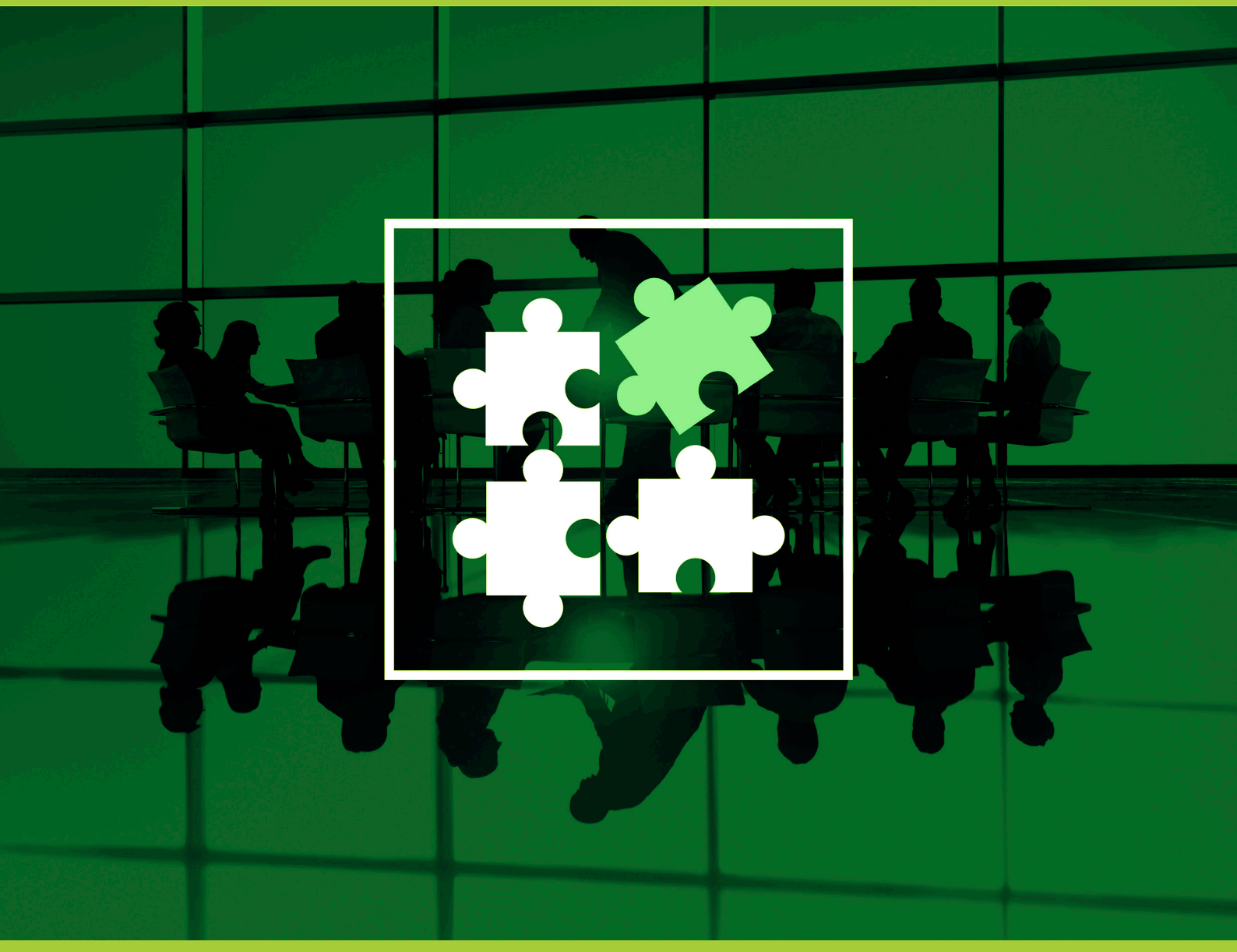
Coalition for Innovative Media Measurement

SOLVING TODAY'S EVOLVING TV MEASUREMENT PUZZLE

Understanding Key Data Components/Approaches and Deciphering Differences in Big-Data Based Currency Metrics

