

## **SAMPLING CONSIDERATIONS WITH SINGLE SOURCE DATA**

**Timothy Joyce and Steve Coffey, Nielsen Household Services, U.S.A.**

---

### **Introduction**

The Nielsen Household Panel was established in the U.S. in its existing electronic form early in 1989. Its members record, by in-home scanning, their purchases of barcoded products which are brought home. (A similar panel using the same equipment has recently been established in the U.K. and in Germany, also by Nielsen.)

Early in 1991, it was decided to make a concerted effort to bring multi-media measurement fully into the U.S. Panel. This was to support custom analysis applications, and also a new media/product syndicated service, which was duly launched later that year, called HOME\*SCAN. These media measurements comprised both a "Reading Poll" for newspapers and magazines, and also questionnaires for the measurement of other media -- the latter obviously being beyond the scope of this particular paper.

In this paper, we describe how the panel is recruited, and discuss questions which have legitimately been raised about the sampling procedure employed. We describe how they have been addressed by a special research project carried out in the last few months, which as a by-product appears to be of much interest as an execution of a candidate for a "Silver Standard" measurement of magazine total audiences.

But first, we shall describe how the basic product purchasing data, and the print audience data, are collected.

### **Product Purchasing Data**

The Household Panel was set up by Coffey, working for what is now known as Nielsen Household Services. It comprises 40,000 U.S. households. Each of these households has an in-home scanner, which any household member can use to record purchases of barcoded items. Recruitment and briefing, as will emerge, are carried out by mail, but there is also continuing interaction with a much used telephone "Help Desk."

The scanner, not much larger than a television remote control, is a powerful microcomputer. It has a small screen which prompts with questions, an alphanumeric keypad, a barcode reader, and a one-way modem speaker by which the data can be transmitted over a regular telephone to the central computer.

After a shopping trip in which any barcoded item has been purchased, the purchases are recorded. These extend far beyond supermarket and drug store items; almost all packaged goods, or FMCG's, are now barcoded.

The scanner asks, for each trip, the store name, or type; the identity of the primary shopper, by name, on that trip and who else -- if anyone -- "went along"; numbers of each item purchased, after the barcodes are scanned; prices paid (in most cases "not needed", because knowing the store, Nielsen receives price files on a daily basis from each chain); and coupons and other deals used.

Once a week, or after completing a survey (another function of the scanner), the panelist is asked to transmit data. It is estimated that the task of scanning adds very little time, each week, to putting away the shopping: Some ten minutes, on average. Panelists are encouraged to continue to participate and to transmit each week by a variety of methods, including a monthly newsletter, sweepstakes, and a "points" program much like a "frequent flyer" scheme with the points redeemable for gifts from a catalog.

These basic marketing data are used by most major packaged goods manufacturers in the U.S. to help manage their businesses. Applications include knowing the demographics of category and brand buyers; understanding brand switching and brand interactions; knowing trial and repeat for new products; evaluating promotions; and much besides. Potential applications concerned with advertising exposure led to the drive to incorporate media measurement -- and we turn now to the relatively new (for the panel) print measurement procedure.

## Print Measurement

Simply put, it is the two-step, logo-card sorting, recent reading procedure which has been employed by Mediamark Research Inc. (MRI) in the U.S. for fourteen years, adapted to self-administration, and to data entry into the scanner, by the device of printing barcodes (assigned for this purpose by Nielsen) below the logos on the cards to be sorted and scanned.

As a footnote to history, this card sorting procedure was devised by Joyce as a candidate for the British National Readership Survey in 1978, and was first piloted for him in London in that year. This British initiative was overtaken by events -- it was adopted in the U.S. by MRI from its inception in 1979.

Each adult in each panel household, in turn, is asked to go through the "Reading Poll", using the scanner to record the data. The first sort of a shuffled deck of cards, is a "6-month screen". After scanning the cards (and recording frequency of reading), those screened-in are separated by publication frequency (weeklies, monthlies, etc.), and are re-sorted to establish recency of reading (the last 7 days, 30 days, etc.). A final scan of the "yes" position cards for the recency sort completes the process.

