MEASURING AND MINIMIZING POSITIONAL AND TITLE CONFUSION BIASES IN PRINT MEDIA AUDIENCE ESTIMATES

Marco Vriens, Barbara Graham, Eric Melton, Tony Incalcatera,
Millward Brown IntelliQuest
Valentine Appel, Consultant

Synopsis

In measuring print media audience sizes using self-administered paper questionnaires two types of biases have been documented or suggested in the literature: positional biases and title confusion bias. This paper contributes to the literature by simultaneously addressing two types of positional biases and title confusion bias and assessing their impact upon screen-in rates and weighted reading frequencies. We also show how these biases can be minimized through the use of a web-based questionnaire design.

Potential Biases in Print Audience Measurement

Self-administered surveys are commonly used in assessing the size of media audiences. In the context of measuring the size of print media audiences, studies have shown or suggested that the position of a title in a survey can affect its screen-in rate and average issue audience estimate. Two reasons for this occurrence have been isolated or hypothesized in past studies. One, when magazines are presented in a fixed order (there is no versioning of the questionnaire) it is possible that the position at which a magazine is presented affects the responses to the screen-in and the frequency of reading questions, as all respondents see the same list of publications in the same order. We call this the **pure positional order bias**. If such a bias exists we can minimize it by randomizing the presentation order of the magazine titles (i.e. the order in which the magazines appear in the survey is randomized across respondents). Secondly, it has been documented that another positional bias occurs when the screen-in question is directly followed by more specific reading-habit questions. For example, a respondent can first be asked whether he/she reads a particular magazine title and then, when the answer is yes, secondly is immediately asked to respond to a reading-frequency question. Respondents quickly discover that they can influence the length of their survey by just saying that they didn't read a particular title realizing that by doing so they will not have to answer the reading-frequency question. This type of bias may be exacerbated when other more detailed questions about their reading habits for the magazines are asked at this point too (i.e. percentage of pages read, ad attractiveness). Once respondents realize the linkage between the screen-in questions and follow-up questions they may consciously decide to alter their response behavior. For example, they may start to 'skip' to answer "yes" to the screen in questions for magazines that they actually did read. The tendency of respondents to do this at some point in the media roster may become stronger as the survey proceeds or as the print media list is perceived to be longer than expected and/or considered reasonable. Alternatively, respondents may just start to randomly 'skip' questions, focusing only on the magazines they read more frequently, or they may just adopt another "short-cut" strategy. We call this the **positional screen-in punishment bias**. The result can be that audiences for titles presented later in the print media roster may be under-estimated. By separating the screen-in and reading-frequency questions and by randomizing the order in which the titles are shown we are able to minimize all aforementioned positional biases. In this study we address the potential biasing effect of a title's position.

Even if the screen-in and reading-frequency questions were separated from each other and the magazine presentation order is randomized an additional bias can occur. Studies have shown that when magazines with similar titles (e.g. PC Magazine and PC World) appear in different places within the media section of the questionnaire, each of the similar titles will generate higher screen-in rates and reading-frequencies as compared to situations where such similar titles are grouped together within the media section of the questionnaire. Respondents can confuse a title that is shown with another similar title that is not (yet) shown. If they did not read the shown title but read the similar not shown title, yet they indicate that they read the first shown title of the similar pair we have what is called **title confusion bias.** We can reduce title confusion bias by grouping similar titles because respondents are "exposed" to both titles in conjunction with each other and hence are less likely to confuse them. We note, though, that situations are conceivable in which audience estimates will be actually higher when similar titles are shown together. This could happen, for example, when a respondent reads both of the magazines with similar titles but one of two magazines is only infrequently read. In such a situation, showing both titles together may actually support the respondents' recall of having read the lesser-read magazine. Either way, we can help respondents give more accurate accounts of their real readership by grouping magazines with similar titles together within the print media roster.

Previous Research on Positional Biases

Order effects or positional biases have been identified in many different survey research contexts (e.g. Day, 1969; Chrzan, 1994). The context of printed media audience measurement is no exception (Whipple and McManamon, 1992). Whipple and McManamon (1992) investigated reported magazine receipt and reported magazine readership of four trade publications. One of four trade publications was compared to one of the other three competing magazines. Each magazine in a pair could be either shown first or second. The authors found a 'primacy order' effect for both reported magazine receipt and reported magazine readership: there was a consistent bias favoring the magazine in the first position. Whitley (1981) found order effects in print media studies. In the British National Readership Survey format publications are put into publication categories: i.e. Daily newspapers, Sunday newspapers, Weekly magazines, and Monthly magazines. The order in which these categories were shown to respondents was rotated. Additionally, within a category, magazine titles could be either shown in forward order or in reverse order. Whitley (1981) found that for monthlies an 18% above average readership was found when presented first in the survey. When this category was shown in the last position in the survey this group of magazines obtained an average claimed readership that was 17% less than the average claimed readership percentage. Within each of the groups that were rotated, publications could be either presented in forward order or in reverse order. For a group of women's magazines, with a monthly publication frequency, he found that for 14 out of the 28 titles in this group an average issue readership index of 140 was obtained. When this sub-set of magazines was presented in reverse order (i.e. they were presented as the last 14 in this category) they had an average issue readership of 104. In the British National Readership Study for each publication in the survey the respondent is first asked a question regarding reading-frequency. Within each category the reading-frequency questions for all the titles in a category are answered first. Note that in the British survey this reading-frequency questions serves as the screen-in question). Depending on the answer to this reading-frequency question a follow-up question is asked regarding recency of reading. Hence, because the "screen-in" and "recency" questions were separated within a category, the results across forward presentation order and reverse presentation order can be interpreted as a pure positional bias when the author looked at the results of categories placed in first rotation order. However, we note that the author did not, or could not, do this analysis. Whitley demonstrated the positional screen-in punishment bias, but not the pure positional bias. Cornish and Meier (1983) show that by separating the screen-in and follow-up questions this positional screen-in punishment bias could be reduced in the British study.

