PDRF – About Propensity Weighting – emma in Australia

Adam Hodgson & Andrey Ponomarev – Ipsos Connect Australia

1. Introduction

It is not news for the research industry that over time, we have to face lower response rates from consumer surveys (Cook, 2000, Holbrook, 2008). It is not infrequent these days, especially when interview length is significant, to observe response rates as low as 10% or below.

In many cases, to make sure that the sample is representative of the universe it terms of a measureable profile, further weighting is considered enough to account for effects of different response rates. Indeed, if it can be safely assumed that the likelihood of the respondent agreeing to the survey is independent of characteristics of interest (collected via a questionnaire), or is independent if controlled for measureable profile, then non-response bias does not affect results of a study that uses proper sampling (Holbrook, 2008).

However, there are studies where there are reasons to believe that the likelihood to take part in a study is affected by the measured characteristics of interest of the respondent, and this cannot be accounted for via proxies such as demographics. It is clear that unless such a relationship is detected and accounted for, survey results will be biased.

Having a practical way to mitigate these effects is of particular importance to the media research industry. Whilst in some market research applications, *levels* of measured characteristics may not be as important as their *relationships and changes* across consumer segments or in time, in media research absolute levels of audiences are one of the most important results of a study.

Additionally, this issue is becoming of higher importance with the development of passive measurement methodologies used to overcome some of the more traditional research problems such as measurement error. Indeed, just as we do not have to worry as much about what questions to ask the respondent, what answer options to present and what stimuli to use, the problem of taking account of the bias introduced by respondent selection itself is becoming ever more important. This is especially relevant since response rates may be further negatively affected by such issues as respondent privacy concerns, the requirement to install applications on their computer or mobile phone, etc., all of which can widen the gap between respondents and the general population.

This article describes one of the possible approaches to overcoming the non-response bias issue, presents its practical application for the Australian emma survey, as well as outlines some practical advice for those seeking to apply a similar technique.

2. What is the salience effect and propensity weighting

The issue noted earlier can be described as *response bias*, whereby a respondent is more likely to participate in the study due to their higher interest in the study's subject (Jephcott, 2012).

People are known to have a *'propensity'* or interest or partiality towards different topics which may be covered in a study.

Most surveys involve asking participants a number of questions around a certain topic or topics. It is often possible to describe a salient feature of the study which would characterise it as whole. This salient feature is often incorporated into the recruitment process for the study, or in the study's introduction. The effect of including this feature will affect respondents differently depending on their *propensity* towards the study topic. This phenomenon can be referred to as the *salience effect.*

For example, we would expect more pet owners to complete a survey about pets than non-pet owners, especially when they are presented with the topic of the study during the recruitment process. Clearly, it follows that the proportion of people having some interest in the study topic would be higher compared to the universe, which would imply that the study's results can be biased.

The mitigation of such an effect raises two practical concerns:

- Measurement of the effect
- Developing and implementing a correction weight or a factor.

Measuring the effect can be achieved by developing a measure which would describe the respondent's propensity towards the study topic. In practice, it could be a general question in regards to relevant attitudes or behaviour. A corresponding benchmark will then need to be obtained from the general population including the study's non-respondents. In the case of the emma survey, it was found practical to obtain this measure of non-respondents during the (CATI) recruit interview.

The relationship between respondents and the total universe through such a measure needs then to be established. Once this relationship is established it can be applied to a more precise measurement of the study's subject matter to obtain final estimates, corrected for the bias introduced by the salience effect. In the case of the emma survey, this relationship was realised with an *adjustment factor*, being the ratio between the propensities of both respondents and non-respondents.

The method applied in the emma survey can be described as *propensity weighting,* whereby weights are produced based on the adjustment factor and respondent's answers to the full survey, which are then applied to the survey sample. The application of the propensity weights are very similar to the usual weighting procedure, with the difference being that one of the weighting targets is obtained by applying an adjustment factor to the survey results.

3. An overview of the emma survey and propensity weighting

emma (Enhanced Media Metrics Australia) is a large (annual 42,000 sample) continuous online survey of the Australian population (age 14+) which measures print readership, other media usage, attitudes, lifestyles and product usage. A monthly database is produced which reports the results of the past 12 months research, which is then distributed to publisher, media agency and other users (Green, 2013). As part of the design of the emma survey the emma Technical Committee (representing the publishing industry, media agencies and Ipsos) set out to measure any possible "response bias" in its readership measurement by asking all respondents contacted in the emma survey recruitment process (CATI) to answer 2 simple questions about their reading of magazines and newspapers. Importantly this includes those who having had the survey explained to them in the recruitment questionnaire declined to take part as well as those who agreed to take part in the survey. From this information the difference measure between readership non-survey takers and survey takers is obtained. Survey results are then adjusted to account for any response bias. In this way the emma survey accounts for the salience effect, by ensuring that both readers and non-readers are properly represented in the readership results. Thus the application of the propensity weights in emma ensures the most accurate estimate of print readership.