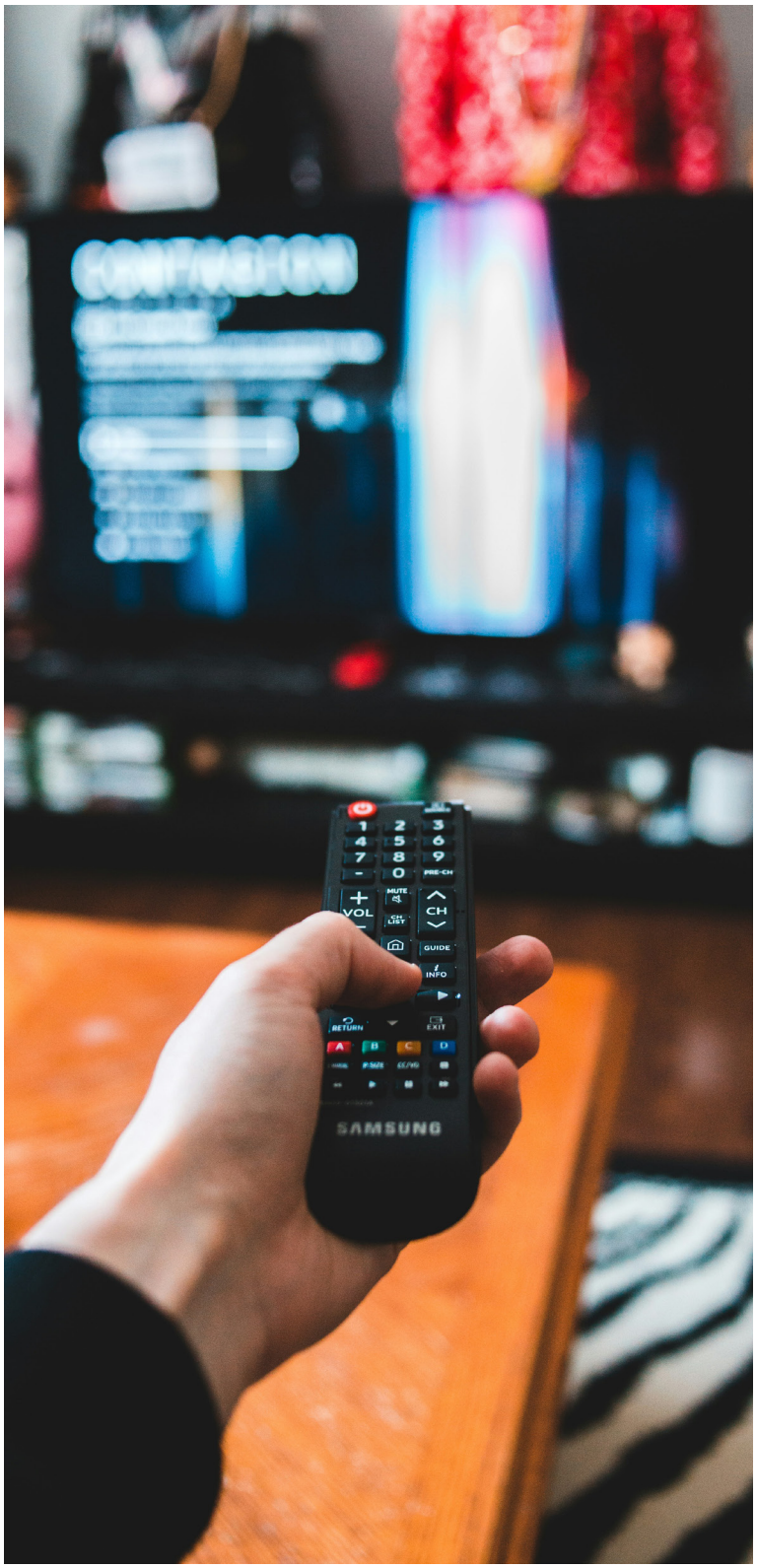
By Josh Chasin and Albert Lau

September 2024



CONTENTS

Preface

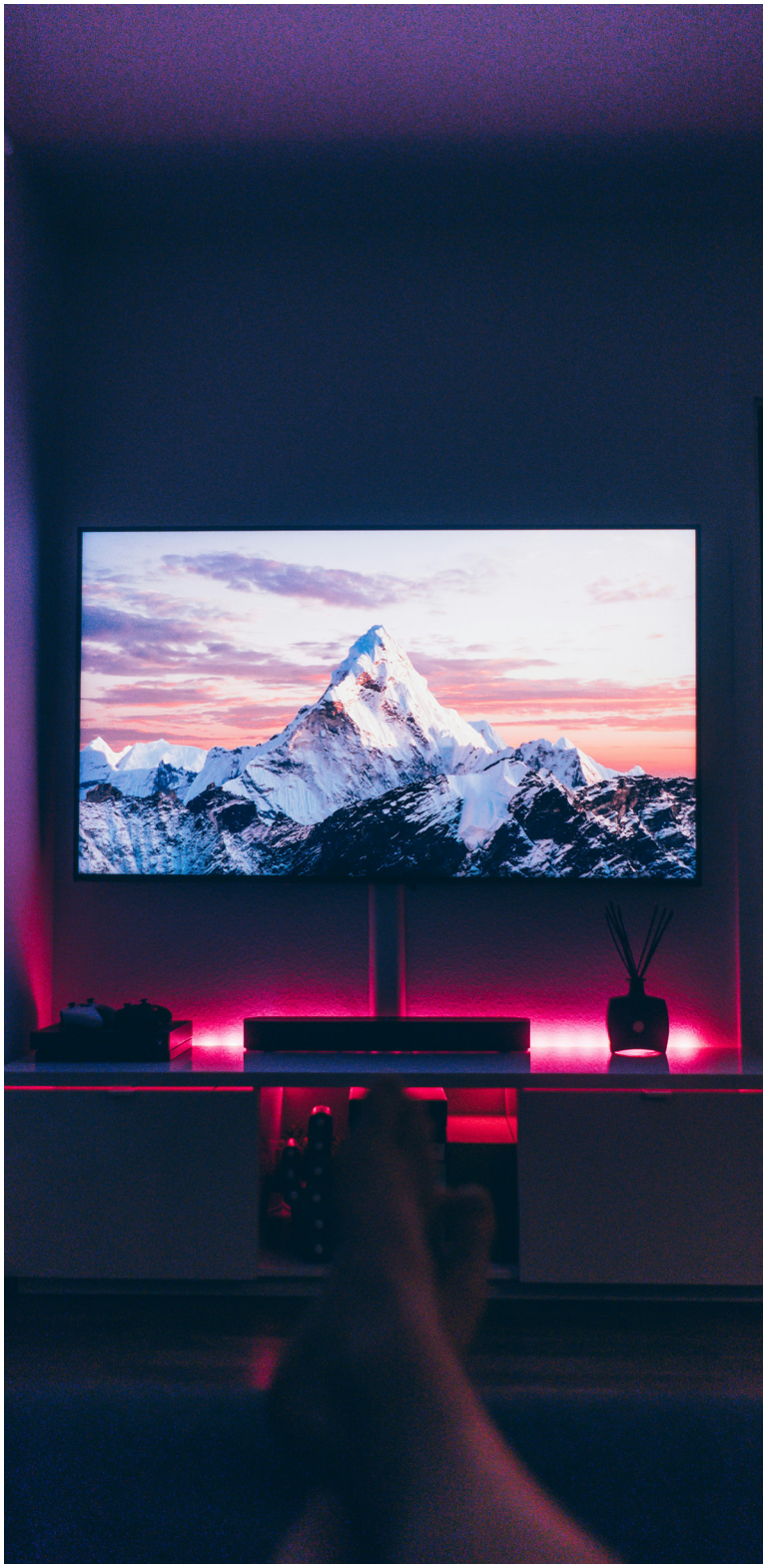


The Coalition for Innovative Media Measurement (CIMM) is a non-partisan, pan-industry coalition of companies from across the media and advertising ecosystem, focused on supporting improvements, best practices and innovations in measurement and currency development, the use and application of new metrics and approaches to understanding the value of media, and data collaboration and enablement.

As part of our program, CIMM commissions papers, thinkpieces and perspectives from industry analysts, experts and thought leaders – to provide insights and occasionally provocative perspectives on critical issues of interest to our Coalition of members.

The studies always involve original research, but unlike our larger studies, are not peer reviewed and do not generally involve a Project Steering Group. The views, thoughts, and opinions expressed in this paper belong solely to those of the author and not necessarily to CIMM, the author's employer, organization, research interviewees and participants, or to any other group or individual.