Pinnel and Appel (1995) discussed two experiments where print media audience estimates obtained under different questionnaire formats were compared. In a first experiment they compared the mean screen-in rates and mean ratings of two questionnaire versions. In one version, a paper-and-pencil questionnaire, each screen-in question was directly followed by three additional follow-up questions. In a second version, a disk-by-mail version, respondents were first asked the screen-in question for all magazines and hence had no knowledge about the consequences of saying "yes" to the screen-in questions. Both versions included 66 publications. The disk-by-mail version led to an average screen-in rate of 19.5% as opposed to an average of 13.1% for the paper-and-pencil version. The mean ratings were 9.0 and 5.7 respectively. Of course this result could partially have been the effect of having two different interviewing formats. In a second experiment they compared the screen-in rates of two paper-and-pencil questionnaires: in one version the screen-in version was followed by three follow-up questions (about reading-frequency, about how closely they read or examined specific advertisements, and about the percentage of the pages they usually look at or read), and a second version in which only the screen-in question is asked. As expected, the latter version, not burdened by a 'penalty' of follow-up questions, resulted in higher screen-in rates. The latter version showed a mean screen-in rate of 13.2%, whereas the 'penalty' version resulted in a mean screen-in rate of 9.6% (a statistically significant difference). The design of this study did not allow the authors to compare audience estimates across the two paper-and-pencil survey versions that differed by whether or not they had the screen-in question immediately being followed by additional reading habit questions.

Appel, Elder and Incalcatera (1999) compared printed media audience estimates under four presentation formats. One format was the traditional paper-and-pencil method, and the three other were web-based questionnaire formats. In one of the web-based versions the screen-in questions were separated from the frequency of reading questions. Each version included 94 publication titles. As expected this version lead to higher audience estimates because the positional screen-in punishment bias did not occur. The authors found an average screen-in rate of 12.2% for the condition where the positional screen-in bias was removed and an average screen-in rate of 10.0% for the condition where the positional screen-in bias was present (this difference was not statistically significant). However, in the web-based version used by the authors seven magazines per screen were presented, and the respondents were asked to indicate which of these magazines they had read in the past six months. It should be noted that they allowed respondents to indicate, with one stroke, that they had not read any of the seven magazines (a "none of the above" appeared on each screen). Placing a "a none of the above" on each screen may have provided respondents with a way to quickly to go over the survey and create a short cut to do the questionnaire. We hypothesize that if respondents had not been given this "escape" possibility that the audience estimates would have been even higher. The study did not address the pure positional bias because in all tested questionnaire versions one title ordering was used. Hence we cannot tell from this study how big the pure positional bias can be.

Previous Research on Title Confusion Bias

McGlathery (1993) analyses the 1980 ARF Comparability Study Data. In this study nine pairs of magazines were included that were either similar in name or in content. McGlathery found that whenever a pair of similar titles was presented together the screen-in level was lower than when these titles appeared separately. The author also cites an experiment, originally published in Folio by Couzens (1988) that showed that title confusion is a real phenomenon. The Couzens' experiment was conducted for Metropolitan Home. Part of the overall sample was shown a deck of logos that included *Metropolitan Home*, but also included a fictitious title *Metropolitan Home & Garden* and the survey format used was similar to the MRI format. The results showed that of those who claimed readership of the genuine Metropolitan Home, three out of four said they also claimed to have read the fake title.

Healy (1997) addressed the issue of "Should titles with a meaningful degree of similarity be grouped together even though this violates 'issue frequency groupings' to minimize over-estimation of the audience due to title confusion. The two experiments presented by this author showed that the audience estimates for American Health and Health were significantly higher when the logos of these magazines were presented alone rather then presented grouped. The explanation given for this was that for respondents who were not sure which of the health magazines they had read would check off having read only one magazine when the four health magazines were shown together. Alternatively, when health magazine titles were shown in separate parts of the survey this level of being conscious about exactly which health magazine they had read would not occur and hence respondents were more likely to check off both magazine titles as having read them.

