

Here's What's Wrong With TV Audience Measurement and How to Fix It

By Steve Sternberg

When it comes to national television audience measurement used for the buying and selling of commercial time, **Nielsen has long been an industry-mandated monopoly**. While potential competitors have periodically emerged over the years, they have failed to sustain enough support to survive – not because Nielsen did anything to impede them, but rather because advertisers and networks have been unwilling to pay for multiple sources to provide similar national data.

There have been a number of times when various issues arose regarding how and what to measure (household meters/persons diaries, people meters, passive meters, portable meters, commercial minute ratings, out-of-home viewing), new technology and devices (VCRs, personal computers, DVRs, tablets, smartphones) or “unusual” fluctuations in reported data (such as sudden TV usage declines among households, kids, or young men). And most recently, problems maintaining a viable sample during the pandemic. But while there have been occasional calls for change, the industry's unwillingness to support any real competition to Nielsen, along with networks and media agencies continuing to sign long-term Nielsen contracts, made the status quo is hard to budge.

A Sternberg Report Sponsored Message



The time is **NOW**
to begin your **Cyber Security Strategy**

Reach out to us for a free Cyber Security Assessment



LEARN MORE

emazzanti.net

844.360.4400

There have been only two instances, more than 20 years apart, when Nielsen was forced to completely overhaul how it reported television audiences – in 1987, when the switch was made from the antiquated household meter/diary system to national people meters, and in 2009, when DVRs and time-shifted viewing led to the switch from program ratings to average commercial-minute ratings (C3).

How those two fundamental changes were implemented shows how much the industry's priorities have changed over time – or really how the speed of change in the access to video content and in consumer behavior has caused good research to take a back seat to rushing new products to market.

There was a time when any change to audience measurement needed to undergo extensive vetting before the industry would consider implementing them. People meters, for example, were evaluated by industry committees for nearly three years – including a full year of side-by-side data with the old meter/diary method – before agencies and networks decided to use it as currency. AGB's attempt to enter the U.S. market with people meters pressed Nielsen to develop its own system – and since AGB did not provide anything substantially different, there was no reason for the industry to pay for two essentially identical measurement services. Of course, back then there were no compelling forces at work to make us move fast.

VCRs had promised but never realized the idea of people programming their own viewing schedules. Even though by the late 1990s, roughly 90% of the country owned a VCR, its major use continued to be renting or buying movies. The average amount of actual time-shifted viewing in primetime never rose above 5%. But even more significant, Nielsen was never able to measure VCR playback among people (although it could measure recording and playback among

households – it ascribed household playback to demos). While this did become a big deal among advertisers and agencies for a few years, the networks didn't see it as a problem. Nielsen eventually just said sorry, nothing we can do, and the industry moved on.

DVRs actually changed viewer behavior and forced the industry to change. Unlike VCRs, the DVR allowed viewers to record television shows without having to use videotape, and both recording and playback was significantly more user-friendly. As a result, there was a substantial amount of time-shifted viewing. Also unlike VCRs, Nielsen was able to measure playback activity. But while VCRs had become the fastest growing electronic device since television itself, DVR ownership grew at a much slower pace. Having been introduced by ReplayTV and TiVo in 1999, it took a decade before DVRs were even in 25% of U.S. TV homes, and another decade before it reached half the country. Nielsen finally started including DVR households in its national people meter sample in January 2006, when the devices were in nearly 20% of homes.

Commercial avoidance through fast-forwarding had become one of the biggest advertiser concerns, and measuring DVR playback and commercial ratings was now a priority. Virtually every study and survey ever done indicated that anywhere between 70% and 80% of VCR or DVR playback involved fast-forwarding through commercials. But there was no actual Nielsen data to prove it.

