Alternative Statistical Techniques to Address Non-Response Bias

Julian Baim, Martin Frankel, Joe Agresti, Jim Collins, Risa Becker and Van Podberesky (GfK MRI)

Introduction

Evidence in the United States (and we suspect elsewhere) shows a continuing deterioration in response rates for all survey modes using probability samples (Krueger and West, 2014). GfK MRI's response rate, for example, has declined substantially over the past 5 years. We also observe these declines for all types of research: government, academic and commercial. The impact of response rate decreases extends well beyond increased survey cost or the need for additional resources; it specifically threatens the validity of the survey findings by introducing the greater likelihood of non-response bias into the resulting estimates. While our efforts to stem the tide of declining response rates focus on changing incentives, increasing the number of callbacks or other interviewing strategies, others have devoted their energies to developing statistical adjustments to confront the **potential of non-response bias**. In regard to national readership surveys, the concern is that non-respondents' reading habits are different, either in aggregate or selectively, than those of study respondents. This paper employs the statistical adjustment strategy by using auxiliary data or paradata in post-survey weighting algorithms to explore differences in reported reading estimates among or between the adjustment procedures.

Statistical approaches to address survey non-response are not a new phenomenon (See Little, 1986) and there is a growing body of literature describing various results of the research (e.g., Krueger and West, 2014; Biemer and Peytchev, 2013). The underlying rationale of these adjustment efforts is to predict the probability of response, using available data from marketing databases for respondents and non-respondents alike, and use these probabilities to correct or address the **potential of non-response bias** in the resulting research. (It should be noted the basic idea to account for the probability of responding in marketing and media studies can be traced to the use of "Nights at Home" (Politz and Simmons, 1949.) In addition to using predictive modeling techniques, research in this area has been further enabled by the growing availability of auxiliary data for an entire sampling frame, helping analysts to classify the selected sample on a number of variables possibly correlated with the survey objectives. The theoretical rationale for these adjustments is predicated on the assumption these additional variables are predictive of both survey response propensities and are covariates of print readership. Our paper follows the prescriptive techniques of prior research and makes use of a number of auxiliary variables appended to GfK MRI's National study sampling frame to evaluate the potential of non-response bias on magazine reading.

Background and Research Design

GfK MRI's National Readership Study (The Survey of the American Consumer) consists of approximately 25,000 in-person interviews per year. Data are weighted to project to the adult population (age 18+) in the 48 contiguous states of the U.S. GfK MRI employs an area-probability sample design with differential selection rates based on income, geography and the number of eligible adults present in the household. Weighting is conducted in 3 stages:

1. Using the inverse of the probabilities of selection
2. Using the inverse of response rates by sample classes
3. Post-stratifying demographic variables to conform to known U.S. Census estimates

The second of these stages is a well-accepted, traditional method of accounting for non-response in survey execution and is based on the assumption that non-respondents and respondents in a weighting class behave similarly for variables under study. While we make use of variables available to the complete sampling frame, the current procedure is a broad-based approach to non-response adjustment. Although this stage is relatively broad-based, the final post-stratification stage includes numerous demographic variables that are historically correlated with media behavior (e.g., household income, educational attainment, etc.)

Our alternative approach to account for and address non-response is also a three-stage process. First we appended a number of variables made available to our entire household sampling frame. These data sources are generally attributed to Census block groups, are accessed from either U.S. Bureau of the Census information or marketing databases, and are appended to all selected households in the sampled block group. The following variables were appended:

- Income sampling strata (i.e., estimated median household income for block groups to create classifications of upper income quartile households, 2nd income quartile households and the bottom half of median household income households)
- Urban/Suburban/Rural designation of the selected block group
- Census region and division
- PRIZM (a geodemographic clustering designation) classification
- Estimated percent of African-Americans in the block group
- Estimated percent of Hispanics in the block group
- Primary Sampling Unit strata

In addition to these auxiliary data, we added number of attempts made to complete an interview at each household, regardless if the final disposition was a complete or not. This data element reflects the difficulty of obtaining an interview and may very well measure the difference in magazine readership, for example, among early responders, late responders and non-responders.

We then conducted a logistic regression with the dependent variable being either a completed interview ("1") or an eligible, but not completed interview ("0"). The logistic regression used all of the above variables as main effects, but did not introduce any second or third order interactive effects into the equation. The equation also used "dummy variables" to allow for evaluation of non-interval level variables. The results indicated that all of the variables were significantly different from 0 at the .05 level[1]

Using the resulting logistic regression equation, we generated a response propensity score (i.e., the likelihood for respondent to complete a survey) for every eligible household. This response propensity score accounted for all the independent variables associated with the household and, in almost all cases, generated a unique score for each household. In line with prior research, we employed two different weighting procedures at the second stage (see above) using these propensity scores. The first (Propensity Model 1) simply applied the inverse of the individual propensity score to all respondents, respectively, in the Spring GfK MRI National 2014 study; the second (Propensity Model 2) created four quartiles of propensity scores and applied the inverse of the average propensity score to all respondents in the respective quartiles. Both of these procedures are alternatives to our current non-response adjustment weighting stage and represent an effort to account for more granular predictors of non-response in generating magazine and newspaper audience estimates.

Analysis

Our analysis was straightforward; we compared the generated magazine and newspaper average-issue audiences from each of the weighting procedures (i.e., Propensity Model 1, Propensity Model 2, Current Weighting scheme for non-response). Since Propensity Model 1 is the most granular, comprehensive non-response adjustment procedure of the three, the print audiences from Propensity Model 1 were deemed as "unbiased" and served as the basis for comparison to the other two methods. In effect, we analyzed the direction and magnitude of differences between the audience estimates from the other two procedures, respectively, and results from Propensity Model 1. At the same time, we recognized the introduction of more discrete weights in the propensity models would likely reduce the overall sample efficiency, thereby potentially increasing the sampling error associated with individual audience estimates while simultaneously reducing bias. By calculating the root mean square error of the estimates, we assessed the joint impact of variance and bias changes for the three alternative weighting procedures.[2]

The *overall total readership differences* between these propensity models and the current non-response adjustment were only 0.001%. More discrete comparisons (Table 1) showed virtually no overall changes in average-issue audiences per title, regardless of the non-response adjustment weighting procedure. Among the 251 publications (magazines, magazine groups and newspapers), the average-issue audience per title for the 251 publications from Propensity Model 1 was +.12% higher than the averages generated from our current weighting procedure. The similar comparisons for average-issue audience for men and

[1] Incidence of Hispanics, African-Americans, income strata, urbanicity, and call attempts were all significant at the .0001 level.