In sum, the positional screen-in punishment bias and the title confusion bias have been clearly demonstrated. The pure positional order bias also has been demonstrated in a printed media context. However, the context was one of two magazines, instead of the dozens of magazines that one usually encounters in printed media audience research. The purpose of the current paper is threefold.

1. First, we want to single out each of the three discussed biases and assess their size in a single web-based environment using a controlled experimental design. It is has been found that average screen-in rates and mean ratings are higher in web-based environment as opposed to paper-and-pencil methods (See Appel, Elder, and Incalcatera, 1999) and hence the size of the various biases may be less dramatic between the cells in our experimental design.
2. Second, we also want to assess the size of each bias in situations where more than one bias may occur, so that we can assess their relative importance.
3. Third, we want to test and demonstrate a possible web-based implementation that minimizes the aforementioned biases. We refer to this latter situation as the minimum bias condition.

Based on previous research we have a number of expectations:

1. We expect print media survey formats where a screen-in question is directly followed by additional reading habit questions to result in lower average screen-in rates and number of reads (for definition see below) as compared to formats where screen-in and follow-up questions have been separated from each other (**positional screen-in punishment bias**).
2. We expect print media survey formats that apply a fixed magazine title ordering to result in lower screen-in rates for those sub-groups of magazine titles that were presented more toward the end of the media part of the survey as opposed to formats where magazine titles have been randomized across respondents (**pure positional bias**).
3. We expect that a questionnaire format that positions similar titles in separate places in a media roster to lead to higher average screen-in rates estimates (**title confusion bias**) or lower average screen-in rates estimates (**reduced memory links effect**) as opposed to formats where similar magazine titles are grouped together.
4. The joint effect of both positional biases and title confusion/memory support is difficult to predict because the two types of biases may have opposite effects.

Experimental Design

Self-administered paper and pencil surveys and in-person interviews do not allow us to measure and reduce all discussed types of biases simultaneously. However, online study executed research has the capabilities to allow us to develop a survey design that deals with these biases simultaneously. The design of our proposed research experiment gives us the opportunity not only to minimize these biases but also to measure the size of these biases.

We designed five different experimental conditions and executed on five matched samples, each consisting of about 600 respondents. Each matched sample was given a different version of a print media questionnaire derived from Millward Brown IntelliQuest's Business Computer Industry Media Questionnaire (CIMS). See below for details on the questionnaire design. We included 50 magazines titles from version 8.0 of CIMS in the web-questionnaire. The experimental conditions are summarized in Table 1 below.

Table 1.
Overview of the Experimental Conditions

	Type of Bias (Present=yes, eliminated=no)			
	Title Confusion Bias	Positional Screen-In Punishment Bias	Pure Positional Order Bias-only	Verbal Description of the conditions
Version 1	Yes	Yes	Yes	Maximum bias
Version 2	No	Yes	Yes	Both positional biases
Version 3	No	No	Yes	Pure positional order bias
Version 4	Yes	No	No	Title confusion bias
Version 5	No	No	No	Minimum bias

The 50 magazines can be categorized in different groups based on their similarity in title and/or content. We have 10 groups: A to J. In the first experimental condition we allow all identified types of biases explicitly. This will be referred to as our maximum bias cell. In this condition our survey instrument has the following features:

- No sets of similar titles are grouped together. This means that magazines in the specific groups, A-J, are not kept together in the presentation ordering.
- The screen-in question is directly followed by a reading-frequency question, and other follow-up questions (i.e. How closely advertising for computer hardware, software, Internet, and communications related products and services in each of claimed magazines were read, and what percentage of the pages usually looks at or reads), and
- No specific action is taken to randomize the order of title presentation in the questionnaire (i.e. a fixed title presentation order).

In the second experimental condition we allow for the pure positional order bias and the positional screen-in punishment bias. Title confusion bias is eliminated. This will be referred to as our "both positional biases" cell. In this condition our survey instrument has the following features:

- Similar titles are grouped together. All magazines within group A were presented together, all magazines within group B were presented together, etc. The groups were presented in alphabetical ordering: A, B, C, etc.
- The screen-in question is directly followed by the reading-frequency question and other follow-up questions, and
- A fixed title presentation order is applied for all respondents.

In the third experimental condition we allow only for pure positional bias. We eliminated positional screen-in punishment bias and we eliminated title confusion bias. This will be referred to as our pure positional order bias cell. In this condition our survey instrument has the following features:

- Similar titles are grouped together in the same way this was done for the second experimental condition.
- Screen-in and frequency question (and other follow-up questions) are separated, and
- A fixed title presentation order is applied for all respondents.

The fourth experimental condition only allows for title confusion bias. We have eliminated both the pure positional bias and the positional screen-in punishment bias. This will be referred to as our title confusion bias cell. In this condition our survey instrument has the following features:

- No sets of similar titles are grouped together.
- The screen-in question are separated from the reading-frequency question (and the other follow-up questions), and
- The order of title presentation in the questionnaire is completely randomized for each respondent.

