

YELLOW PAGES INDUSTRY DIRECTORY USAGE MEASUREMENT

WHAT SURVEY METHODS/AND SAMPLING FRAMES ENSURE RELIABLE AND ROBUST ESTIMATES?

ABSTRACT

Advertising Research Guidelines for Yellow Pages Usage Measurement were established in 2003 to “*Encourage and guide the creation of research comparable to other media that will provide the necessary information to produce ratings for: a) advertisers to support the purchase of Yellow Pages directories; and b) publishers and agencies to facilitate the sale of advertising in Yellow Pages directories.*” The ultimate goal is to achieve usage measurement systems that provide the Yellow Pages industry with overall and specific directory usage share that is credible and comparable to other mass media. The Yellow Pages Market Reporter (YPMR) syndicated service from Knowledge Networks/SRI constitutes the industry standard across almost 300 DDAs adhering to ARF guidelines. A predictive model approach has been newly introduced into the Yellow Pages research community by Directory Share Ratings (DSR) that requires scrutiny as to the adherence to ARF guidelines for audience measurement. A review of the prediction model methodology employed by Directory Share Ratings raises some important issues/questions about the reliability and robustness of their data.

SOUND METHODS FOR YELLOW PAGES AUDIENCE MEASUREMENT

The Yellow Pages Market Reporter (YPMR) syndicated service by Knowledge Networks/SRI is grounded in sound survey research methods and adheres stringently to the ARF guidelines for usage measurement. In fact, the lead researcher of KN/SRI was a member of the committee who helped establish the ARF guidelines. Estimates from YPMR constitute the industry standard in regard to quality. However, it is not evident from published materials by Directory Share Ratings¹ that their products are grounded in methods that meet minimum requirements for credible and accurate reporting of directory share results. The context of these statements is that sound survey methodology in any research arena requires the following components of quality be sufficiently addressed:

- **Projectability:** In a bona fide survey, the sample is scientifically chosen so that each person in the population has a measurable chance of selection – a known probability of selection. It is not selected haphazardly or only from persons who volunteer to participate. This way, the results can be reliably projected from the sample to the larger population with known levels of certainty/precision, i.e. standard errors and confidence intervals for survey estimates can be constructed.
- **Coverage:** (a) Ensure that the right population is indeed being sampled (to address the questions of interest); and (b) Locate (or "cover") all members of the population being studied so they have a chance to be sampled. The quality of the list of such members (the "sampling frame") whether it is up-to-date and complete is probably the dominant feature for ensuring adequate coverage of the desired population to be surveyed. Where a particular sample frame is suspected to provide incomplete or inadequate coverage of the population of interest, multiple frames should be used. (c) Ensure in multi-stage sample selection that each stage has adequate coverage.
- **Minimizing Bias due to Nonsampling Error :** Response rates need to be high, questionnaires accurate in terms of content and manner administered, and all other interviewer and fielding methods administered in ways to minimize biases due to nonsampling error.
- **Validation:** Results need to be validated using available benchmarks and replicated to ensure accuracy and consistency. This is especially true when new methodologies are being introduced into the research community.
- **Reliability:** Although sampling errors can be readily estimated using probability sampling methods, they do not reflect the total error of a survey statistic or estimate, which is a function of many different features of a given survey. Survey professionals practicing at their best carefully seek to balance these various types of error in the design and conduct of a particular survey, in order to minimize the total error given the budget or resources available.

¹DSR Website: www.directoryshareratings.com/

Table 1 summarizes adherence to components of survey research quality (and ARF standards for Audience Measurement as a by-product) by KN/SRI's YPMR methodology and DSR's prediction model estimates.

Table 1. Adherence to Components of Survey Research Quality

Standard	KN – YPMR	DSR
Projectability	Yes	No
Coverage	Yes	No
Minimizing Bias due to Nonsampling Error	Yes	No
Validation	Yes	No
Reliability	Yes	No

In the sections below, we will assess both the YPMR and Directory Share Ratings methods and describe in more detail the success and/or limitations of the DSR prediction model approach in each of the critical requirements listed in Table 1.

A. PROJECTABILITY

The KN/SRI-YPMR methodology employs a probability-designed random digit dial (RDD) sample (where random 4-digit numbers are used within residential telephone exchanges) sample with results projectable to the adult population of the DDA within known statistical tolerances. Basically, sample exists with known selection probabilities for 290 individual DDAs. Thus estimates are projectable at the DDA-geographic levels with known estimates of sampling error.

Conversely, the DSR methodology does not include sample in all DDAs that are reported. This is a serious limitation. The zip codes that are selected (less than 1% of all zip codes – 200/31,000) result in NO sample in a large number of DDAs, i.e. 6,300 DDAs have no sample. Therefore, projection for those DDAs are based on demographics and/or similar configurations in other DDAs, but they are not truly projectable since sampling theory requires sample to actually exist in the reported geographic area.

Selecting one or two zip codes to represent consumer behavior throughout an entire DDA is problematic as directory shares often vary across different geographies within a book's scope. As an illustration, in Book "A"'s Big City DDA which has 89 zip codes, the directory share for the Book "A" Big City directory varies from a high of 56% in the downtown DDA to 32% in the suburban DDA.

Another flaw in the projectability of the DSR estimates is the fact that the questionnaire administered appears to ask the required questions only of the person answering the telephone. There appears to be no attempt at a random selection of one person to interview among all persons in the household which is a normal sampling procedure to ensure randomness and projectability across all persons. Otherwise, the respondents will skew female, non-minority and older which will then skew the share estimates.