A word about newspapers. It was found, in pretesting, that the self-administered "interview" flowed much more smoothly if they were handled along with magazines. So the logo deck (actually with printed titles rather than logos for newspapers) includes together with some 150 magazines, the national newspapers and the local newspapers circulating in the household's neighborhood -- all computer generated and laser printed -- no mean feat.

Pretesting of the procedure was, of course, most carefully done, and while in the field, further comprehension checks have been made which assure us that the "poll" is being responded to properly. This work is available to any interested party. Cooperation rates among the panel, incidentally, are very high: generally in excess of 90 percent.

## Product/Reading Data

Before turning to the sampling issue, we should say a little about the applications of the product/reading data. Besides custom applications, we have set up, as noted above, the HOME\*SCAN syndicated data base. This aggregates all purchasing data for each individual adult in the panel households over 12-month period, and links these to the corresponding Reading Poll period (and to other media and demographic data as well).

The strength of HOME\*SCAN is, of course, the precision and detail of the purchasing data. We know, for each adult (male or female, by the way) whether s/he purchased packaged detergent as a category, and Tide as a brand, in the last 12 months; the volume purchased, in ounces; the dollars spent (and therefore the average price paid); and for the category, the dollars spent where a coupon was used.

We don't have to rely on data, supplied only by the homemaker, on numbers of washloads in the household in the last seven days. We think the differences in precision and detail speak for themselves.

Of course, the data can be aggregated in all kinds of ways, such as by manufacturer and by store. We can tell *Family Circle* magazine how many of their readers bought Tide; how much they bought and how much they spent; and how much they spent on Procter & Gamble products in total, or in Safeway stores in total.

A vast step forward in data for media planning, buying and selling for packaged goods -- but is it?

What about the sampling issue?

## Sampling Procedure

Now we turn to how the panel is recruited. As a preface to this, we should remember the original and still by far the most important purpose of the panel -- which is to supply high quality longitudinal purchasing data. We want, for example, to look at brand switching over long periods -- even from year to year. This necessitates a panel of people who are willing to perform a task which, while not difficult or time-consuming, must be performed with some care, week in and week out for a long time -- in fact, people prepared to make a commitment.

In practical terms, this sort of panel can't be established in classic media research terms as a high quality, high response rate, strict probability sample. If we were to try, we would find that the higher we drove the initial response rate, the greater would be the rate of drop out, to a quite unacceptable degree. We stand to be corrected, but we are not aware of a single counter example to this in the world.

But nevertheless, we want a representative, and in all key respects, a valid sample. The classic solution to this, and the one used here, is quota sampling.

The panel is recruited by direct mail (within a sample of some 3,000 areas, which is a strict area probability sample, at that stage). Those expressing interest are instructed in some detail as to what they are being asked to "commit" to do. Those still willing to be panelists are then signed up on the basis of filling a very detailed set of quotas defined in terms of demographics and geography.

If a panel household ceases to transmit data on a regular basis, or fails to pass reasonable data quality control tests, it is dropped and instructed to return its scanner. In fact, up to some 25 percent of the panel drop out each year, and are replaced -- but this is a satisfactorily low proportion.

What is known about the validity of the panel, in terms of the data which it supplies? So far as the marketing data are concerned, the proof of the pudding is in the eating: manufacturers who use the data find they agree quite adequately with "known" sales data at any one time, and also track well with them over time: otherwise they obviously wouldn't be users. This has been studied in general terms too (again, we'll be happy to supply the details to any interested party). Comparing like with like for hundreds of categories, i.e., purchases from supermarkets, the Household Panel accounts on the average for 85 percent of sales as recorded by Nielsen Scantrack store scanner data. Some of the balance is institutional purchases; some is individual purchases not brought home; the desired true coverage (i.e. household purchases brought home) is clearly very high. And the correlation across these hundreds of categories of total dollar sales is 0.99. We are not aware of any serious questions being raised about the use of quota sampling to recruit and maintain the panel from the point of view of its marketing data. But the introduction of media measurement brings new constituencies and interests into the mix.