The fifth experimental condition we have eliminated all biases simultaneously. This will be referred to as our minimum bias cell. In this condition our survey instrument has the following features:

- Similar titles are grouped together in a similar fashion as in experimental conditions 2 and 3. The ordering of groups was randomized though in this condition.
- The screen-in question is separated from the reading-frequency question (and the other qualitative readership questions), and
- Within the constraint that similar titles are kept together, all titles or groups of titles are presented in random order.

Web-Questionnaire Design

The five experimental conditions were translated in to five web-based questionnaires using SPSS Quancept Web program. The CIMS Version 8.0 business questionnaire was used as a template. The 50 titles were selected from the initial 91 titles measured in V8.0, eliminating titles with smaller circulation and including as many as possible with similar titles. One magazine title was included that is not in the CIMS V8.0 version: The New York Times Sunday Edition. In the appendices we show the groups of similar magazines in the grouped questionnaire version, and we show an illustration how the magazine titles may be presented when the magazine titles are ungrouped. We also show two representational screens of versions that differ with respect to whether they have separated screen-in question and follow-up questions or screen-in questions immediately followed by reading-frequency questions. In the original CIMS questionnaire, the media reading questions are based on paper-and-pencil where each page of the questionnaire contains 23 publications. In our experiment respondents were shown seven publications per screen. In order to move onto the next screen respondents were required to respond to every question on the screen. Additionally, in our experiment the respondents were asked to respond to the media reading questions and some follow-up questions detailing what the scope of their influence in purchasing certain technologies for the purpose of weighting and projecting the data to CIMS version 8.0 business universe estimates.

Sample

A sample of 7,500 was drawn from IntelliQuest's Technology Panel. Millward Brown IntelliQuest's Technology Panel is a panel of about 27,000 Technology Purchase Influencers. The Panel has been constructed from a wide variety of sources. The drawn sample was not a pure random sample from the panel. It was drawn in 11 segments that closely approximate the composition of the CIMS supplemental Dun & Bradstreet Office Intensive business sample based on work site size and job function. Each respondent was randomly allocated to one of the five experimental conditions. Hence, for each experimental condition 1500 respondents were invited to participate in the survey. In return for their participation respondents were offered the equivalent to $7.50. Reminder emails were sent 48 hours after the initial invitation to those who had not yet completed the survey. The number of valid completes in the five experimental cells were 602, 606, 601, 600, and 604 respectively.

Measuring the biases

The difference between cell 5 and cell 2 is an estimate of how large the full positional bias effect is (the effect of presentation order in tandem with coupling the screen-in and frequency questions). The difference between cell 5 and cell 3 is an estimate of how large positional bias effect as a result of presentation order-only is. The difference between cell 5 and cell 4 is an estimate of how large the title confusion bias is. The difference between cell 2 and cell 3 is an indication of how big the screen-in punishment bias is. The difference between cell 5 and cell 1 is an estimate of how large the joint effect of all biases is when they operate simultaneously.

Results

Table 2 below shows the results for the average screen-in rates and the mean number of reads across the five experimental cells.

Table 2
MEAN SCREEN-IN RATES AND MEAN READS ACROSS EXPERIMENTAL CONDITIONS

	Experimental Conditions				
	Cell 1 Maximum Biases	**Cell 2 Both Positional Biases***	**Cell 3 Pure Positional Bias**	**Cell 4 Title Confusion Bias**	**Cell 5 Minimum Bias**
Total sample	602	606	601	600	604
Screen-In Rates					
Mean**	6.69***	7.01	10.04	9.01	9.26
Standard deviation	5.69	5.59	7.03	6.96	6.70
Number of Reads					
Mean	3.64	3.78	4.80	4.44	4.69
Standard deviation	6.69	5.59	7.03	6.96	6.70

* **We note that the magazine presentation order was the same in experimental conditions 2 and 3.**

** **The means in the table are computed across magazines. For example under experimental condition 1 we found that respondents screen in about 6.69 times on average.**

*** **All differences are statistically significant (P< 0.05) except that means under experimental condition 1 are not statistically different from the means in experimental condition 2. The means in experimental conditions 3, 4 and 5 are not statistically different from each other. We used an analysis of variance to test for overall statistical significance of the treatment factor, and then applied Post-hoc tests using the Bonferroni method to adjust for the number of treatment levels.**

Table 2 contains results on two statistics. For each experimental condition we computed:

- The average number of titles screened in across respondents. For each respondent we counted the number of titles screened in and then computed the average of this number across respondents.
- A weighted reading-frequency statistic referred to as the number of reads. Respondent can claim different reading frequencies, i.e. 4 out of 4, 3 out of 4, 2 out of 4, 1 out of 4, and less than 1 out of 4. We assign to each reading-frequency claim a probability: the probability for 4 out of 4 is 1, the probability for 3 out 4 is .75, etc, and the probability for less than 1 out of 4 is 0.10. To obtain the number of reads statistic we multiply each probability with the number of respondents that have claimed the corresponding reading frequency, then we add up the different weighted probabilities.

