HOW SHOULD NATIONAL TV AUDIENCES BE MEASURED?

Nielsen has had a number of issues recently, including problems maintaining contact with its national peoplemeter panel, which caused some audience understatement. This led to its loss of accreditation from the MRC. Nielsen has also been faulted for its tardiness in finding a way to measure the audiences of streaming services, as well as viewing using digital devices. Sample size has been another issue; many question whether a 42,000-home panel is large enough to deal with an increasingly fragmented TV/video world.

In response, Nielsen has developed a totally new "big data" service" that will merge its existing peoplemeter panel with device usage data from millions of ACR-enabled "smart set" and set-top box panels to produce more "granular" program and commercial "viewing" estimates for virtually every TV/video platform or network, including those with very small audiences.

The prospects of a much larger, "census-style" panel is intriguing to those who hope to refine TV targeting capabilities and move away from sex/age-based audience metrics like adults 18-49 or women 25-54, which have been in use for more than 50 years and are still used today as TV's primary buying and selling "currency." Others also hope that having a huge sample and linking its audience findings to "outcomes" will lead them to the holy grail of "attribution"—in effect, telling them which commercial placements on specific platforms or individual channels generate the best return-on-investment (ROI) for advertisers. Finally, there are those who believe audiences across platforms, networks and channels need to be "deduplicated," so that advertisers will know the reach and frequency of their cross-platform buys.

While hopes are very high for a totally new national TV rating service, and there is growing sentiment favoring the "big data" approach, there are problems as well.

PROBLEMS WITH "BIG DATA" RATINGS

One of the ongoing concerns about Nielsen's peoplemeter panel—aside from its relatively small size—has been whether it is truly representative of the total population. According to some estimates, only 30% of the homes that are invited to participate in the panel agree to do so. While statistical "sample balancing" techniques reweight the panel so that, in aggregate, it resembles the total TV home population by sex, age, race, household size, region, etc., this does not mean that you have a perfect panel. The people who agreed to participate may not be representative of all people with similar demographics.

A case in point concerns the amount of TV that panelists consume. Researchers have long known that if the purpose of a survey—whether it's about TV viewing, cooking, or travel plans—is obvious to the would-be respondent at the outset, those who rarely or never participate in the activity will tend to opt out to a greater degree than their opposites, who are frequent participants.

When a TV home is asked to join a panel where all its TV sets are monitored, it's pretty clear to the household head what the purpose of the panel is. Does this create a heavy viewing bias in a TV rating panel? The only evidence we have seen on the subject comes from a decades-old study by Bill Harvey, then at Hooper. Here, a sample of adults was interviewed and asked how much TV they normally viewed. Then they were invited to join a meter panel that would monitor the activity on their TV sets. Those who

agreed to participate were 10-15% more likely to be heavy TV viewers than those who declined. We suspect that, directionally, this bias is still true today.

There are other questions regarding the "big data" approach. For example, who has vetted these panels? Are the homes selected on a random probability basis or do they take any home that agrees to have its sets monitored? If so, what's their opt-in rate? If it's less than 30% how do we know that they are representative of all ACR homes? How often are these homes checked to account for newly acquired ACR sets or those that break or are discarded? What about "dumb" sets which many people still use, especially in non-ACR homes? To what extent are these panels turned over to "refresh" them when a family moves or otherwise drops out and replacement homes are recruited? How do the TV set usage findings compare from one ACR panel to the other? If they differ significantly how do we reconcile such variations as they are all claiming to represent all ACR homes?

Similar questions apply to set-top-box panels. Here, we are using set usage data from homes that have stuck with "pay TV" while many others have cancelled their subscriptions. Are the two population groups very similar or are those who still have STBs more likely to be older, heavy viewing, adults? Does sample balancing effectively deal with this issue?

What about obtaining viewer data? As we have noted, it appears that Nielsen's new "big data" service will meld together the findings from various ACR and STB panels—involving many millions of households—in order to obtain large sample bases for individual TV show episodes as well as commercials. To produce viewer estimates, viewer-per-set factors will be derived from the small people meter panel, probably using about 25,000 of its homes where such information is gleaned. These will, no doubt, be matched show by show and commercial by commercial with the big data set usage results to produce program and commercial "viewing" estimates. Which means that the sample sizes for the set usage element will be huge—hence, in theory, very stable—but not so the viewer projections, especially if sliced and diced by sex, age, income, etc.