In Europe, few eyebrows would be raised when faced with quota samples for product/media purposes. Many such surveys employ them, a good example being the British Target Group Index (TGI) -- Joyce's original foray (1968) into product/media research for media planning/buying/selling purposes. It employs quota sampling.

In the U.S., the classic media research position has tended to demand strict probability sampling. In our view, if a high response rate for the purpose concerned can be achieved this is still correct. It is achieved (at high cost) by single-shot surveys such as those conducted by MRI and SMRB.

But when respondent cooperation is required to be both careful and long-term, it strikes us that a fixation on probability sampling can become unreasonable. To take an example within Nielsen: the NTI national peplemeter panel has a response rate among basic households (i.e., not counting substitutes), in terms of those recruited, installed and in-tab at any one time, of only 35 percent (source: CONTAM). Yet the industry seems to have made it clear that they prefer this to what would be closer to a quota method which might have other advantages.

Enough of theoretical discussion; what are the practical questions so far, specifically, as magazine measurement is concerned -- and can they be addressed?

### The HOME\*SCAN Levels

The publication of the first HOME\*SCAN total audience estimates, for 145 magazines, naturally put the discussion in a more practical frame. What emerged was that the HOME\*SCAN estimates are even higher than those of MRI, which of course are themselves well known to be systematically higher than SMRB's, especially for monthly magazines. Currently (Fall 1992 Release) the HOME\*SCAN levels are, in the aggregate, 23 percent higher than MRI's.

As just one example of the sizeable differences between the services in the U.S., here are the estimates from their most current releases of the adult audience for the *Reader's Digest*

	Millions
HOME*SCAN	65.6
MRI	50.9
SMRB	39.8

The correlation between HOME\*SCAN and MRI levels is high -- 0.97. Some of the larger differences are systematic, as for these titles where our estimates are much higher than MRI's:

**Skin**  
Playboy  
Penthouse

**Tabloid**  
National Enquirer  
Star

**Popular Culture**  
Us  
Soap Opera Digest

We ourselves believe that it is possible that this is a result of the relative anonymity of the HOME\*SCAN procedure, not disadvantaging titles which some true readers might not want to acknowledge to an in-home interviewer that they read.

But even removing them, a substantial aggregate difference remains. It could be argued in favor of HOME\*SCAN, and we shall indeed argue it, that it has the advantage that the respondent can complete the Reading Poll in her/his own time, without feeling "rushed" by an interviewer or by circumstances; that s/he can read and re-read the instructions until s/he is confident they are fully understood; and further that the rather long battery of "reader quality" measures, present in MRI for good reasons, are absent from HOME\*SCAN because they are beyond its scope. Each of these factors could operate to get a better count of total audience than a lengthy personal interview can.

The counter argument is, of course, that HOME\*SCAN may be producing higher reading estimates as a consequence of a literacy bias in the panel.

We have always acknowledged that there may be such a bias. It is necessary to be literate to function in the panel. The most difficult households to recruit and retain are certainly less well educated, and despite filling of quotas and subsequent sample balancing, it could quite well be that it is the brighter of these groups that are actually recruited to the panel and remain in long enough to qualify for the HOME\*SCAN data base (remembering the 12-month purchasing data requirement).

With this in mind we resolved to quantify the "literacy bias" -- if there is one.

### Telephone Survey

An obvious step was to conduct a telephone survey among two samples, one a subset of the Household Panel (drawn from the panel, weighted to simulate actual HOME\*SCAN sample balancing), and one a strict probability sample of the U.S., drawn by random digit dialling (RDD) methods. The questionnaire would establish levels of magazine reading on some basis, to compare Panel and RDD overall responses. The research would be double-blind, in that interviewers would be given a "meshed" list of telephone numbers to call, not knowing that it in fact comprised two lists, which would subsequently be separated on tabulation based on ID numbers -- and Panel members would have no idea that their interviews had anything to do with Nielsen.