In Table 2 we expect significant differences between experimental conditions 2 and 5 because the difference between results under these conditions indicates the size of the screen-in bias. Experimental condition 2 also contains the pure positional bias: but we note that the effect of pure positional bias cannot be detected at the average levels, shown in cells 3 and 5, because there is no reason why average screen-in rates would be affected by a random ordering. To detect pure positional biases we have to look at individual (groups of) magazines. The screen-in punishment bias is a bias that has been consistently found in the literature. As expected the differences between condition 2 and condition 5 are statistically significant, and in addition the difference is very meaningful: When the screen-in punishment bias is removed respondents claim to have read more magazines as opposed to the condition where the screen-in bias is present: the difference being 7.01 versus 9.26 represents an increased readership level of 32%. This is within the same order of magnitudes found in the previous studies, which ranged from 20% to 50%.

The results shown in Table 2 on title confusion bias are less clear: the statistics shown in cell 4 and 5 are not statistically significant from each other. The minimum bias condition and the title confusion condition have average screen-in rates that are 9.26 and 9.01 respectively. However, title confusion bias as was discussed earlier can operate in different ways: it can increase screen-in rates (in which case we have an over-estimation of audience) and it can also decrease screen-in rates (which was only to be expected for a small number of journals that were hypothesized to be similar in title or content and hence could be confused.

The average results presented in Table 2 are not able to identify pure positional bias, because in both a random presentation format as in a fixed presentation format positional bias occurs. In the random presentation format it just affects different magazines whereas in the fixed format it affects most likely the same magazine titles. To shed further light on the positional biases we therefore looked at the average screen-in rates of sub-groups of magazines. This may also shed further light on title

confusion biases. Recall that in experimental conditions 2 and 3 the same order of magazine title presentation was used. Magazines are grouped together in classes, A, B, C, D, E, F, G, H, I and J, and are shown in this order in experimental conditions 2 and 3. We calculated the average screen-in rates for each of these sub-groups of magazines: they are shown, for each experimental cell, in Table 3.

Table 3
MEAN SCREEN-IN RATES FOR SUB-GROUPS OF MAGAZINES

	Experimental Conditions				
	Cell 1 Maximum Bias	**Cell 2 Both Positional Biases**	**Cell 3 Pure Positional Bias**	**Cell 4 Title Confusion Bias**	**Cell 5 Minimum Bias**
Total sample	602	606	601	600	604
Mean Screen-In Rates					
Class A (4 publications)	0.89	1.14	1.30	1.13	1.17
Class B (3 publications)	0.28	0.33	0.38	0.39	0.39
Class C (6 publications)	0.99	1.09	1.55	1.25	1.40
Class D (5 publications)	1.32	1.07	1.40	1.23	1.38
Class E (6 publications)	0.43	0.58	0.73	0.63	0.71
Class F (6 publications)	0.42	0.54	0.92	0.64	0.72
Class G (5 publications)	0.29	0.27	0.54	0.44	0.48
Class H (6 publications)	1.22	1.21	1.99	1.88	1.69
Class I (5 publications)	0.50	0.38	0.60	0.75	0.67
Class J (4 publications)	0.36	0.39	0.63	0.69	0.67

We inspected the results presented in Table 3 specifically for positional biases. Comparing the results under experimental cell 2 and 5 again confirm a strong indication of a positional screen-in bias effect. Consider the following differences between cells 2 and 5:

Difference for magazine group A: 1.17-1.14=0.03 Decrease of 2%

Difference for magazine group F: 0.72-0.54=0.18 Decrease of 25%

Difference for magazine group J: 0.67-0.39= Decrease of 42%

These results are again consistent with the three previous studies that have identified positional screen-in bias effects and show positional screen-in bias becomes very severe for magazines toward the end of media part of the questionnaire. The differences in Table 3 between cell 3 and cell 5 are an indication of the existence of a pure positional bias effect. Most of the screen-in rates under cell 3 are higher than under cell 5: This is contrary to our expectations. In the discussion section we will offer some post-hoc hypotheses as to why we found this result.

The results in Table 3 also don't seem to indicate strong title confusion effects. The results in Table 3 above show groups of magazine titles for which some titles in each group are likely candidates for title confusion, and others less so. For example, one group consists of e-week, PC Magazine, PC World, and Smart Computing. In this group it is reasonable to assume that PC World and PC Magazine are potential candidates to be confused. The other titles in that group are less likely to be confused. Hence, to investigate this we calculated average screen-in rates under the different experimental conditions for those specific magazine titles that had a higher likelihood to be confused. In Table 4 (see below) we show the results for a number of these selected magazine pairs. We have given the specific magazines dummy names such as A1 and A2: these would represent the two most likely to be confused magazines for group A, etc. The pairs of magazines selected for this analysis included among others: PC World and PC Magazine, Forbes and Fortune, Family PC and Maximum PC, etc. Title confusion bias should be evident from the differences between cells 4 and 5 in Table 4. However, the screen-in percentages between cells 4 and 5 do not show any striking differences. This is partly contrary to what has been found previously in the literature. However, as argued in the introduction we believe that identifying 'title confusion' by comparing the results under grouping similar magazines with

results found under a condition where similar magazines are presented in separate positions may be more complicated that has been thought thus far. For those respondents who actually read both of the similar magazines there is no title confusion possible because they actually did read both magazines. In such a situation another effect can occur though: grouping both similar magazines together may actually benefit memory recall and hence lead to higher screen-in estimates for the grouped version.