[2] The root mean square error helps assess the tradeoff between reducing bias, for example, and decreasing reliability. If an alternative weighting procedure increases variance at a higher rate than it reduces bias compared to the current method, the recommendation is generally made not to employ that alternative weighting scheme.

women were -0.23% and +0.06%, respectively. The comparable analyses between Propensity Model 1 and Propensity Model 2 also showed insignificant, negligible differences (see Table 1).

We also examined whether the absence of overall differences masked systematic variances for particular magazine genres. The analysis (see Charts 1-6) clearly illustrates that most of the magazines clustered around the center, reflecting no differences, and very few magazines experienced audience changes of + or -5%. No magazine genres demonstrated any systematic differences in audience among the different non-response adjustment techniques.

TABLE 1
Change in Average-Issue Audiences: A Comparison of Non-Response Adjustment Procedures
(Total = 251 Publications)

	Propensity Model 1/ Current Procedure (Adults)	Propensity Model 1/ Current Procedure (Men)	Propensity Model 1/ Current Procedure (Women)	Propensity Model 1/Propensity Model 2 (Adults)	Propensity Model 1/Propensity Model 2 (Men)	Propensity Model 1/Propensity Model 2 (Women)
Change in Audience (%)	+0.12%	-0.23%	+0.06%	-0.13%	+0.05%	-0.38%