4. Propensity weighting in emma in more detail

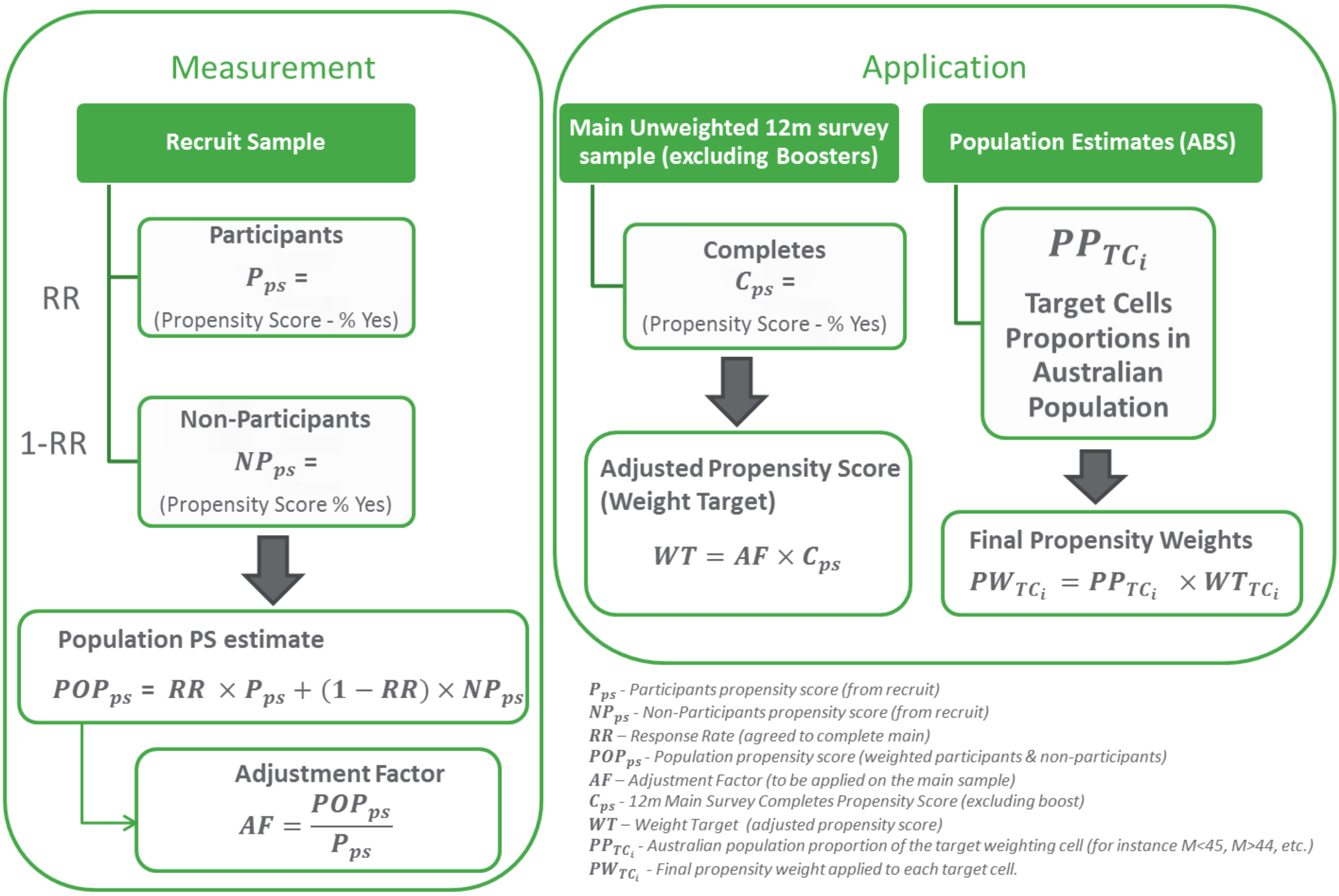
Propensity weighting has been used in emma from the launch of the survey fieldwork in 2012 to ensure that results accurately reflect both readers and non-readers of newspapers and magazines. The approach to propensity weighting used in emma was developed at the request of the emma Technical Committee by the respected media research statistician Jonathan Jephcott (Jephcott, 2012). The approach was designed to reduce survey bias by accounting for response bias, which could potentially lead to an overestimate of magazine and newspaper readership as reported in the emma survey. As advised by Jonathan Jephcott, in order to measure the salience effect, the (CATI) recruitment survey contains two questions about past week readership of newspapers and magazines for the emma main survey as follows.

