of 'Read in past 12 months' CASI/CATI

(SummoScanner 1999-2nd quarter = 100; face to face interviews without default)

	1	2	3	4	5	6
	Interview-NSS			Nipo		
	total	male	female	total	male	female
All titles	78	72	82	106	102	108
All newspapers	79	77	81	92	92	93
National newspapers	82	79	85	102	103	101
National morningpapers	82	83	79	105	104	106
National eveningpapers	87	66	120	87	110	75
Regional newspapers	75	73	76	80	77	83
All magazines	78	71	82	109	105	111
Magazines ex. Broadcasting magazines	77	70	82	110	107	112
Broadcasting magazines	79	78	80	101	89	109
Newsweeklies	64	49	85	95	100	88
Weeklies	107	117	103	138	141	136
Motorclub magazines	56	56	55	74	87	61
Car magazines	75	79	61	123	131	94
Youth/music magazines	59	46	71	112	79	136
Football/sports magazines	65	65	63	84	79	90
Home and garden magazines	62	43	70	85	78	88
Monthlies	94	89	95	120	96	131
Parenting/health magazines	74	37	89	125	133	120
Computer magazines	94	92	92	124	113	158
Other	52	43	57	74	67	77