Chart 1

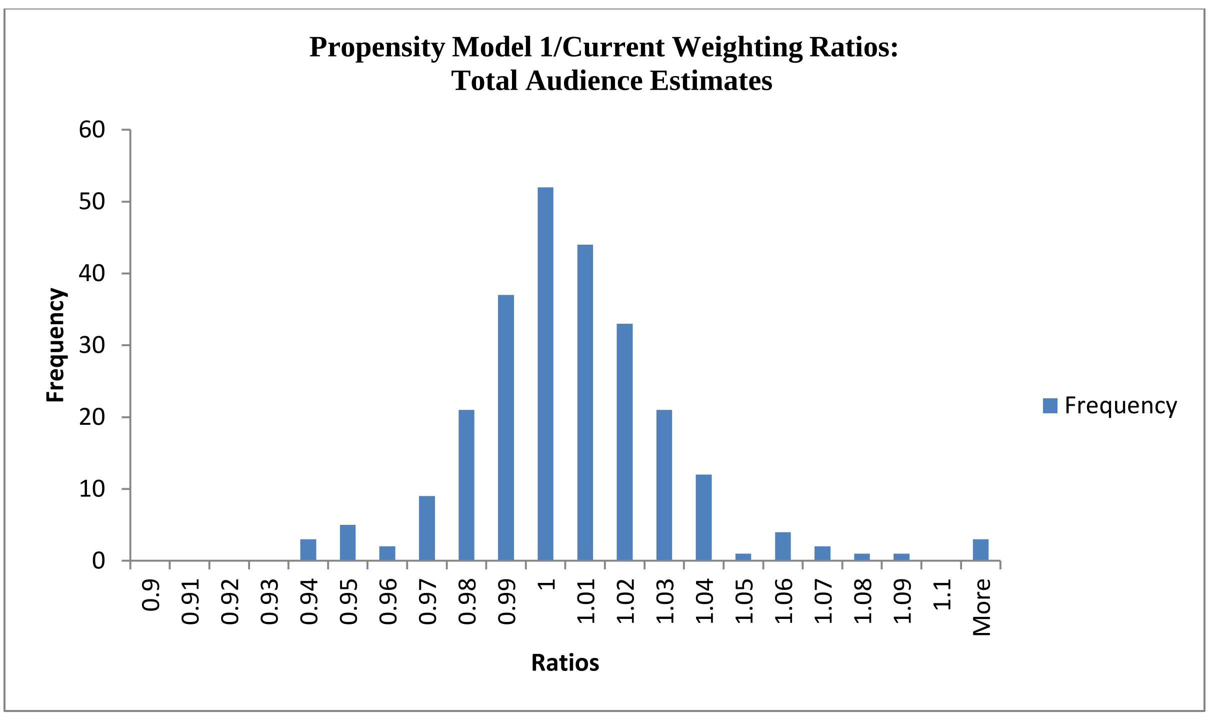


Chart 2

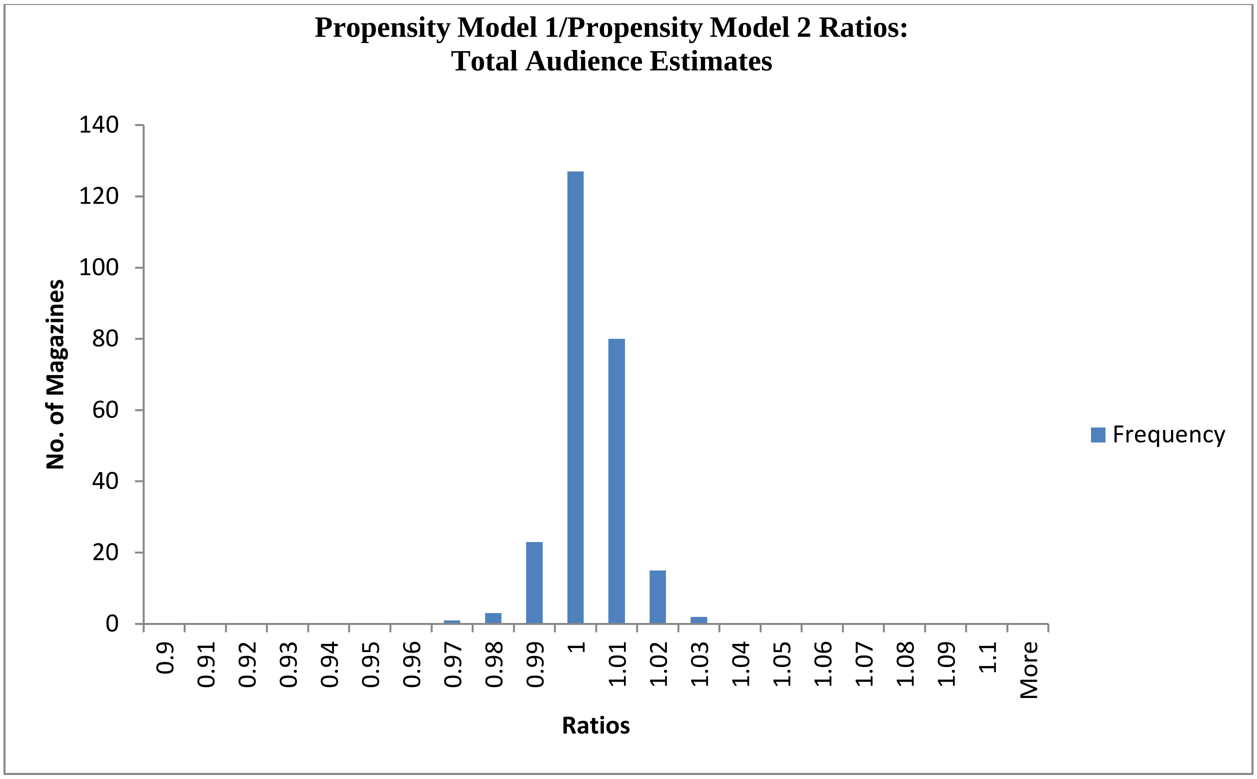


Chart 3

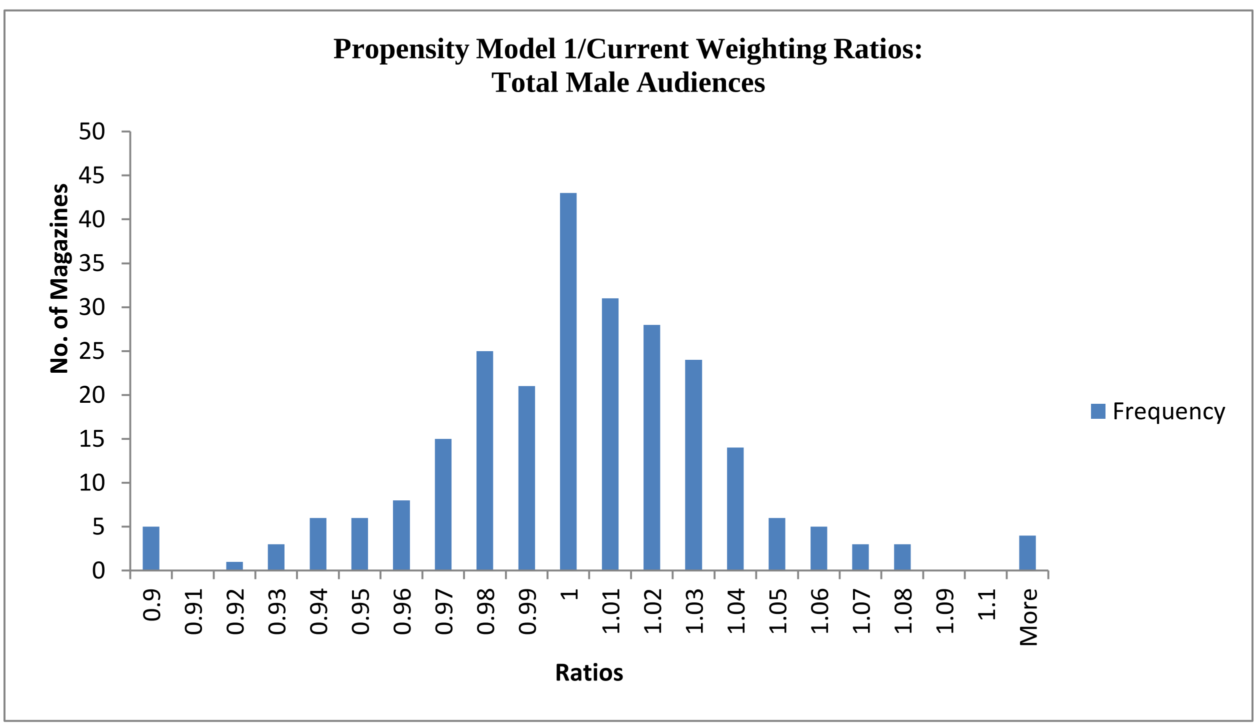