N1.Have you read a newspaper in the last week?
N2.Have you read a magazine in the last week?

The reason for the question simplicity is to attempt to gain answers from as many people as possible, especially survey non-respondents. Another benefit of having a simple measure of respondent's propensity is easier operational application.

This design obtains comparative scores P_{ps} and NP_{ps} for both respondents who go on to participate in the main survey (participants) and those who choose not to participate (non-participants) respectively. Using these scores, estimates are then made for propensity across the total population (POP_{ps}), using response rate as an input for weighted average:

$$\boldsymbol{POP_{ps} = RR * P_{ps} + (1 - RR) * NP_{ps}}$$



Adjustment factor, as a *measure* of salience effect is then calculated as:

$$AF = \frac{POP_{ps}}{P_{ps}}$$

In 2012 the propensity adjustment factors were calculated as follows based on the results of the emma respondents (participants) and the weighted population number POP_{ps}

Type of Publication	Participants Score (P_{ps})	Population Score (POP_{ps})	Adjustment Factor (AF)
Newspapers	79.2%	71.5%	0.90 (71.5/79.2)
Magazines	57.4%	47.1%	0.82 (47.1/57.4)

Source: *Jephcott, 2012*

For the calculation of the final weighting target, unweighted estimates of newspaper and magazines readership (C_{ps}) from the emma survey are taken and the following propensity weighting targets (WT) achieved.

Type of Publication	Unweighted estimate from emma survey sample (C_{ps})	Adjustment Factor (AF)	Final weighting target (WT)
Newspapers	84.2%	0.90	75.9% (84.2*0.90)
Magazines	65.3%	0.82	53.6% (65.3*0.82)

Source: *Weighting Specs for emma project*

All the calculations above are made at a level of demographic chosen to control for, in case of emma project, at the Gender x Age level. However, for ease of representation, examples above are shown on a total level. The weighting targets (*WT*) obtained for each of the groups are then applied to its relevant population proportions PP_{tc} in order to obtain final weighting targets $PW_{tc.}$

Aggregated results of PW application on emma database can be seen in the table below. As seen, PW application has a significant effect on estimates obtained in the readership database. Industry sources have endorsed these adjustments as a credible improvement.

Type of publication	Change in total readership after PW application
Newspapers (M – F)	-11%
Newspapers (Sat)	-11%
Newspapers (Sun)	-11%
Magazines	-17%
NIMs	-10%

Source: *Jephcott, 2012*

Why an adjustment factor?

One might ask a reasonable question "Why would you use an adjustment factor and apply it to the full survey estimates, instead of using a POP_{ps} number directly as a weighting target?" This comes back to the very reason why propensity weights are used. The idea behind the procedure described above is to obtain a proxy measure of the difference between respondents and non-respondents, so that this ratio can be applied to the results of the full survey, rather than obtaining a ready-to-use population propensity estimate. Even though the question "Have you read newspaper in the last week?" is quite simple to answer, we have found that this is not an ideal descriptor of true respondent's behaviour in regards to newspaper readership. The emma survey utilizes a 15-minute long questionnaire about newspaper readership, facilitated by rich stimuli. Clearly, such a questionnaire will provide a better measure of whether a respondent has read a newspaper in the last week, compared to one simple question. Some discrepancy between these two estimates is to be expected, at least because during the recruit interview respondents are not presented with the same amount of recall stimuli compared to the emma questionnaire.