Table 4
MEAN SCREEN-IN RATES FOR SELECTED PAIRS OF MAGAZINES

	Experimental conditions				
	Cell 1 Maximum Bias	**Cell 2 Both Positional Biases**	**Cell 3 Pure Positional Bias**	**Cell 4 Title Confusion Bias**	**Cell 5 Minimum Bias**
Total sample	602	606	601	600	604
Mean Screen-In Rates					
Title A1	29.4%	42.6%	47.1%	42.3%	42.2%
Title A2	36.2%	39.4%	45.3%	42.2%	44.0%
Title B1	7.0%	8.3%	9.5%	9.5%	10.8%
Title B2	7.0%	8.8%	7.8%	12.2%	9.6%
Title F1	4.7%	11.2%	18.8%	14.3%	13.1%
Title F2	11.8%	12.2%	20.0%	12.8%	17.4%
Title G1	11.6%	11.7%	19.0%	15.0%	16.9%
Title G2	5.8%	5.0%	13.1%	8.7%	9.9%
Title I1	8.6%	6.3%	9.5%	10.5%	9.9%
Title I2	10.0%	3.8%	7.0%	12.8%	9.8%

Conclusions and Discussion

The issue of bias continues to be an issue for many syndicated print media studies. We believe that we have shown how web-based questionnaire designs allow us to minimize the effects of the three biases explored in this paper. Our experiment provides the industry not only with the "potential" benefits of on-line interviewing techniques, but also with a relative measure of biasing effects when on-line interviewing is not possible.

Our experimental results showed a very strong positional screen-in punishment bias effect: magazines presented later in the media part can be under-estimated by as much as 40%. Although the size of such effects may vary across studies we note that previous studies have found effects in between 20% and 50%!

Our experimental results did not support our expectations regarding the pure positional biases. In the literature we found little support for this expectation. We can only speculate as to why this is. One Post hoc explanation could be that once the screen-in punishment effect is removed respondents just have to answer the "yes/no" screen-in questions: the task for the respondent then is in reality quite manageable (especially in our case where we had included the 50 larger titles).

The results regarding title confusion bias were also partially contrary to our expectations. As with the pure positional bias, the available empirical evidence for this bias in the literature is sparse and non-experimental. This provides a weak case to begin with. However, we also hypothesized that the issue of grouping similar magazine titles versus ungrouping similar magazine titles may also have other effects that title confusion. Grouping may not only help with eliminating title confusion bias (by removing a negative effect) it may also help the respondent by supporting memory recall: This is more likely in a case where a respondent reads both involved magazines but where one of the magazines is only infrequently read.

The current study does not explore any of the formatting issues associated with web questionnaires or other issues such as the number of pixels, the quality of the screens as they appear to the respondent. Nor does it explore issues involving the presentation of logos (color versus black and white), or covers versus logos. We believe these aspects to have ignorable effects. However, in addition to the biases discussed in this paper, a number of additional survey design elements still need to be investigated: the position of a title in relation to the overall length of the media part of a questionnaire, the total number of titles included, respondents' knowledge or expectation about the number of titles ahead of them, and the perceived length of the overall survey may all potentially lead to biasing effects. It has been found that longer surveys may have detrimental effects upon the overall quality of the data by leading to increased item non-response and increased use of undesired response styles (see Vriens, Wedel, and Sandor, 2001). Future research needs to assess these issues. Biases in print medium audience measurement have or may have an effect on the estimated screen-in rates and hence on the average issue audiences. This leads to sub-optimal decisions when the audience size results are used to estimate the exposure distribution of media schedules (e.g.

Danahar, 1988), a distribution needed to estimate reach, frequency and GRP of media schedules. This was not investigated in this study, and will be addressed in follow-up research.

It is often said that the more we know, the more we need to know. It is clear that our understanding of the biases explored in this experiment has helped to further the cause for "audience measurement" using web-based executions. Based on the results and arguments presented in this paper we have a strong case for web-based executions of print media audience measurement:

1. First, in web-based questionnaire design it is more difficult for respondents to gauge the length of the media part of the survey resulting in higher screen-in rates.
2. Second, a web-based questionnaire designs allow us to separate the screen-in questions from the follow-up questions and randomizing the presentation order of the included magazine titles while keeping similar titles grouped together. This as is shown in our paper reduces the (potential) negative effects of three biases: the screen-in punishment bias, the pure positional bias and the title confusion bias.

References

Appel, V. (1994), "Length of Screening Interval and Print Media Audience Estimates", **Journal of Advertising Research**, 34, 5, 22-26.