Chart 4

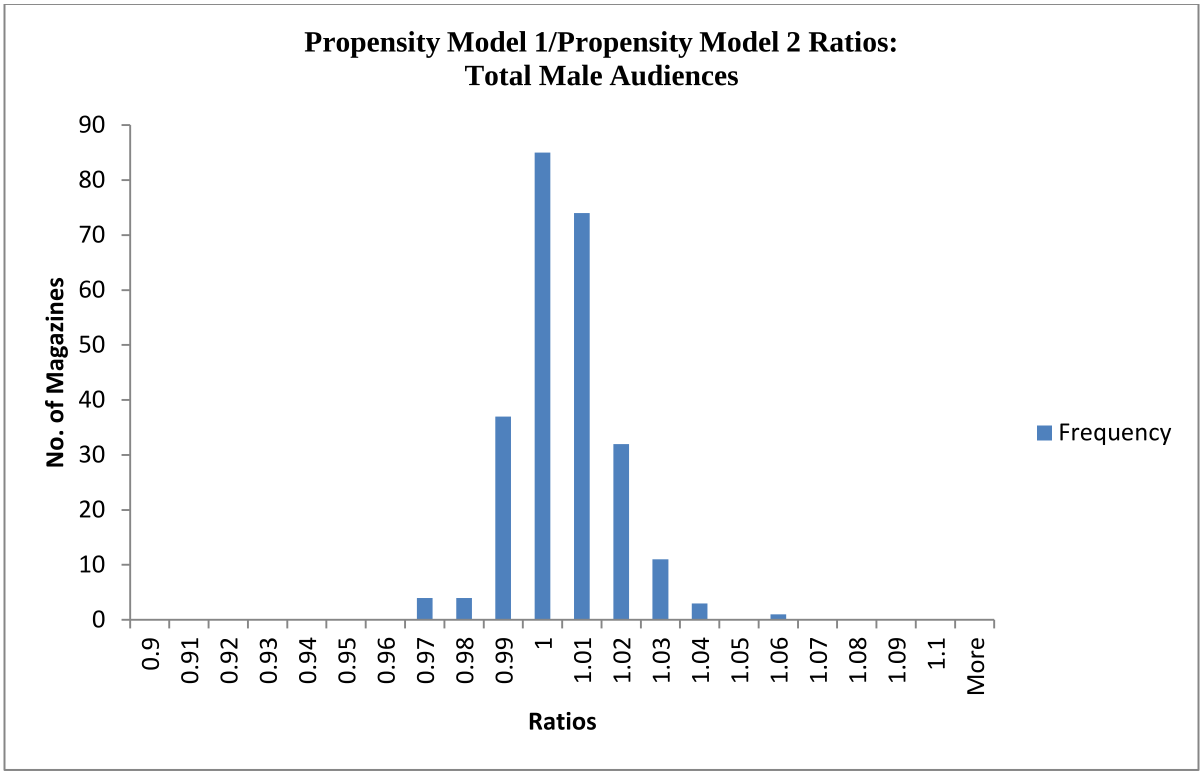


Chart 5

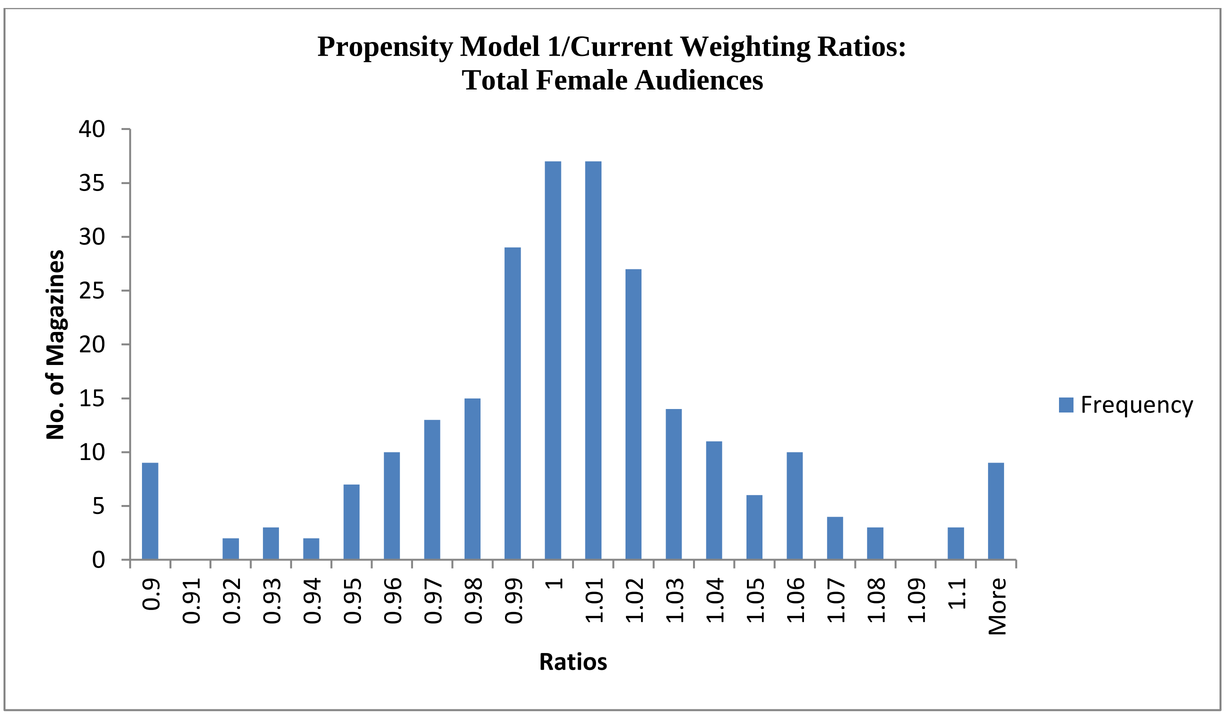
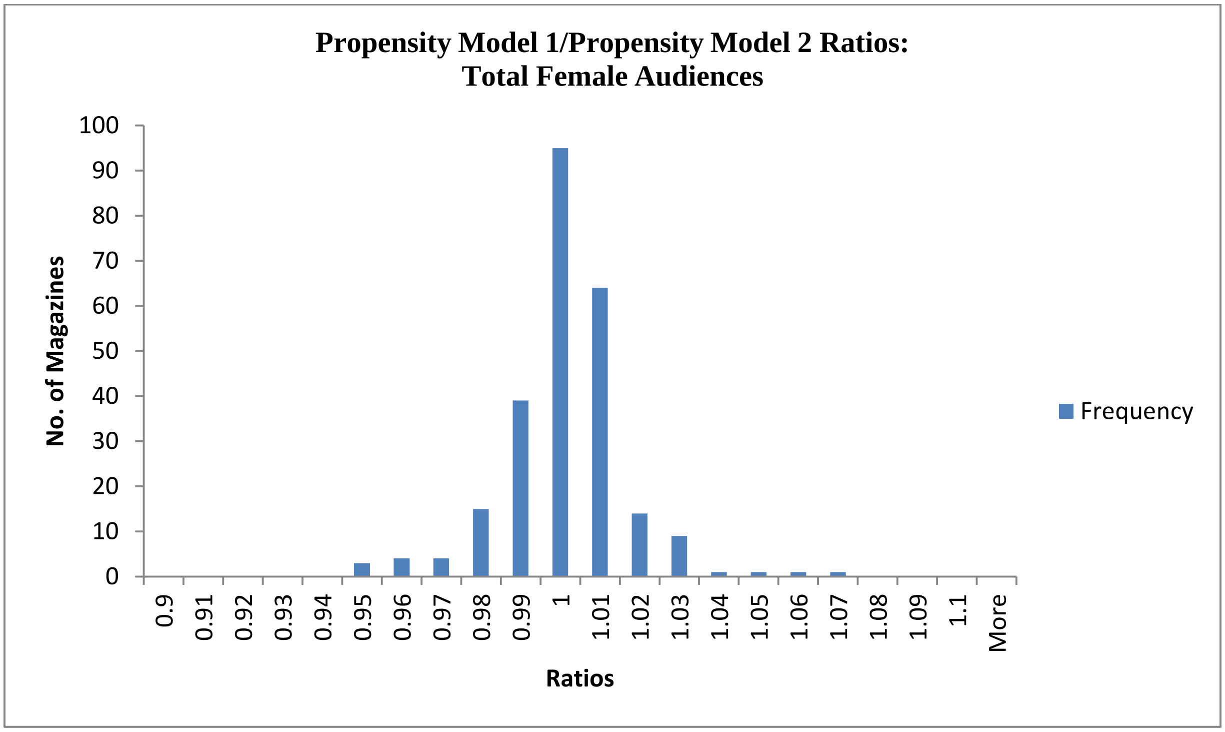