On further consideration, it occurred to us that the research could also have a somewhat bold second objective: namely to test a candidate "Silver Standard" first read yesterday (FRY) approach, to yield data which could be directly compared with published HOME\*SCAN and, indeed, any other service's levels.

For this purpose, we took as a model a survey to establish levels of reading days which was designed by Joyce, and described by him at the New Orleans symposium in 1981. This asked respondents to volunteer the newspapers and magazines they had read or looked into yesterday, unaided, with prompting by time of day and places visited out of home. For each publication for which reading was claimed, the number of issues read yesterday was established (one only, of course, in 90 percent or more of the cases).

To this line of questioning we added the "reading days" question which proved, in the ARF Gold Standard Work as validated by observational tests, to "work" very well in terms of establishing first read yesterday -- FRY -- with accuracy. That is, for each publication issue read yesterday, we asked on how many different days (not counting today) that particular issue of that publication had been read. The response "one" meant, of course, "FRY" -- first read yesterday.

This is a necessarily very summary description of a survey approach concerning which, again, full details are available to interested parties.

1,020 interviews were made with our panel members, and a further 1020 interviews with an RDD sample, each at the rate of exactly 17 a day over two 30-day periods during November 1992/February 1993, avoiding the holiday season. Fieldwork was conducted by Westat, the parent of Crossley Surveys which was our contractor. They achieved the highly respectable response rates of 78.5% for panel members and 70.0% for the RDD sample. One adult was randomly selected for interview in each household, with no substitutions allowed.

## Survey Results

In the table of results below, we show data for the 145 magazines measured by HOME\*SCAN, in terms of their FRY levels from three sources.

First, the FRY levels inferred from HOME\*SCAN itself. These were, of course, derived by aggregating the total audience estimates by publication frequency, and then dividing the number of days in the publication interval, e.g., 7 days for weeklies, 30 days for monthlies, and so forth.

Second, FRY levels projected directly from the subsample drawn from the (weighted) Nielsen Household Panel.

Third, FRY levels projected directly from the RDD sample.

In the table, some very broad types of publications are broken out in addition to the total. As to frequency, "Weeklies" includes the handful of biweeklies. "Monthlies" includes triweeklies and bimonthlies. As to sex orientation, a (HOME\*SCAN) audience composition of 60 percent or more qualifies a publication for male/female orientation, as the case may be. As to circulation size, "large" means 3 million or more. Here are the results:

	FRY LEVELS					
	Inferred from HOME*SCAN(F'92)		Panel Interviews		RDD Interviews	
	mm	%	mm	%	mm	%
All Titles(145)	71	100	71	100	67	100
Frequency						
Weeklies etc.	40	56	36	51	37	54
Monthlies etc.	31	44	35	49	31	46
Sex Orientation						
Dual	19	26	22	31	17	25
Male	19	27	18	25	24	35
Female	33	46	31	43	27	40
Circulation						
Large	36	51	36	52	35	52
Small	35	49	34	48	32	48
(Unweighted count)			(354)		(346)	

One further technical word about the table: to facilitate calculations of sampling variances, we show at the foot of the columns of percentages for the Panel telephone interviews, and for the RDD telephone interviews, the unweighted FRY aggregate counts (out of 1,020 interviews in each case). It should be kept in mind, however, that the samples are not simple random samples (SRS) -- they involved some clustering and some weighting (especially by numbers of adults in each household, since only one was selected for interview) -- and also that the variances are not actually binomial, since a respondent could FRY more than one magazine/issue. An effective sample base (ESB) of some 250 in each case would be nearer the mark. Which is to say that sampling variances, as with all FRY procedures, are unfortunately LARGE.