It has been suggested that this problem might be dealt with, at least in part, by using the known characteristics of the residents of each "big data" household. Sorry, but that doesn't fly, since you can't assume that a 35-year-old woman who lives in one of these households was the viewer whenever the set was in use. Attempting to model such behavior run into difficulties because homes with adults aged 25-54 usually have more residents (children and adults) as well as more sets, which are used, collectively, compared to older households where there are often only one or two residents and fewer sets. As a result, ACR set usage tallies may suggest that 25-54-year-old adults, particularly those with above average incomes, are viewing many shows when, in fact someone else in their home was the actual viewer. In like manner, viewing by older adults would be understated as their sets are not in use as often. Yet viewer studies paint quite a different picture: older adults are by far TV's most frequent viewers while younger/upscale adults are usually much less likely to be watching.

In short, what you get with "big data" panels is a far larger sample size but not necessarily the same degree of stability for viewer projections which will continue to be based on very small panels.

This brings us to the matter of including digital devices to capture viewing. Will every device in the big data homes be monitored or will only a sub-sample of such households be used for this purpose? Perhaps

additional panels of mobile device users, for example, will be melded into the mix. But this raises many of the same questions we have already asked.

Last but hardly least, will users of a "big data" rating service be getting anything approaching an idea of who is actually watching—eyes on screen—any portion of program content or any commercial, as well as how much dwell time they devoted to said content. Here, the answer is clear: no, they won't be getting such vitally important information. All they will know is that program or commercial content appeared on a TV screen. That's all.

The fact is that none of the TV rating services to date have intended to measure viewing behavior on a minute-by-minute basis, let alone second-by-second. Even when peoplemeters were introduced by Nielsen in 1987, panel members were merely asked to indicate whether they were "watching" the program whenever a channel was selected. They were also supposed to notify the system any time they left the room or otherwise stopped watching, but most panel members didn't bother to record such non-viewing situations. Failing such notifications, the peoplemeter system simply assumed that, once identified as a "viewer," the panelist watched every second of content—including commercials—as long as the channel wasn't changed.

ENTER ATTENTIVENESS

As far back as the 1950s, there were clear indications that many viewers didn't watch every commercial that appeared on their TV screens. At the time, some major-market newspapers, feeling threatened by television, obtained minute-by-minute usage readings from their municipal water boards that showed advertisers that water usage spiked when commercials were aired. The implication was that program viewers used commercial breaks to get a drink or visit the bathroom. Later, sophisticated research that involved the use of cameras or teenage/college students who "spied" on their parents while they watched TV shows, confirmed the reality of commercial avoidance.

More recently, TVision's panel of 5,000 homes has utilized electronic "eye cameras" to monitor program viewer behavior when commercials appear on their TV screens. Typically, of those who were in the room just prior to a commercial break, 30% absent themselves when the break begins, while another 30% who remain in the room aren't looking at the screen. This means only 40% of the program audience watches an average ad message. Those who do watch see only about half of the average commercial's content.

Also obvious from this body of research is the fact that attentiveness is not a constant. There are significant variations based on the skill of the ad message in capturing and holding a viewer's attention; what's being advertised; the demographics of the audience; the nature of the programming; the degree of ad clutter in the break; commercial length; how recently the same message was seen; and so on.

Yet attentiveness is <u>not</u> part of the design of any of the existing national TV rating surveys, nor is it part of any of the proposed alternatives, including the "big data" panels currently in favor. Just as in the past, it will be assumed that if the TV set or digital device presents commercial content on its screen, the "viewer" was "reached" by the message.

THE IDEAL NATIONAL TV RATING SERVICE

So, what should the ideal national TV rating service be able to do?

Attentiveness A Must

First and foremost, we think it should determine exactly what proportions of program and commercial content were viewed—and we mean actually looked at—by those whose TV/video device usage is being monitored. And the only way to get at that is by incorporating an attentiveness measurement, probably using a method much like TVision's for every set (including "dumb" sets) in every household. A panel would be required to do this; you can't measure attentiveness using set top box data.

Panel Size

We also think that a much larger sample than Nielsen's current 42,000-home peoplemeter panel is required to provide stable attentive audience measurements. This is because so much of the "viewing," particularly for commercials, will prove to be inattentive. We estimate that a panel of about 200,000 homes for in-home TV usage is probably mandated.