Chart 6



Even though both propensity models had very little impact on total audiences compared to the current non-response adjustment algorithm, it was still possible that audience profiles were affected by the more comprehensive propensity models. Since median household income estimates for magazines are particularly important to publishers, we examined the differences in median household income estimates among the three models. Similar to our total audience findings, Table 2 shows median household incomes for the 251 titles varied minimally among the three methods. Propensity Model 1 showed a modest 0.78% average increase in adult median household income for the 251 titles. The comparison between the two propensity models showed a -0.09% relative difference in median household income averaged across the 251 titles.

TABLE 2

Change in Median Household Income Estimates: A Comparison of Non-Response Adjustment Procedures

(Total = 251 Publications)

	Propensity Model 1/ Current Procedure (Adults)	Propensity Model 1/ Current Procedure (Men)	Propensity Model 1/ Current Procedure (Women)	Propensity Model 1/Propensity Model 2 (Adults)	Propensity Model 1/Propensity Model 2 (Men)	Propensity Model 1/Propensity Model 2 (Women)
Change in Median HHI (%)	+0.78%	+1.49%	+0.54%	-0.09%	-0.30%	-0.07%

A further examination of the distribution of median household income differences by title (Charts 7-12) clearly demonstrated the absence of any systematic differences across the titles. Simply stated, the introduction of more granular non-response adjustments failed to impact currently reported profiles and provided renewed confidence in our current procedures.