The main comparisons which the survey was designed to make possible emerge from the top row of the table. From HOME\*SCAN, one would infer an aggregate FRY audience across all measured titles of 71 million. This happens to agree exactly with what was found as a direct FRY measurement by the telephone survey with panel members (after weighting to make it directly comparable with HOME\*SCAN, which is projected to U.S. Census demographics). This indicates that the HOME\*SCAN recent reading procedure yields the same levels as the telephone FRY procedure which was used here.

The second comparison is between the 71 million aggregate from the panel telephone interviews, and the 67 million aggregate from the RDD telephone interviews. The panel aggregate is 5 percent higher than the RDD aggregate. This is, of course, in the direction we would expect if there is a literacy bias, causing more magazine reading, in the panel; but the difference is not large, and in the case of this survey, it is not significant. By contrast, the comparable aggregate FRY estimate inferred from MRI for the 145 titles is 57 million, which is 15 percent below the RDD level, a significant difference. Some of the 145 titles are not measured by SMRB, but for the titles in common, the aggregate FRY estimate inferred from SMRB is 28 percent below the RDD level.

To summarize the differences between the services and the RDD level, indexing the latter at 100, for common titles:

HOME*SCAN	105
RDD	100
MRI	85
SMRB	72

Looking at the percentage compositions of aggregate FRY by type of magazine, comparing the panel sample with the RDD sample, none of the differences is statistically significant with the exception of aggregate audiences for male-oriented books, where RDD is significantly higher.

In summary, we think we are entitled to conclude that while the quota sample on which HOME\*SCAN is based may involve a literacy bias, it is slight. The RDD "Silver Standard" does, in fact, yield total audience estimates in the aggregate which are closer to HOME\*SCAN than to any other service.

## Conclusion -- A "Silver Standard"?

For us, perhaps the most interesting aspect of the telephone survey was the opportunity to experiment with a candidate "Silver Standard" approach to establishing total audience levels, and we conclude with a short discussion of this.

Over some years in the U.S., the Advertising Research Foundation has developed and validated a "Gold Standard" approach, using actual magazine issues in personal interviews, with a FRY procedure. The validation has been by observation, or a "Platinum Standard", in test situations some of which were most ingeniously contrived to detect errors. The "Gold Standard" passed these tests -- over all these tests on the average, exhibiting 1 percent overclaiming and 6 percent underclaiming.

Unfortunately, in practical terms the Gold Standard would be extraordinarily expensive to execute for real-life audience measurement. There are several reasons for this. First, a given interview can in practice only cover two, three or at most four titles (with multiple issues of each). Second, the samples used would have to be very large, especially for monthlies, given the need to multiply the FRY estimates by 30. Third, the sample would have to be a high quality high response rate national probability sample, very carefully controlled: equal numbers of interviews would have to be completed each day as nearly as possible. Test issues would have to enter the interviewers' kits immediately on publication in order not to miss the earliest readers -- and stay in as long as they were still accumulating new first-time readers.

While some might dispute this, we believe it would cost tens of millions of dollars to get adequately precise estimates for just a handful of major titles.

This is not to say that the "Gold Standard" work did not have a lot of value as development work. Specifically, the same (or very similar) observational tests could be used to evaluate, improve upon, and (one might hope) in due course validate a "Silver Standard" approach to be executed by more economical means, with many more titles covered per interview.

Telephone research has the advantage that high response rates can be obtained from dispersed samples at reasonable cost, and that the number of interviews each day can be strictly controlled. Further, a large number of titles can be studied per interview.

What we have executed is just one candidate "Silver Standard". Reading claims were unaided by magazine title. It has the advantage that the interview is brief and straightforward, and that all magazines are measured.

It is easy to devise alternatives. An obvious one would be to ask, concerning a rotated roster of titles, when each one was last read (not counting today). A "yesterday" claim would be followed with questions about number of issues read yesterday and reading days for each issue, just as we did with our method to get FRY.

These and perhaps other alternatives could be validated (or improved upon), just as the Gold Standard personal interview was. We hope this paper will stimulate further work on these lines.