Commercial Pod studies paved the way for C3. As the industry started to publicly grapple with how to deal with delayed DVR viewing and fast-forwarding, some media agency analyses were already underway. My Audience Analysis group at MAGNA Global conducted extensive commercial pod studies over four consecutive years (2005-2008), examining the dynamics of commercial pod versus program performance for broadcast, cable, and syndication. We recorded hundreds of television shows across seasons, dayparts, and genres. We separated program segments from

commercial pods for each telecast, and matched it up with Nielsen's NPower minute-by-minute ratings. We shared our reports with advertisers, networks, Nielsen, and the press. These studies gave the industry confidence that commercial-minute ratings were consistent enough from year-to-year to use as marketplace currency – which led to Nielsen implementing C3 in 2009.

Nielsen switched to C3 with virtually no research into the methodology, or whether it actually accounted for fast-forwarding. While an extensive amount of analysis went into determining whether “commercial-minute” ratings were viable to use as marketplace currency, virtually no research was done by Nielsen to determine the proper methodology for calculating them. Speed to market was considered more important than evaluating the measurement methodology. The MAGNA Global Commercial Pod Studies defined commercial minutes as minutes containing at least 30 seconds of commercials. Nielsen decided to define commercial minutes as those containing even one second of commercial time – and then weight averaging those minutes based on their commercial duration to arrive at the “average commercial-minute” rating.

In 2006 and 2007, Nielsen held a series of industry-wide meetings to figure out how to report DVR playback and commercial-minute ratings. The biggest debate at the time was not on methodology, but rather on what to report. Buyers (advertisers and their agencies) wanted only live viewing to be used as marketplace currency. Sellers (broadcast, cable syndication) obviously wanted a full seven days of delayed DVR viewing included in the reported ratings. The final meeting on the subject, in which the C3 compromise was established, was the first time that Nielsen had invited top research, buying, and sales executives in the same room to discuss audience measurement methodology and related business issues. It is also the moment that Nielsen shifted away from being a pure research company – at the following year's national client meeting, **Nielsen actually proclaimed that getting a product to market was now more important than validating the research beforehand.**

Those of us who were involved in the discussions on how to best measure commercial-minute ratings, recall that C3 was designed as a one- or two-year band-aid. The industry's major computerized pre- and post-buy analysis systems, Donovan Data and Media Bank (which have since merged to create Media Ocean), could not yet handle individual minute measurement. They still needed audience data in the format of Nielsen's MIT tapes, which used average ratings data. Advertisers wanted individual commercial ratings as soon as they were feasible. I had been working with both Donovan and Media Bank, and both did have the capability of providing pre- and post-buys based on individual minutes the following year.

Even though research was starting to clearly demonstrate fundamental flaws in C3 measurement, we were told it didn't really matter individual commercial measurement would soon be the marketplace currency. I recall telling whoever would listen at the time that given the chaos that ensued with the switch to C3, and the additional expense and staffing required to handle moving to actual commercial measurement, that C3 would likely be the end point. That was 12 years ago.

It should also be noted that the vast majority (if not all) of Nielsen executives did not (and probably still do not) know exactly how C3 is calculated. Many have probably forgotten, that when the "C" streams of ratings data were first made available, Nielsen sent an advisory to clients saying that you couldn't simply subtract Live viewers from Live + 7 Days to calculate time-shifted viewing. Since that's what everyone was doing, and it made logical sense to do so, many of us research types started to wonder why. It raised some uncomfortable questions about exactly how these calculations were done, because Nielsen quickly stopped saying it. If you specifically asked them they would tell you, but otherwise they never brought it up again.

While many people think the methodology for calculating C3 takes fast-forwarding through commercials into account, and only measures exposure to the commercial minute, this is actually not entirely correct. It has to do with the way Nielsen measures viewing to the average minute. Rather than counting the majority of viewing to the average minute, it takes the plurality of viewing. In other words, the channel viewers tuned to the longest during the minute. If, for example, in a given clock minute, you are watching 20 seconds of ABC, and 10 seconds each of four other networks, you are counted as watching ABC for the entire minute. If there is no plurality for that minute, Nielsen goes back to the previous minute and uses that plurality for both minutes.