Introduction – overview and approach



US television consumption behaviors have been rapidly changing in the past decade, making measurement increasingly challenging as the landscape becomes more complex, leading to legacy TV measurement companies struggling to keep pace. Historically, TV program (content plus ad time) viewership has been used as a proxy for advertising exposure. In 2006, Nielsen began publishing delayed viewership with the advent of DVRs. In the meantime, their Average Commercial Minute (ACR) metric was matured enough to allow the industry to create the C3 metric as a measurement and transactional product, which combined the measurement of estimated commercial viewing time live as well as delayed up to 3 days (to manage sales promo windows). The advent of the C3 metric in 2007, which incorporated ACM, officially created a distinction between TV content and advertising exposure. Meanwhile, digital ads including digital video and CTV are primarily measured by tracking ad exposures directly from the ad serving process.

Due to the differences in distribution processes, the same episode of the same program will carry different spot loads depending on the distribution channel (live linear; VOD; FAST player), with varying measurement methods.

As the US TV advertising industry migrates from traditional sample-based TV measurement methodologies (i.e. panels, diary sweeps, surveys) to solutions built from various big data sets, the process and criteria to evaluate quality and accuracy logically change.

With the growing availability of big TV data sets, which have reduced barriers to entry in the measurement marketplace, the US now has multiple currency measurement providers – more than at any other time since the first TV services launched. These different measurement vendors leverage different data sets and (often, very different) methodologies, resulting in very different measurement outputs that often, diverge to varying degrees.

Multiple and diverse data sources are a notable feature of the current marketplace. In this paper, we identify and review the most critical methodological challenges currently facing big data-based currency-grade measurement providers, offering some guidance on how best to evaluate potential quality issues. In a time when alternative currencies and measurement solutions built off different data sets and methodologies produce different outcomes, it is essential to identify some of the opportunities for the industry, collectively, to help measurement vendors to address these methodological challenges. Ultimately, we hope this paper can help to point the way toward codification of some best practices and, potentially, collaborative industry initiatives designed to address some of the most critical methodological challenges.

Our approach

The authors of this paper have extensive experience of reviewing big data methodologies and measurement outputs, one from a vendor perspective, the other from a buy-side perspective. Both have several decades of TV research and measurement experience providing unique perspectives on best practices in various measurement aspects. They both have pioneered changes in the industry and understand the steps needed for success. We believe that our collective (but different) experiences give us a unique perspective on current challenges, one that we hope will be of value to the industry.

In addition, we were able to secure input from a spectrum of industry experts from across the measurement marketplace, including representatives from the buy side, the sell side, and from each of the four primary providers of currency-grade measurement. Their input is greatly appreciated and helped shape and solidify several of the key points we are making on the measurement challenges.

With their input, we were able to establish what we believe is a broad consensus about the key elements that account for significant differences in the outputs of big data-based measurement solutions – and the most important errors, if not adequately addressed.

Ultimately though, the opinions expressed in this paper are those of the authors and should not be taken to reflect the views or opinions of our contributors or our previous employers.

About the authors

Josh Chasin is currently Principal at KnotSimpler, a consulting practice specializing in the development and monetization of data assets as syndicated measurement services. Josh has extensive experience at syndicated audience measurement vendors, including Vice President, Marketing, New Ventures at Arbitron; Chief Research Officer at Comscore; and Chief Measurability Officer at VideoAmp. He has been involved in the successful design and management of audience measurement offerings in network TV, local TV, network radio, local radio, newspapers, magazines, Out-of-Home, the Internet, and cross-platform video. He is a past winner of the Erwin Ephron Demystification Award and was a 2012 Advertising Age Media Maven.

Albert Lau is a seasoned advertising agency media researcher with years of measurement experiences. Most recently at Omnicom Media Group, Al was tasked with the assessment of all the major multi-currency products launched in the market and provided topline assessment on how the marketplace could be impacted. His detailed assessment covered several years of newly produced viewership metric in which he covered various metric, data, and methodology changes. These changes allowed him to provide a unique data perspective across vendors on how different and challenging national TV projections could be. Earlier in his career at GroupM, he helped usher the C3 metric and created a roadmap on how the industry could adopt a new and sustainable currency metric at scale with all the major stakeholders' blessing, which included marketing clients, TV networks, agency planning and buying leadership, and Nielsen's sales and production teams.