KN/SRI-YPMR's method enumerates all adult male/female residents in the contacted household and randomly selects one of these persons to interview. Hence, all adults in the household have known selection probabilities.

B. COVERAGE

Sample for the KN/SRI-YPMR is selected via standard probability based RDD methods with approximately 92% coverage of all residential households – only non-telephone households are missing from the RDD sample frame. The approximate 8% of households not covered in RDD sampling is comprised of households without landline phones (2%) or cell phones only households (6%). Those households without any phone line may be less pertinent to directory share estimates and cell phone only households represent a small coverage issue (6%) and are unlikely to introduce bias into share estimates due to the small contribution to the total.

The DSR sample is selected from listed residential phone numbers only. Hence, the sample excludes unlisted numbers either because of a recent move or a request that the number be unlisted. As a result, according to information from Survey Sampling, Inc., one of every three (33%) households in the U.S. is unlisted. This unlisted rate is notably higher in more urban DDA's like the Bronx at 54%, Philadelphia at 50%, Chicago at 48%, Seattle at 41% and Atlanta at 39%. Coverage error in the DSR sample is quite high, which can lead to biased directory share results.

C. MINIMIZING BIAS DUE TO NONSAMPLING ERROR

Reducing bias due to nonsampling error is critical. A study can be based on perfect sample in terms of efficiency and coverage, but produce inaccurate results if fielding methods are not random and implemented well.

In the DSR methodology, the questionnaire asks the respondent to identify the spine code of the directory last used, it does not mention a procedure for handling cases where the respondent (a) cannot find or see the spine code (which is non-random event), (b) provides a non-existing number, (c) provides an out of area book number or (d) provides the means for correcting a transposed or incomplete number. The KN/SRI-YPMR method addresses these issues in the questionnaire and interviewer/respondent interaction. Also, in the DSR methodology, there are no additional questions to help classify the book actually used (e.g., directory title, town or area covered, pub month, pub year or publisher). The DSR data will have more item nonresponse, larger response errors and dirty data.

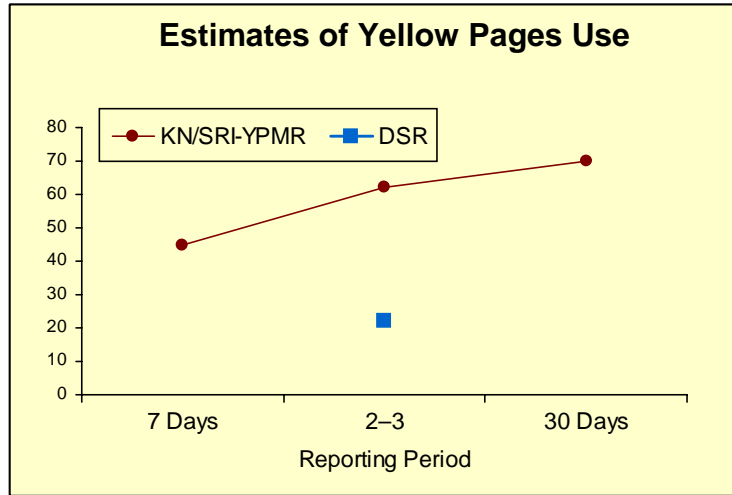
The specific search item of the last reference in the DSR methodology is not specified. As a result, the directory brought to the phone is more likely to be the “most available” book versus the directory actually used.

D. VALIDATION

When a potentially new methodology is being introduced, it is imperative that the results be compared to benchmark data to the extent possible and data evaluated for consistency from one fielding to the next. Sufficient data exists from the KN/SRI-YPMR products for use to evaluate the DSR methodology.

For example, DSR reported 24% of respondents had used a Yellow Pages Telephone directory within the past 2–3 weeks (which in and of itself is a poorly worded reference period question). Conversely, the KN/SRI-YPMR question asks the respondent when they last referred to Yellow Pages of any telephone directory either at home, at work or elsewhere. KN/SRI-YPMR estimates

of usage are higher than the DSR estimate – 49% and 71% respectively for past 7 day and past 30 day use. Just applying a simple linear interpolation to the 7 day and 30 day reporting to get to a comparable 2–3 week reference period suggests real usage at about 63%, not 24%. That is a gross understatement of Yellow Pages use from DSR by approximately almost 40 percentage points. The chart below visually illustrates the comparison:



E. RELIABILITY

In order to calculate estimates of reliability (standard errors and confidence intervals), the study sample needs to be selected using probability methods as in the KN/SRI-YPMR approach. As described above, for a large number of DDAs in the DSR approach, sample simply may not exist or sample sizes are less than 30, due to zip codes with no survey coverage, which obviates calculation of reliability estimates. Due to the non-probability sampling approach for DDA level estimates, one cannot legitimately calculate a standard error for the DSR directory shares nor can one assess the statistical significance of differences from one directory share to another. Consequently, estimates or projections of usage at the DDA level are inappropriate.

CONCLUDING REMARKS

Balancing quality with cost is the biggest challenge in media research. However, the driving goal should be to first identify the desired reliability for key outcomes of the study and build the study design and fielding around those goals. To this end, the KN/SRI-YPMR methodology can be held as a gold standard for estimating directory share ratings, adhering stringently to standard survey research methods and to ARF guidelines. Serious design flaws exist in the DSR methodology – results are not statistically projectable, reliability is not measurable, biases in the fielding are not measured/controlled, and results do not match up with industry standard benchmarks.