Chart 7

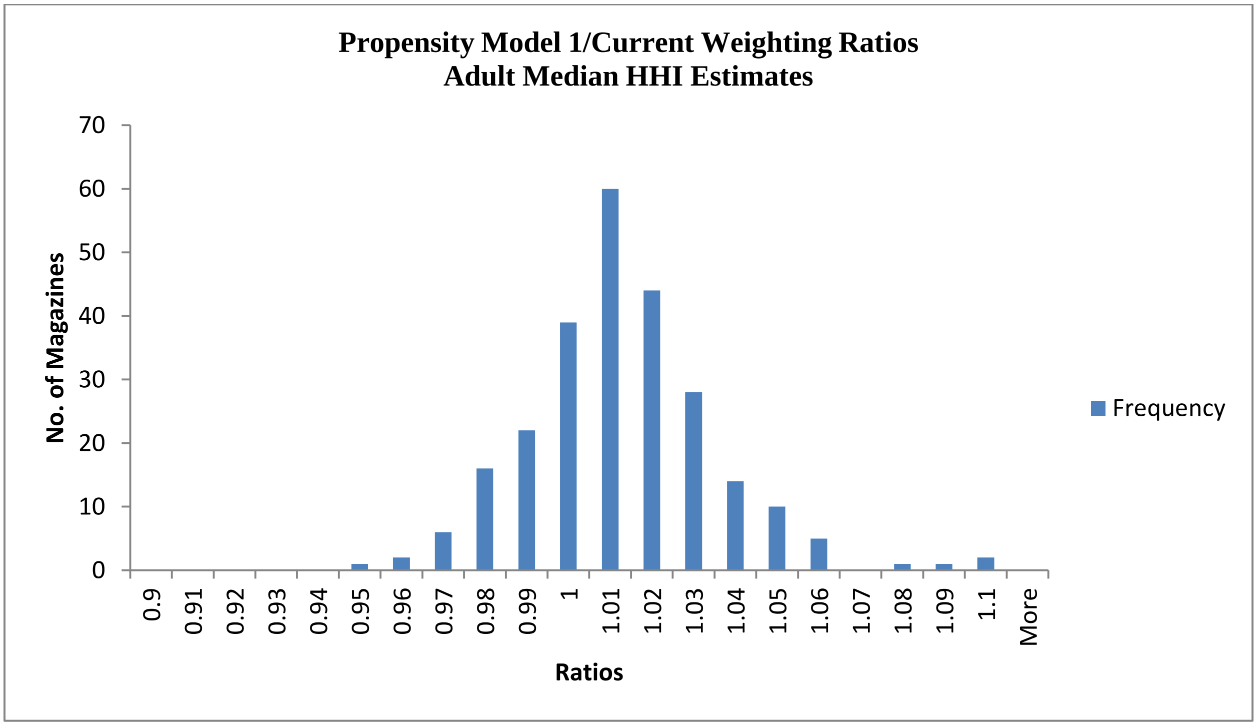


Chart 8

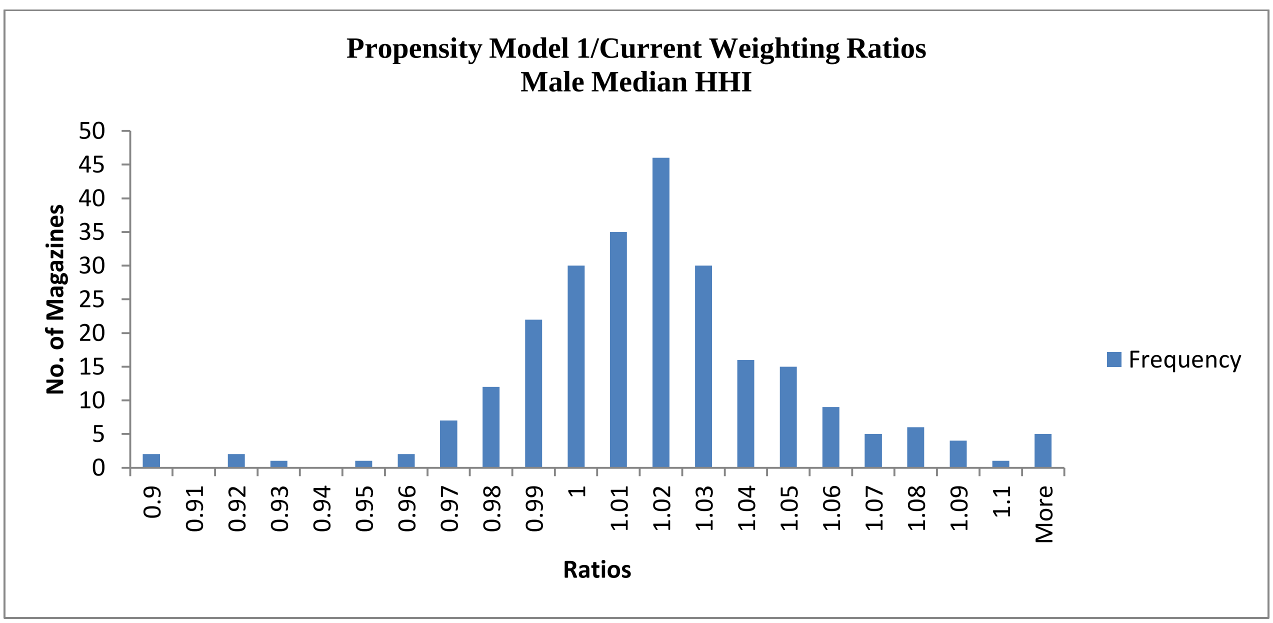


Chart 9

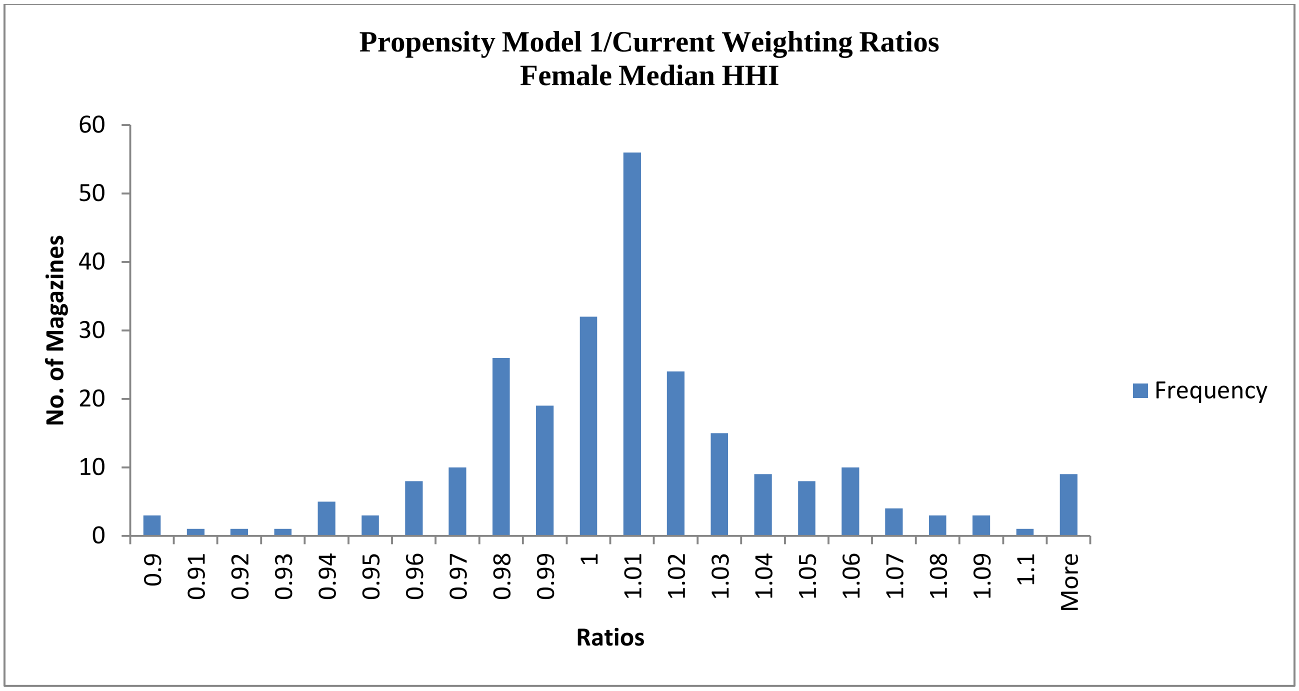


Chart 10

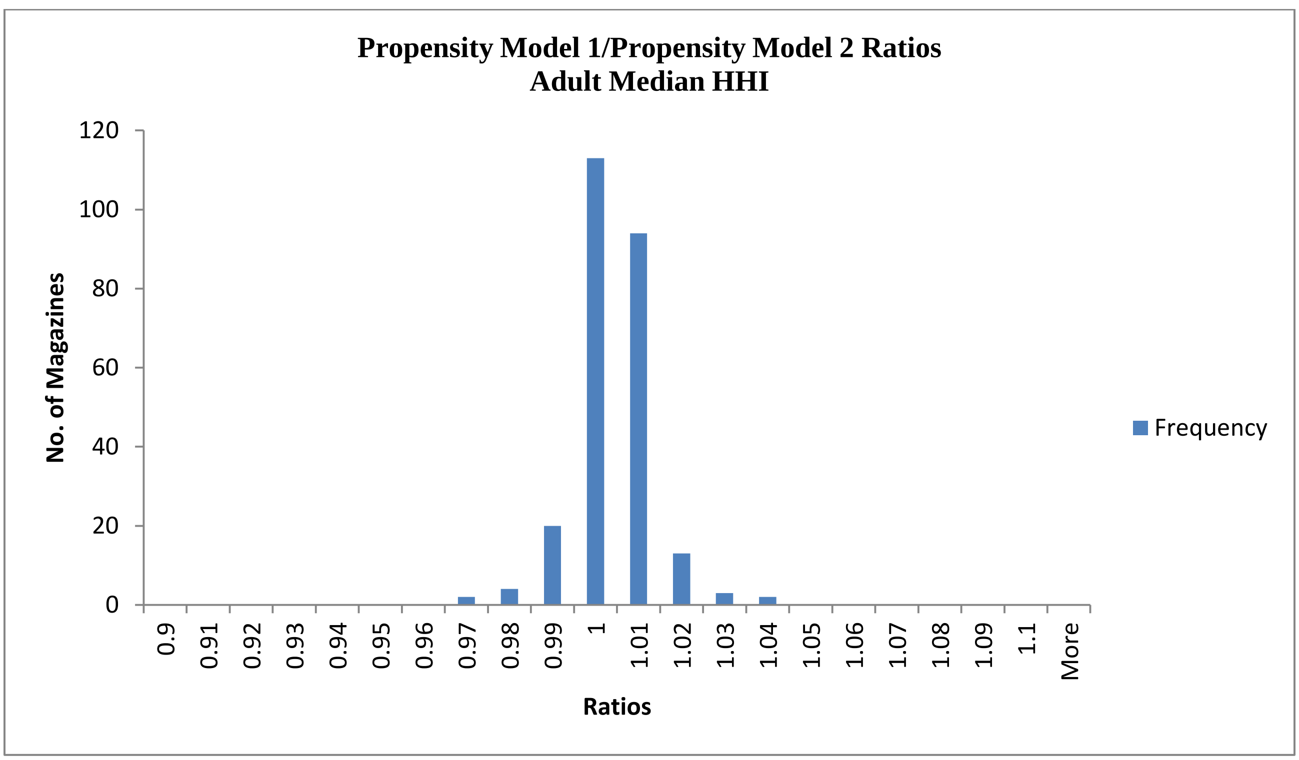


Chart 11

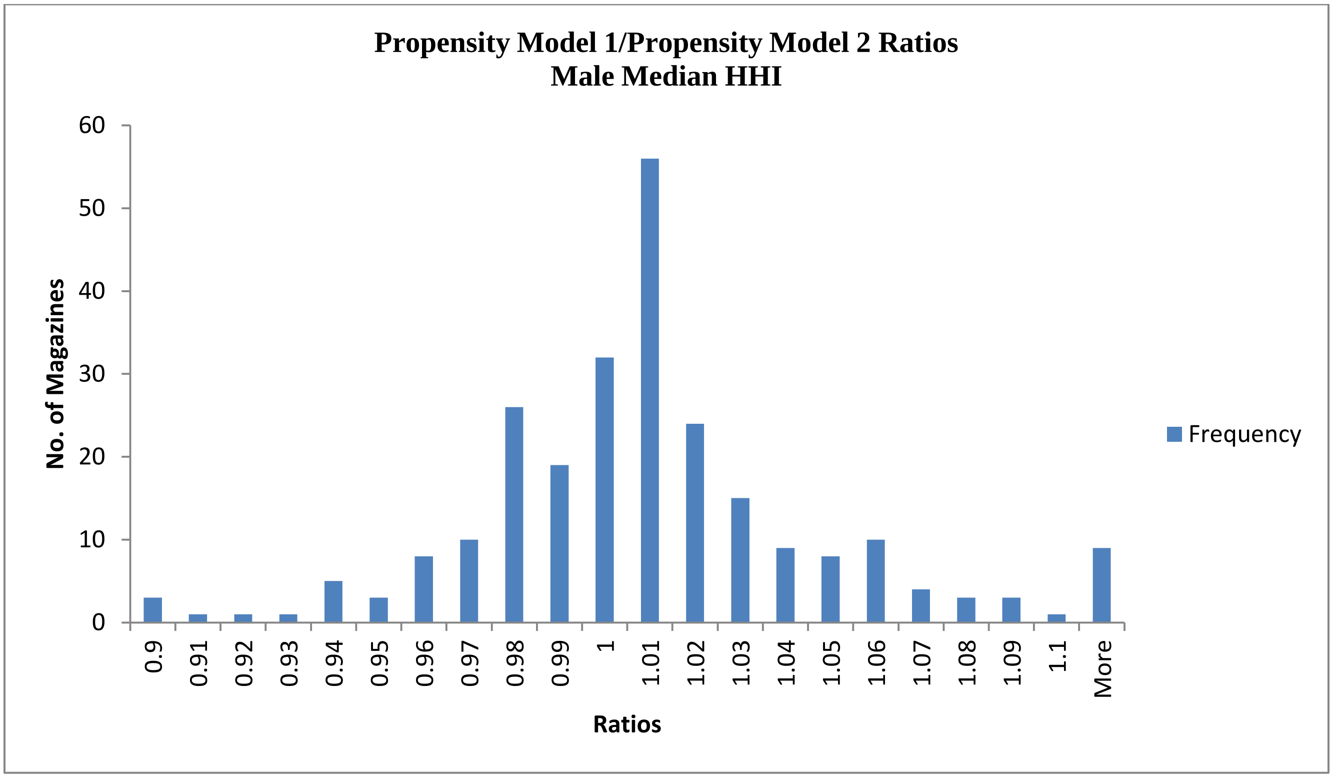
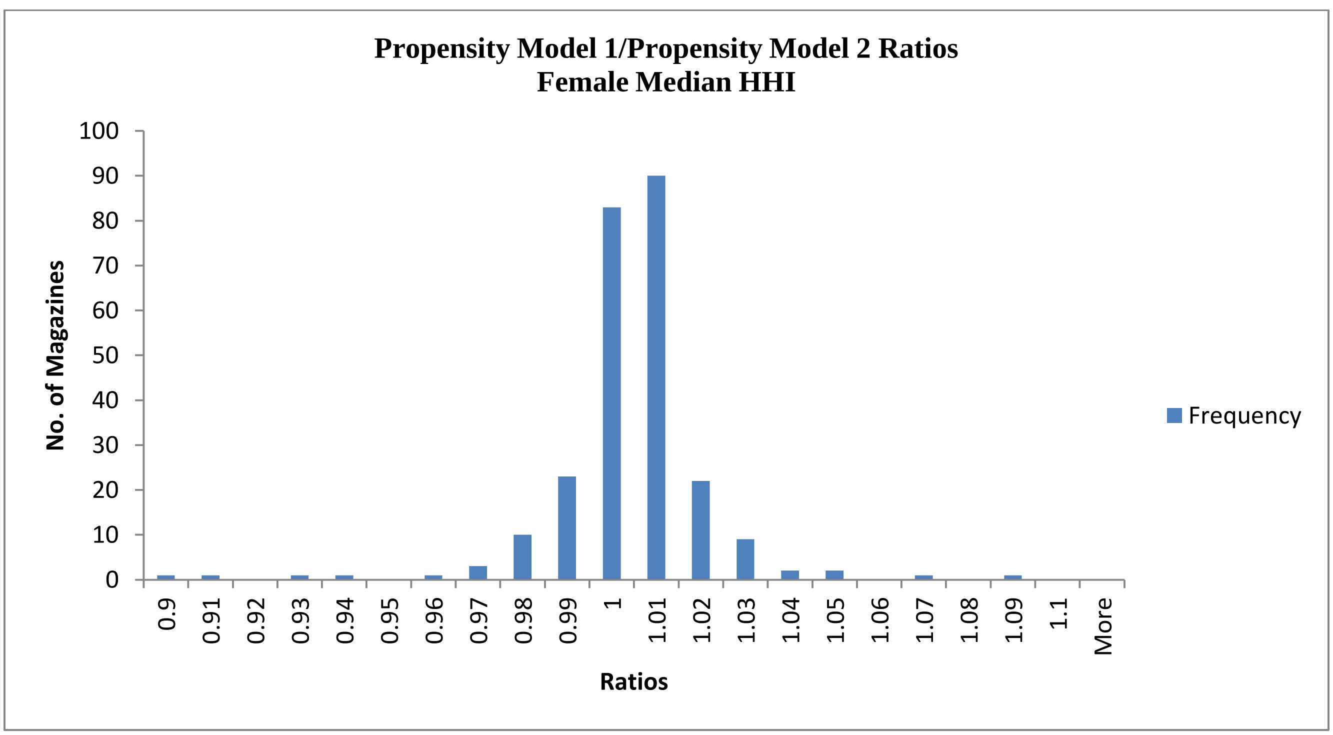


Chart 12



We conducted one additional analysis of the alternative non-response adjustment methods by comparing the root mean square error of estimators from Propensity Model 1with those from the Current Weighting adjustment. This term incorporates the variance of the estimator and its bias. For this analysis, we assumed estimators generated from Propensity Model 1were unbiased and the difference between respective publication audiences in the Current Weighting Scheme and Propensity Model 1 to be the bias term in the Current Weighting adjustment method. The results showed the root mean square errors for the 251 magazine audience estimators, respectively, were higher in the Current Weighting method in 51% of the cases and lower in 49% of the cases. In addition, audience estimates from Propensity