		Benchmark estimated "no"	Benchmark estimated "yes"
Full survey	"no"	1124 (9.5%)	724 (6.1%)
	"yes"	633 (5.3%)	9358(79.1%)

Accuracy	0.843
Precision	0.927

Source: *Jephcott, 2012*

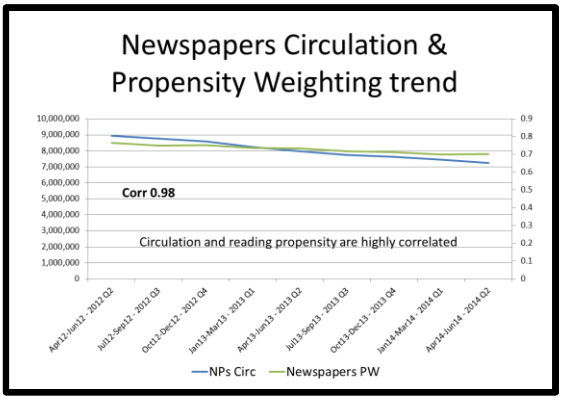
Even though the number of inconsistencies are of a relatively small order, we found that the propensity estimate is significantly higher in the full survey, even when measured across the same respondents, and so it is more appropriate to apply AF to the full survey measure as compared to taking POP_{ps} as the weighting target.

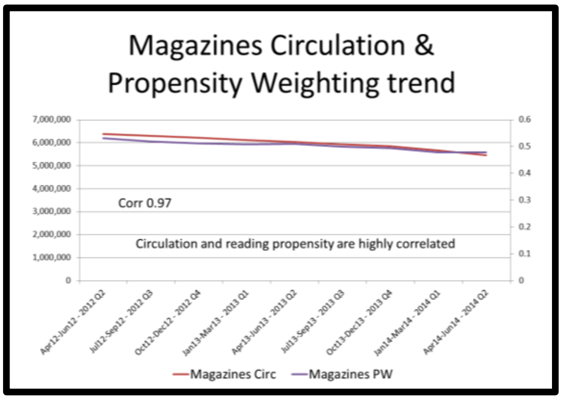
5. Updating weighting targets and adjustment factors

In his original 2012 paper Jonathan Jephcott advised that based on tracking six month's data, it was apparent that even though propensity scores change over time, benchmark based adjustment factors change slowly, so they could be treated as constants over significant periods of times.

At the same time, it became clear that the C_{ps} number, or full-survey based estimate, changes rapidly over time. In Australia print readership is now reducing significantly and so weighting targets need to be updated every month with the WT=AF x C_{ps} formula.

In order to validate this approach, a comparison was made between readership levels and circulations by quarter from April 2012 to June 2014. Encouragingly high correlations were found between the trends of both metrics across the period being 0.98 for newspapers and 0.97 for magazines (see below).





As mentioned, it was not expected that Adjustment Factors would be changed much at the beginning of the project. In order to check this assumption, a review of changes in C_{ps} and AF was undertaken in Nov 2014. The following dynamics were observed:

Type of Publication	$Cps_{\ 2012}$	Cps_{2014}	ΔCps	$\Delta Cps/Cps_{2012}$
Newspapers	84.2%	79.9%	-4.3%	-5.1%
Magazines	65.3%	60.7%	-4.6%	-7.0%

Source: PW review for emma project, 2012

Type of Publication	$AF_{\ 2012}$	$AF_{\ 2014}$	ΔAF	$\Delta AF/AF_{\ 2012}$
Newspapers	0.90	0.88	-2%	-2.3%
Magazines	0.82	0.80	-2%	-2.5%

Source: PW review for emma project, 2012

As seen from the tables, the dynamics in C_{ps} are more noticeable then dynamics in Adjustment Factors. But the tables clearly indicate that change in AF, though not as significant as in $C_{ps,}$ is still observable and significant.

Since weighting targets are affected by both:

$$\mathrm{WT} = \mathrm{C}_{\ \mathrm{ps}} * \mathrm{AF}$$

And change in weighting targets is almost equally dependant on changes on the two factors.

$$\frac{\Delta\mathrm{WT}}{\mathrm{WT}} = \frac{\Delta\mathrm{C}_{\ \mathrm{ps}}}{\mathrm{C}_{\ \mathrm{ps}}} + \frac{\Delta\mathrm{AF}}{\mathrm{AF}} + \frac{\Delta\mathrm{C}_{\ \mathrm{ps}}}{\mathrm{C}_{\ \mathrm{ps}}} * \frac{\Delta\mathrm{AF}}{\mathrm{AF}} \approx \frac{\Delta\mathrm{C}_{\ \mathrm{ps}}}{\mathrm{C}_{\ \mathrm{ps}}} + \frac{\Delta\mathrm{AF}}{\mathrm{AF}}$$

It is clear that change in Adjustment Factors cannot be ignored in line with the change in unweighted readership levels. As such these should be updated regularly, though possibly with a lower frequency compared to the $C_{\ ps}$ updates.