Acknowledgements

Many thanks to the following contributors:

1. Eric Cavanaugh
2. Marzena Del Gaudio
3. Pete Doe,
4. Noelle Huynh
5. Adria Jewell
6. Tameka Kee
7. Bharad Remesh
8. Travis Scoles
9. Michael Vinson
10. Jon Watts
11. Brian West
12. Leslie Wood
13. the ARF team

(Paul Donato, Garg Deepak, Samantha Zhang).

As noted above, the opinions expressed in this paper are those of the authors and should not be taken to reflect the views or opinions of our contributors or those of any other companies named in the paper, above.

Key takeaways

Big-data measurement providers face six critical methodological challenges:

1. **Assessing the impact of identity** within data and the role that it plays as a building block and source for demography. Match-rates, quality of demographic data, and accuracy of identity resolution are, collectively, an extremely important determinant of quality.

2. **Addressing data footprint coverage biases,** such as geographic and demographic skews in various big data sets, and determining how best to quantify, correct, and calibrate the data.

3. **Onboarding, cleaning, and combining** big data assets.

4. **Sourcing, applying and analyzing accurate metadata** to provide key contextual elements in content and campaign identification.

5. **Integrating linear and digital** (streaming) data together from multiple sources and producing deduplicated cross-platform metrics such as reach.

6. **Establishing processes and methods for addressing coverage gaps and shortfalls** in data sets (e.g. Over-The-Air TVs, houses without Wi-Fi), determining what's left out or underrepresented in the big data building blocks.

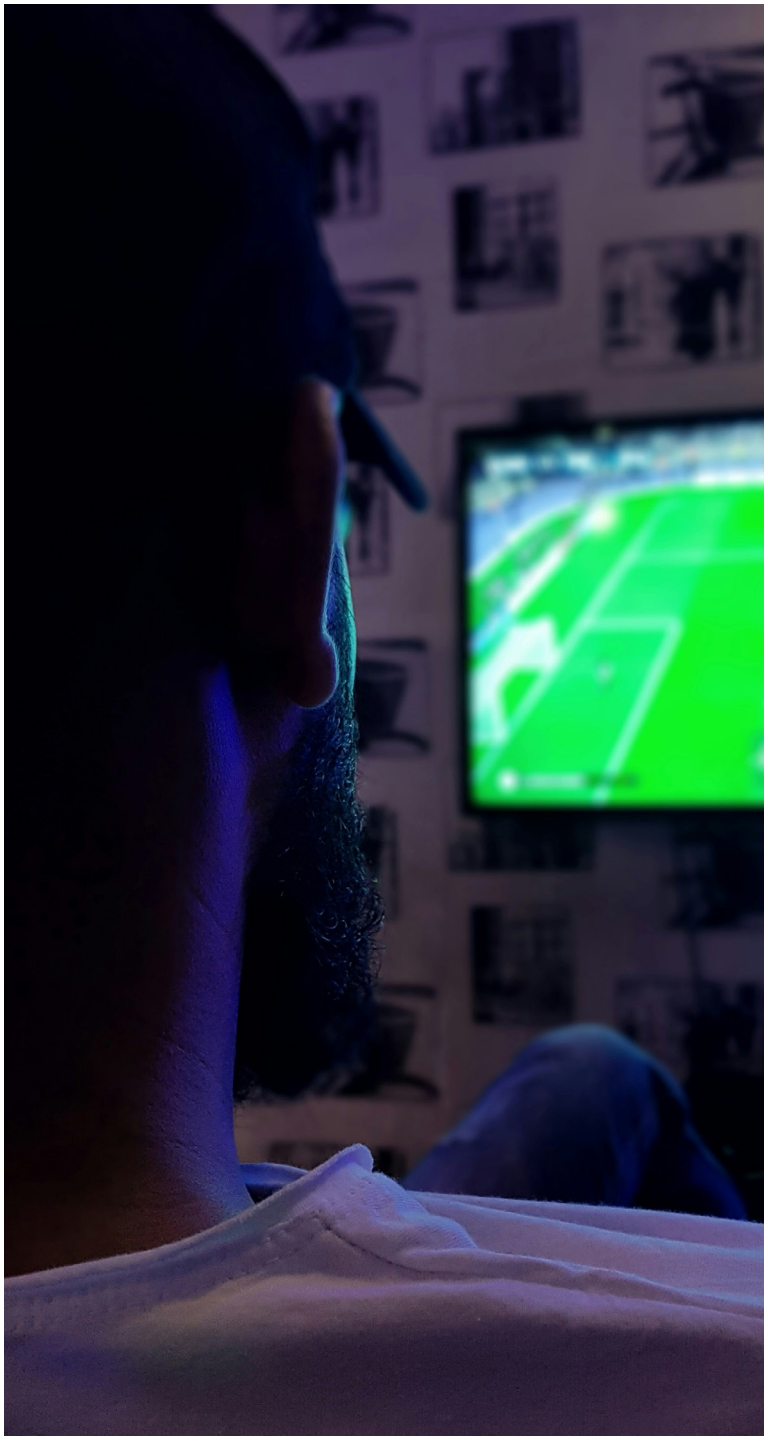
All measurement companies have their own approaches to addressing these challenges, with varying results. However, simply addressing these challenges does not automatically mean that a measurement solution will be of high quality. All four major competing currency providers (Comscore, iSpot, Nielsen, VideoAmp) are engaged with the MRC to varying degrees. The MRC audit process is "table stakes" for a media transactional currency, meaning an MRC the accreditation process is a requirement for a service to be used in advertising transactions. The MRC audits against standards, but today many key components of big data currencies either lack standards (e.g. Identity) or have standards that are due for a refresh (e.g. Set-top Box standards).