Appel, V., A. Elder, and T. Incalcatera (1999), "Measuring Print Audiences Via the Internet", **Proceedings of the Worldwide Readership Research Symposium**, Florence, pp. 509-516.

Advertising Research Foundation (1980). "ARF Comparability Study on Methods of Estimating Magazine Audiences-Report No. 2", New York, ARF.

Chrzan, K. (1994), "Three Kinds of Order Effects in Choice-Based Conjoint Analysis", **Marketing Letters**, 5, 2, pp. 165-172.

Cornish, P. and E. Meier (1983), "An Experiment with Grouped Titles", **Proceedings of the World Wide Readership Conference,** Montreal, pp. 215-219.

Couzens, M. (1988), "Audience Measurement: Inching Toward a Solution", **Folio**, April.

Danahar, P. (1988), "A Log-Linear Model for Predicting Magazine Audiences", **Journal of Marketing Research**, 25, pp. 356-362.

Day, R.L. (1969), "Position Bias in Paired Product Tests", **Journal of Marketing Research**, 6, 1, pp. 98-100.

Healey, T.E. (1997), "Title Confusion Versus Issue Frequency in a Mail Survey", **Proceedings of the Worldwide Readership Research Symposium**, Vancouver, pp. 147-152.

Mallet D.T. (1993), "The Relationship Between Screen-in Rates and Readership Levels in MRI and SMRB" **Journal of Advertising Research**, 33, 1: 18-22.

McGlathery, D. (1993), "Does Title Confusion Affect Magazine Audience Levels"? **Journal of Advertising Research**, February, 24-37.

Pinnell, J. and V. Appel (1995), "How Computerized Interviewing Eliminates the Screen-In Bias of Follow-Up Questions", **Proceedings of the Worldwide Readership Research Symposium**. Berlin, pp. 117-123.

Vriens, M., M. Wedel, and Z. Sandor (2001), "Split-Questionnaire Designs: A New Tool for Survey Design and Panel Management", **Marketing Research**, Summer, pp. 15-19.

Whipple, T. W. and M.K. McManamon (1992), "Primacy Order Effects in the Measurement of Trade Magazine Receipt and Readership", **Journal of Advertising Research**, September/October, pp. 24-29.

Whitley, E. (1981), "Some Rotation Effects in the British Survey", **Proceedings of the World Wide Readership Symposium,** New Orleans, pp. 247-249.

Appendix I: Grouped versus Ungrouped Magazine Title Presentation

Sample: 50 Publications Grouped

	Title	Frequency
A.	1. eWeek (formerly PC Week)	(weekly)
	2. PC Magazine	(biweekly)
	3. PC World	(monthly)
	4. Smart Computing	(monthly)
B.	1. FamilyPC	(13X/year)
	2. Maximum PC	(monthly)
	3. Yahoo! Internet Life	(monthly)
C.	1. Interactive Week	(weekly)
	2. InternetWeek	(weekly)
	3. Internet World	(biweekly)
	4. Network Computing	(biweekly)
	5. Network Magazine	(monthly)
	6. Network World	(weekly)
D.	1. CIO	(23X/year)
	2. Communications News	(monthly)
	3. Computerworld	(weekly)
	4. InformationWeek	(weekly)
	5. InfoWorld	(weekly)
E.	1. Business 2.0	(biweekly)
	2. Fast Company	(monthly)
	3. The Industry Standard	(weekly)
	4. Red Herring	(biweekly)
	5. Upside	(monthly)
	6. Wired	(monthly)
F.	1. Business Week	(weekly)
	2. The Economist	(weekly)
	3. Forbes	(biweekly)
	4. Fortune	(24X/year)
	5. Inc.	(18X/year)
	6. Smart Business (formerly PC Computing)	(monthly)
G.	1. Barron's	(weekly)
	2. Investor's Business Daily	(daily)
	3. Kiplinger's Personal Finance Magazine	(monthly)
	4. Money	(monthly)
	5. SmartMoney	(monthly)
H.	1. The New York Times	(daily)
	2. The New York Times	(Sunday Edition - weekly)
	3. USA Today	(Mon-Fri only)
	4. The Wall Street Journal	(Mon-Fri only)
	5. Newsweek	(weekly)
	6. Time	(weekly)
I.	1. Golf Digest	(monthly)
	2. Golf Magazine	(monthly)
	3. Men's Health	(10X/year)
	4. Men's Journal	(monthly)
	5. Sports Illustrated	(weekly)
J.	1. Entertainment Weekly	(weekly)
	2. The New Yorker	(weekly)
	3. People	(weekly)
	4. Vanity Fair	(monthly)