Nielsen's rules for allocating minute-by-minute viewing were designed for a time when virtually no one had remote control and it actually took a minute or more to get up and change the channel. In the 1960s and 70s, plurality of viewing to a given minute was pretty much the same as viewing to the entire minute. Individual minutes were not reported, so it didn't matter how Nielsen measured them. They were simply aggregated up to average program ratings. No one ever thought about how viewing to a single minute was derived.

When it comes to time-shifted viewing, however, the flaws in C3 methodology become apparent. With live viewing, there are multiple channels. During a commercial break, you can stay on that channel for 20 seconds and then scan several other channels during a given clock minute. When you DVR a program and play it back, however, there is only one channel. The majority and plurality of viewing are therefore the same. If you are playing back a program, watch 20 seconds of a commercial minute and fast-forward the remaining 40 seconds, your viewing falls into something Nielsen calls AOT ("all other tuning"), and the commercial minute will receive no credit for your viewing (because both the majority and plurality of tuning is not to the commercial minute). You would need to be tuned to the majority of a clock minute to be counted in delayed DVR viewing. So, Nielsen's

methodology for counting live commercial minutes and delayed commercial minutes are not only different, they are both fundamentally flawed.

Nielsen still can't accurately measure streaming, and won't be able to until its sample reflects individual streaming service penetration and until it can measure viewing on all platforms and devices. Netflix's two minutes of viewing data is meaningless gibberish that tells us nothing about how many viewers are actually watching anything. The press should stop repeating (as opposed to reporting) these bogus numbers provided by streaming services. Is *Squid Game* really Netflix's most popular series ever? Maybe. If ABC said 50 million people watched *Grey's Anatomy* this month without any third-party verification, I doubt anyone would report it as a real number. In reality, there are no streaming data that is anywhere near being comparable to linear ratings reported by Nielsen or ComScore. The fact that Nielsen only reports on titles from Netflix, Hulu, Prime Video, and Disney+ (it just added AppleTV+), and only measures streaming on a TV set, is an ongoing embarrassment. Netflix has repeatedly said that Nielsen's limited reported streaming data is "not even close" to the actual numbers.

A number of advertisers, media agencies, and industry trade organizations have joined NBCUniversal's Measurement Innovation Forum, which is designed to evaluate current audience measurement methodology and come up with alternatives to Nielsen. In August, NBCU sent out a request for proposal (RFP) to more than 50 companies (including Nielsen) to help "build a new measurement ecosystem for us that reflects the future." Ironically, this is not in response to the basic flaws in current C3 measurement, but rather was spurred by the problems Nielsen encountered maintaining its national people meter sample during the pandemic – which resulted in reported network audience declines that seemed illogical to say the least.

This seems somewhat similar to the Council for Research Excellence (CRE). I was one of the founding members of this group, formed in 2005, which consisted of 40 top research executives from major Nielsen clients – representing advertisers, media agencies, broadcast and cable networks, syndication companies, local television stations, and industry trade associations. There are obviously some significant differences.

The CRE was designed as an independently operated group, which was basically set up so Nielsen could avoid government intervention into how it measures media audiences. Its stated goal was to evaluate Nielsen measurement methodology and recommend ways to improve. Nielsen funded the research at a cost of about \$3.5 million per year (although having no say in the projects we undertook). The CRE's first major effort was the landmark *Video Consumer Mapping Study*, which still stands as the best original research into consumer media habits I've ever seen. The study was conducted in 2008 by the Ball State University Center for Media Design and Sequent Partners. This study cost about \$2.5 million, a number simply out of the ballpark for any agency or network to conduct on its own. At the time, it seemed as though the CRE was the only industry entity capable of doing real good independent research, designed solely to advance how audiences are measured, without any sales-related agenda or bias – in retrospect, perhaps a naïve notion. While there were several more interesting analyses produced by the CRE over the years, it never lived up to its initial promise and Nielsen pulled funding at the end of 2017.

The NBCU-led effort is obviously not funded by Nielsen, and it remains unclear how much any accepted proposal will cost and who will fund it. Also unclear is whether this is designed to find an alternative to Nielsen or a supplement to Nielsen.