A refreshing of industry standards, crafted by key stakeholders like publishers, agencies, researchers, and marketers, would help the industry to align around best practices, allowing measurement customers to accurately assess the quality of the solutions they are using.

In the main section of this paper, we will provide our unique perspectives on the practical implications of each of the methodological challenges that we have identified and what end users, the customers of the measurement solutions, should be focusing on. With different, competing methodologies and data sets yielding different results, measurement users will need to undertake careful assessments and due diligence to apply output into the current working environment with confidence, unless this work is performed centrally.

Before we explore the specific challenges that the industry faces, there are several changes in the marketplace with media consumption and measurement approach. The following sections will provide a brief overview on the marketplace that led to industry's need to explore alternative video measurement approaches.

1. A changing market environment for TV measurement in the US



Change #1: The Media Environment

In the past 30 years, the explosion in the number of linear viewing choices has resulted in high levels of viewership fragmentation and a far more complex, diverse and competitive media marketplace. The phenomenon of streaming has served to rapidly accelerate this fragmentation; today, virtually everything is available all of the time. According to the ARF's DASH study for 2023, 90% of internet-connected households have 10 or more connected devices (TVs, gaming consoles, phones, tablets, computers, Smart Speakers.) Many of these devices comes with easy-to-adopt apps from thousands of content providers, including all the popular SVODs, AVODs, vMVPD, and FAST channels.

This level of fragmentation has made traditional panel-based measurement projection approaches far more challenging, given the complexities associated

with reliably measuring long-tail TV networks and streaming apps, let alone individually streamed programs. The migration to narrowly defined targets further fragments viewing beyond the ability of a panel to robustly provide measurement.