Sample: 50 Publications Ungrouped

Title	Frequency
1. Smart Business (formerly PC Computing)	(monthly)
2. Golf Magazine	(monthly)
3. Upside	(monthly)
4. InformationWeek	(weekly)
5. Fortune	(24X/year)
6. Newsweek	(weekly)
7. Computerworld	(weekly)
8. Inc.	(18X/year)
9. Communications News	(monthly)
10. Barron's	(weekly)
11. eWeek (formerly PC Week)	(weekly)
12. CIO	(23X/year)
13. Money	(monthly)
14. The New York Times	(daily)
15. Investor's Business Daily	(daily)
16. InfoWorld	(weekly)
17. Business 2.0	(biweekly)
18. The Wall Street Journal	(Mon-Fri only)
19. Network World	(weekly)
20. SmartMoney	(monthly)
21. Network Magazine	(monthly)
22. Golf Digest	(monthly)
23. Internet World	(biweekly)
24. Men's Health	(10X/year)
25. InternetWeek	(weekly)
26. Time	(weekly)
27. Sports Illustrated	(weekly)
28. Interactive Week	(weekly)
29. Yahoo! Internet Life	(monthly)
30. FamilyPC	(13X/year)
31. The New York Times	(Sunday edition-weekly)
32. Entertainment Weekly	(weekly)
33. Forbes	(biweekly)
34. Men's Journal	(monthly)
35. Smart Computing	(monthly)
36. The New Yorker	(weekly)
37. The Economist	(weekly)
38. Maximum PC	(monthly)
39. Fast Company	(monthly)
40. People	(weekly)
41. PC World	(monthly)
42. Business Week	(weekly)
43. Wired	(monthly)
44. Vanity Fair	(monthly)
45. Red Herring	(biweekly)
46. Kiplinger's Personal Finance Magazine	(monthly)
47. USA Today	(Mon-Fri only)
48. PC Magazine	(biweekly)
49. Network Computing	(biweekly)
50. The Industry Standard	(weekly)

Appendix 2. Screen Shots of Questionnaire Versions: Screen-In and Follow-Up Questions Linked and Separated.

Figure A2.a
Screen shot of the questionnaire version used in cell 1. The screen-in and follow-up questions are linked.

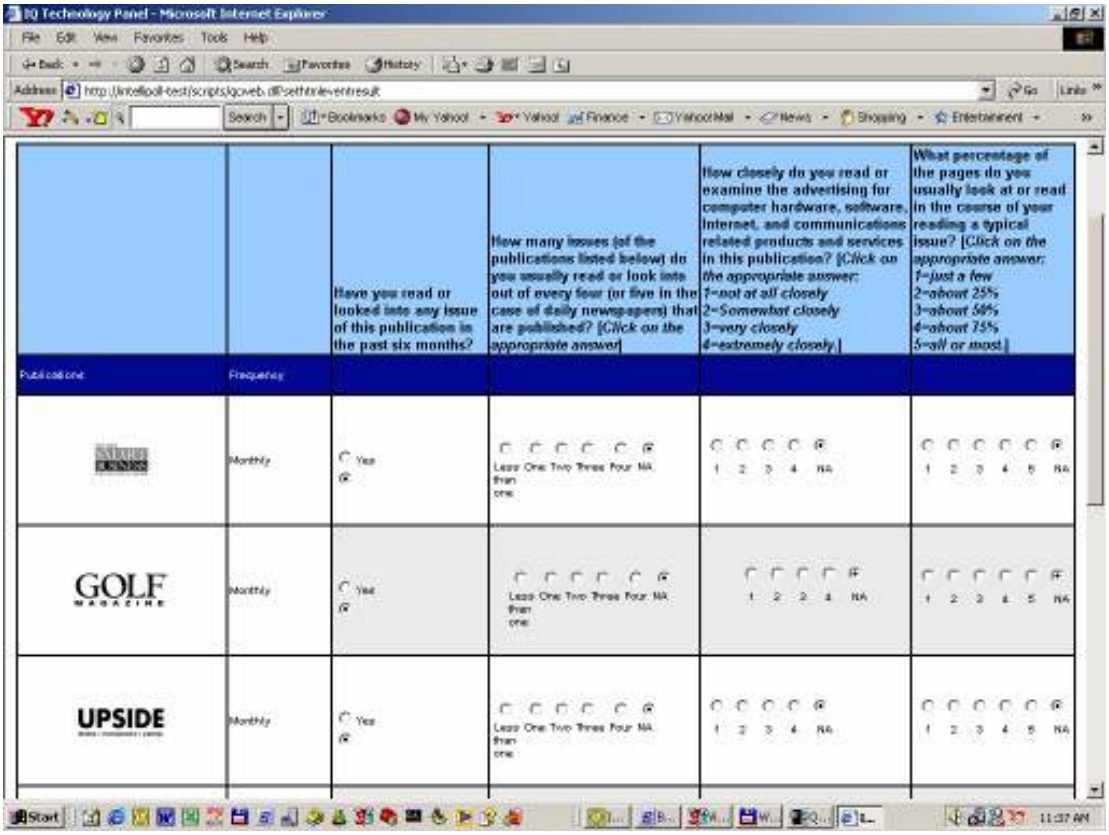

Figure A2.b
Screen shot of the questionnaire version for cell 2. The screen-in and follow-up questions are separated.