So, how do we “fix” TV/video audience measurement? If the real (as opposed to PR) goal is to actually build a better system, there are three basic steps that need to be taken. Here is my proposal to NBCUniversal's Measurement Innovation Forum:

The first step should be to determine what current Nielsen measurement is doing right and wrong.

- Develop a sample of 300 or so industry researchers, people who are adept at detailed work and understand how audience measurement works. Nielsen (and others) would meter their TVs, DVRs, and any other video viewing capable devices they own. They would also be given portable meters to measure their exposure to out-of-home video and audio.
- Set up several different scenarios, from the simplest to the most complex viewing environments, for what and how each participant will view video content and commercials, in-home and out-of-home during a single day.
- Participants would then log in their video and advertising viewing (down to the second). They should write down their media behavior in minute detail, what they are watching, what platforms they are using, whether they are time-shifting, when they are fast-forwarding through commercials, when they switch channels, which commercials they are seeing (and hearing), etc.
- Compare their actual viewing to what the meters report. We would then be able to address weaknesses in audience measurement, see how much real-world viewing is not being captured by current methodology, and gauge the differences between such things as commercial exposure (the opportunity to view) and actual viewing, and between individual commercial viewing and C3.

Not only will this provide, for the first time, a look at how accurate reported ratings are, but it will also tell us exactly where improvements to audience measurement need to be made (which was the original but unfulfilled purpose of the CRE).

The second step is to determine whether samples or panels are even adequate to measure national video viewing in today's increasingly splintered media environment.

I would argue as currently structured they are not.

- The whole purpose of a TV sample is to be representative of the total U.S., with the idea that you and your demographic cohorts have similar viewing habits. Your cohorts are determined based on characteristics that Nielsen has determined impact viewing behavior – such as age, sex, race/ethnicity, presence of children, income, education, language spoken at home, etc. Also important is whether or not you have cable/satellite/telco (i.e., pay TV services) or DVRs. On occasion, Nielsen has added certain things to the list, but even then, it was always with the impact on linear TV viewing in mind.
- In the past, broad platform mattered more than individual platform. In other words, whether or not you have cable does not require the sample being representative of every cable system, because that does not significantly affect viewing choices or behavior. Nor is it important to know what type of DVR you have, since that will not really impact your viewing behavior and how much you fast-forward through commercials. Whether or not you subscribe to streaming services, however, is not the same. If you only have Netflix, the programming available to you, and your actual viewing habits, will be dramatically different than if you only subscribe to Paramount+. Likewise, if you bundle Disney+, Hulu, and ESPN+, and add HBO Max, your viewing habits will be substantially different than if you have Netflix, Amazon Prime Video, Peacock, and Apple TV+. And there won't just be differences in what you stream, there will be significant differences in your linear TV viewing as well. **The only way to have a meaningful television sample, is to accurately represent the total U.S. based on**

every possible combination of streaming service and cable/satellite/telco subscriptions. This means that the make-up of Nielsen's national and local samples could need to change dramatically every year. The days when all Nielsen needed to worry about was cable and DVR penetration are gone forever.

The third step is to decide if a single source or multiple sources are preferable for measuring multiple current and future platforms. I used to be firmly on the side of believing single-source measurement was the way to go. I'm no longer so sure that any one company is capable of measuring all video viewing on all platforms.

A final note in the category of "be careful what you wish for" - **understand that having more accurate video viewing data is likely to make few people happy.** Although advertisers do want the most accurate data, they and their agency representatives need stable (i.e., projectable) audience data in order to effectively plan and buy commercial time. Networks, on the other hand are interested primarily in higher ratings. While the high cost of supporting a competing national measurement service to Nielsen is one factor in continuing to only have a single currency, the main reason is that none have demonstrated they would produce higher ratings.

The only real way to put pressure on Nielsen to make fundamental changes or to actually go with an alternate measurement system is to move toward one-year contracts. As long as all the media conglomerates continue to sign long-term 5-to-7-year deals, not much is going to change.