6. The importance of AF precision

As seen from the methodology and results described above, application of PW weights has a significant effect on estimates obtained from the readership database. In emma's case, a 10% to 17% reduction in readership estimates was observed, depending on the title.

As such, an important issue is precision of the adjustment factors themselves, as this is a measure which affects all of the estimates obtained in the database, and so any variance or bias in the Adjustment Factor is "contained" in all other estimates obtained in the database.

Since AF is used to calculate weighting target for the study, any estimate *E* in a database where propensity weighting is used, can be thought of as a function of unadjusted estimate *U* and Adjustment Factor *AF:*

$$\boldsymbol{E = f(U, AF)}$$

Hence,

$$\boldsymbol{Var(E) = f(E(U), E(AF), Var(U), Var(AF), Covar(U, AF))}$$

So it can be seen that any bias or variance in Adjustment Factors is inherited in any other database estimates.

The Adjustment Factor variance is highly dependent on the sample of non-participants, thus the quality of all estimates in the database is hugely dependant on this number. This can be demonstrated via the following:

$$\boldsymbol{AF = \frac{POP_{ps}}{P_{ps}}}$$

Since

$$\boldsymbol{POP_{ps} = RR * P_{ps} + (1 - RR) * NP_{ps}}$$

clearly,

$$\boldsymbol{AF = RR + (1 - RR) * NP_{ps}/P_{ps}}$$

For the purpose of this exercise, RR can be treated as constant. Hence,

$$\boldsymbol{Var(AF) = Var\left[(1 - RR) * \frac{NP_{ps}}{P_{ps}}\right] = (1 - RR)^2 * Var(\frac{NP_{ps}}{P_{ps}})}$$

Clearly, since $\boldsymbol{NP_{ps}}$ and its variance are dependent on the non-respondent sample, AF and its variance are dependent on this parameter as well, and the salience of this dependency grows as the RR decreases. In other words, the lower the response rate for the survey, the more important non-respondents answers are for the final estimates in the database.

In summary, it is of extreme importance that the researcher choosing a propensity weighting approach should pay particular attention to the quality of the AF estimate. This is because in further analysis, any estimate in the database has a variance component introduced by the use of AF. This is quite different from traditional weighting approaches, where population estimates are obtained from third-party census-like sources and do not inherit variance.

7. Subgroup analysis

One of the most important considerations when using propensity weight procedures is considering differences in salience effect across subgroups of interest. Indeed, we would expect that if a *salience effect* exists in the recruitment process, the impact of a salience effect on response rates may differ depending on the subgroup of respondents.

Some groups, for example, might be generally more likely to fill in the survey; hence the magnitude of salience effect is relatively small compared to other groups. Another factor to consider is possible different levels of propensity towards the topic by subgroup. Where this difference is low, the effect of salience on sample composition would be relatively low compared to other subgroups.

Since salience effects can be different across the subgroups, before applying propensity weights one should check for significant differences in adjustment factors across the subgroups, and apply propensity weighting by subgroup where required. Having said that, the earlier point about variances still stands, and a balance between detailed analysis and high variance needs to be found.

In emma's case, the most notable differences were observed at the Gender & Age level.

For Newspapers:

	Non-Respondents		Respondents			
	Base	Propensity NP_{ps}	Base	Propensity P_{ps}	Propensity POP_{ps}	Adjustment Factor (*AF*)
Male 14-44	439	46.7%	3343	57.6%	48.1%	0.83
Male 45+	995	71.7%	2672	80.5%	72.8%	0.90
Female 14-44	341	45.2%	3136	51.9%	46.0%	0.89
Female 45+	1601	74.3%	2970	79.5%	75.0%	0.94

Source: Newspapers Weighting Specs for emma project

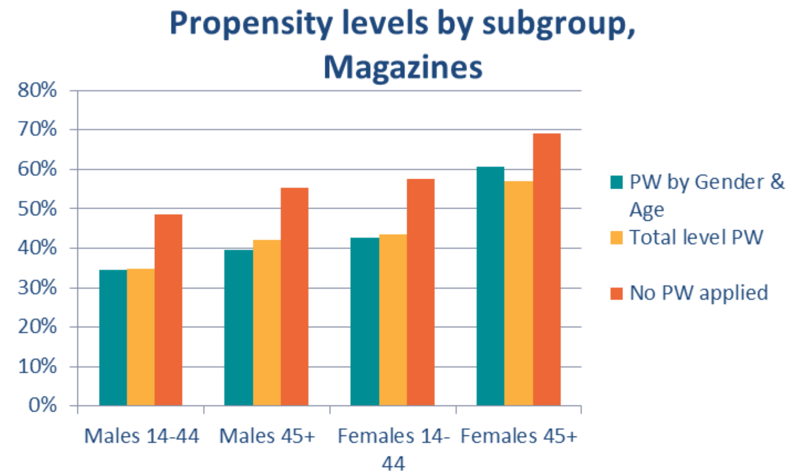
For Magazines:

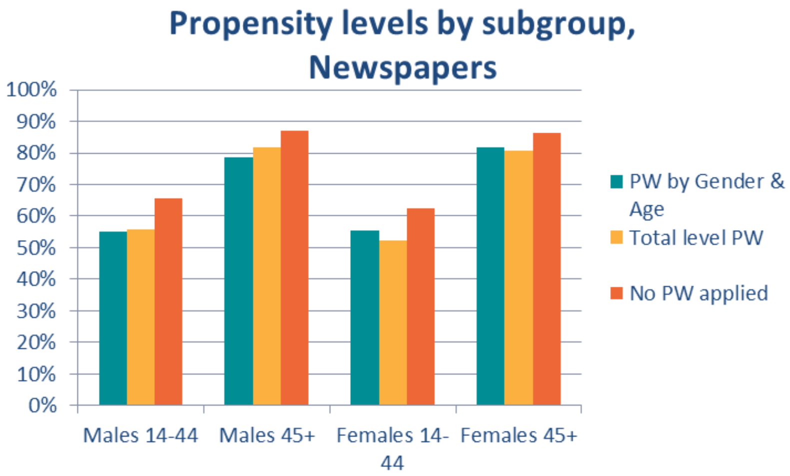
	Non-Respondents		Respondents			
	Base	Propensity NP_{ps}	Base	Propensity P_{ps}	Propensity POP_{ps}	Adjustment Factor (*AF*)
Male 14-44	438	24.4%	3343	35.0%	25.8%	0.74
Male 45+	995	34.9%	2672	50.3%	36.8%	0.73
Female 14-44	339	32.2%	3136	43.3%	33.5%	0.77
Female 45+	1599	54.0%	2970	61.7%	54.9%	0.89

Source: Magazines Weighting Specs for emma project

It was noticed that there is a significant difference between subgroups, both in propensity levels and magnitude of salience effect. Unfortunately, it is hard to relate this difference to a particular reason; however one might theorize that it might be interconnected with some of the possible difference reasons outlined earlier.

In order to validate the application of propensity weights by age and gender, the following comparisons were made by gender & cells separately between the unweighted database, propensity weights applied on total level, and propensity weights applied by gender & cells separately:





As seen from the charts, general level propensity weight application reaches the goal of estimate corrections on a broad level. However, splitting the weighting by appropriate groups adds value by more precisely correcting audience estimates, as well as facilitating more accurate profiling of the database.

8. In Conclusion

This paper illustrates a case study of the application of a propensity weighting approach to correct a response bias in a large-scale readership study. The approach taken is by asking two simple questions during the recruitment interview; finding out differences of propensity levels between respondents and non-respondents for the study; then applying this ratio to more precisely measure propensity levels among respondents, to finally arrive at weighting targets which are applied to the database during the routine weighting process.

Some of the authors' learnings about the subject matter are presented in this paper, including an approach to updating weighting targets calculated, the importance of the AF measurement precision and, finally, an understanding of differences across subgroups of respondents.

The procedure itself is efficient operationally, does not require significant time or resource investment, and can be used quite generally across market research studies.

The authors believe that the application of such a methodology can be beneficial to a number of studies across the industry and hope that this paper provides some clear directions and recommendations for its application.

References:

Holbrook, A.L., Krosnick, J.A and Pfent, R (2008) **The Causes and consequences of response rates in surveys by news media and government contractor survey research firms.** Ch 23, Advances in telephone survey methodology

Cook, C, Heath, F and Thompson, R (2000) **Meta-Analysis of Response Rates in Web- or Internet-Based Surveys.** Educational and Psychological Measurement **December 2000** vol. 60 no. 6 **821-836.**

Jephcott, J (2012) **Propensity weighting and the Salience Effect**

Green, A, White, H (2013) **Australia: Building a 21st Century Readership Survey**, Print & Digital Research Forum, Nice