Model 1 were generally more variable (over 60% of the cases) than the comparable estimates from the Current Weighting approach. This finding suggests that reducing biases by small margins is offset by the increased variance of the estimates.

Conclusions

Applying statistical adjustments to survey data, especially in a period of declining response rates, is clearly a valuable, available option to address the potential of non-response bias. Consistent with other research efforts that have used auxiliary data, we have appended a number of variables to the GfK MRI national sample to investigate their use in future studies. Unlike a number of studies, we have not shown any significant effects on the published audience estimates and profiles generated by our current non-response adjustment and post-stratification adjustments. There are a number of possible explanations for these findings:

- Our current procedures, which encompass a substantial number of variables correlated with readership in the post-stratification process, account for much of the non-response adjustments made by using the auxiliary data.
- The propensity model does not include other potential auxiliary data (e.g., the number of magazines subscribed to by eligible households) that are covariates of magazine reading.
- Even accepting the extremely modest differences in projected magazine audiences produced by incorporating these propensity models, the introduction of additional variance to these estimates essentially offsets any gains from the reduction in bias.

References

Biemer, Paul P., and Andy Peytchev. 2013. “Using Geocoded Census Data for Nonresponse Bias Correction: An Assessment.” *Journal of Survey Statistics and Methodology* 1:24-44.

Krueger, Brian S. and Brady T. West, 2014. “Assessing The Potential of Paradata and Other Auxiliary Data for Nonresponse Adjustments.” *Public Opinion Quarterly* 78:795 – 831.

Little, Roderick J. 1986. “Survey Nonresponse Adjustments in Estimates of Means.” *International Statistical Review* 54:139-57.

Politz, Alfred and Willard Simmons 1949.”An Attempt to Get the Not-At -Homes into the Sample Without Callbacks.” *Journal of the American Statistical Association* 44:9-32.