But how do we account for digital devices, which currently represent about 15% of total "TV" viewing? This type of activity will continue to grow, so addressing this issue is a must.

A solution might be to track every digital device anyone in the TV set panel uses. Most devices have built-in cameras, so this would be the most likely method for observing digital viewer behavior, but this will require panelists to give their consent. In addition, will the cameras record images that can satisfactorily confirm attentive viewing? As we understand it, work is being done on these issues for smartphones, so there is no reason that, pending their resolution, this type of monitoring can't also be applied to tablets, laptops, and PCs.

Cost

A 200,000-home panel, with all devices monitored constantly to determine who is watching any portion of the content appearing on the screens, is going to be very costly.

At present, it's estimated that Nielsen gets about 75-80% of its national TV revenue from TV networks, cable channels and syndicators, while ad agencies account for most of the remainder. Nielsen typically charges its time selling clients between .5-1.5% of their ad revenue for its service. What if the cost of the ideal system we have described doubles or even triples the current expense? Will Nielsen's customers be willing to absorb such increases in their research budgets?

The answer is probably no, which means that a less-than-ideal compromise solution will be needed. This will most likely necessitate the creation of a split panel, with the larger proportion—perhaps 75% of the total—measured only by device usage, while the rest measures both device usage and attentiveness. The latter's attentiveness findings would then be statistically merged with the larger device usage-only panel data.

Low-rated Programs Still A Problem

Even with 200,000 homes supplying data, there will still be very low rated programs, channels, and platforms. But it is simply unrealistic to double or triple the size of the panel just to provide for this contingency; the small fry can't afford it, and the larger platforms, who don't need the added sample, won't pay for it.

Because of the way most TV is sold, this is not as serious a problem as some think. Typically, a seller guarantees audience tonnage—GRPs or "impressions"—for an entire schedule of commercial placements that involve many shows and many episodes, not each placement individually. Therefore, even if the average attentive commercial viewer projection for a very low rated program is based on only 50 panel members, over the course of the entire schedule in that venue, 500 or more panel members may have watched one or more of a brand's commercials on a number of occasions. This means the total GRP or impression projections for a low rated seller are more statistically reliable than those for any individual spot in the schedule.

WHO SHOULD PAY?

This brings us to our final point. Why should the only function of the ideal national TV rating service be to supply an improved audience currency for time buying and selling? Aren't there other uses for such a service, and those, aside from buyers and sellers, who could benefit? And, if so, why must the time sellers bear the majority of the costs?

There are two uses of the data that we think are of particularly great value to other parties. The first—insights that advertisers could glean about how their ads and those of the competitors work—is the most important. For the first time, advertisers would be able to determine the true reach and frequency patterns for their ad campaigns, based on who watched their commercials and for how long per exposure. Similar findings for rival brands could be tracked and deduplicated; one brand could see how many consumers who watched its messages also viewed those of a competitor. Commercial wearout could also be tracked based on declining viewer dwell times. And wouldn't it be helpful when casting decisions are being made, to see how an actor or celebrity spokesperson under consideration scored in attentiveness in prior ads? What about the "internals" of a brand's commercial performance; how well does the opening scene of an ad message perform in generating attentiveness? The possibilities are endless.

Why wouldn't advertisers, who currently pay nothing for TV ratings, be interested in such ad impact findings? Isn't this reason enough for them to consider funding the ideal national TV rating service?

The second use of data from the ideal rating service concerns program development, again, with attentiveness being the key. A program producer or a network that licenses shows from outside suppliers could duplicate many of the analyses we've cited for advertisers, only this time the focus would be on program content, not commercials. So why couldn't a portion of the cost of the ideal national TV rating service be borne by program creators, talent agencies and, of course the program execs at the TV networks, cable channels and streaming services?

WHAT'S NEXT?

Being realists, we believe that the "big data" approach will win, and that attentiveness will probably not be part of the overall design. We also believe that TV commercial time sellers will continue to pay 75-80% of the costs, which is why they will get what they want: large numbers of device-based "impressions" to market to eager buyers. And the attribution folks will have a field day trying to pin an ROI value on every commercial placement on every platform, until they finally realize that the "audience" data they are using is too wildly inflated for this to be done on a realistic basis.

It's a shame, but one can dream…can't one?