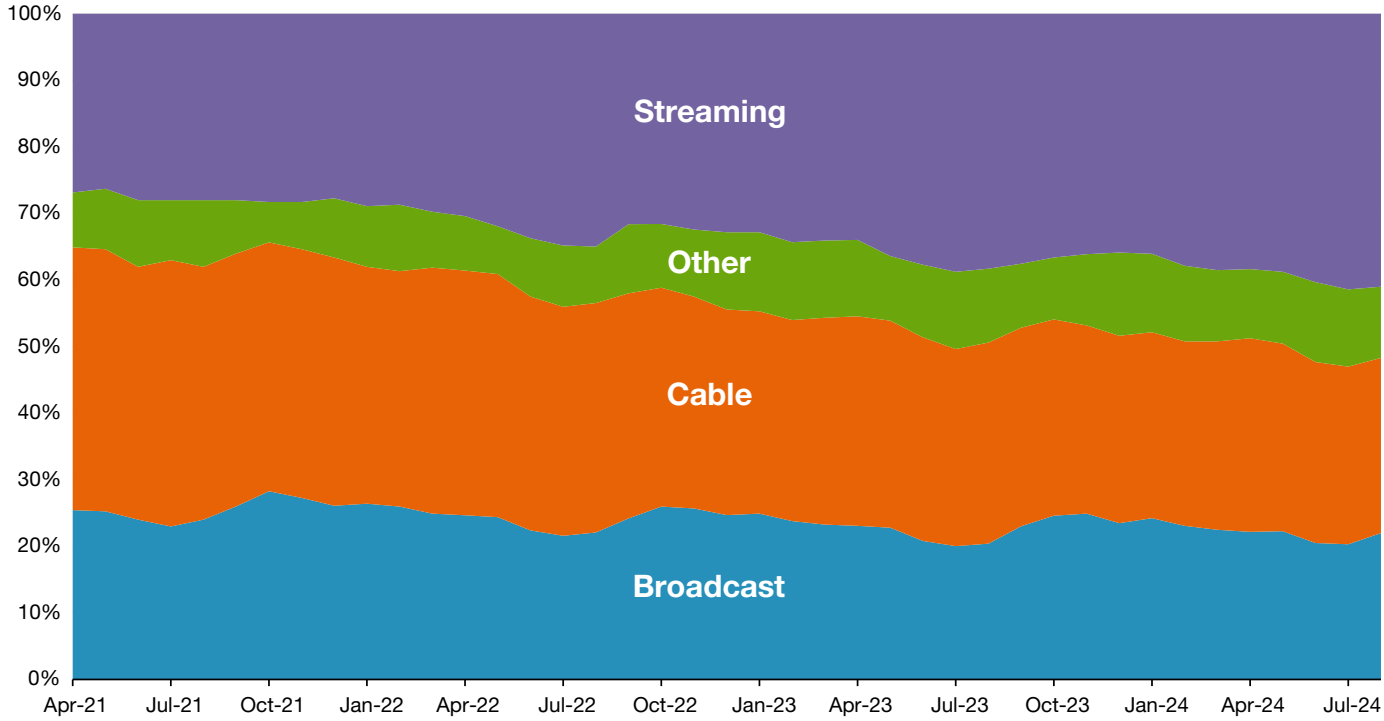
According to DASH, 75% of US households have a smart TV and even more have internet connected devices and STBs with return path data. Likewise, DASH reports that 92% of US households have at least one smart phone. Traditional cable/ satellite penetration in the US was in excess of 85% ten years ago compare to about 50% of US households now subscribe to a pay TV bundle (and this includes vMVPDs such as YouTube TV.). Streaming services like Netflix, Hulu, Amazon Prime, and YouTube account for significant viewership share, and traditional program provider companies (i.e. networks) have responded by launching their own streaming platforms with different degrees of success. The big winner in 2023 was ad-supported VOD (AVOD); DASH reports US households subscribing to an AVOD package (including lower-priced tiers of Netflix and others) almost tripled, from 17% in 2022 to 45% in 2023.

The combination of cord-cutting and streaming disruption has led to a significant decline in linear TV viewing during the last decade, with viewing becoming far more widely distributed, especially in the case of younger viewers. In March 2023, only 57% of all content viewership was attributed to traditional TV sources (Cable and OTA), according to Nielsen Gauge. Streaming on television went from little or no reportable coverage 10 years ago to 43% of TV viewing today. This change in behavior over this short a period of time was dramatic, and the TV measurement infrastructure struggles to keep up. Panels, designed to measure a handful of broadcast viewing options, are not purpose-built to measure today's fragmented media consumption, including a myriad of digital streams, accessed in a non-linear fashion. Thankfully, several upstart companies like VideoAmp and iSpot created their own unique measurement solutions to challenge the traditional panel-based approach. With more competition in place, all the measurement companies must evolve and move towards more comprehensive and encompassing solutions to address the ever-changing viewership behavior.

While the majority of available impressions remain on linear TV channels (with higher ad loads and no ad-free options), new ways to watch the same episode of the same program have proliferated (live linear; the network streaming platform; the cable operator's VOD offering), resulting in different viewers seeing different spot loads while watching the same show.

Finally, and non-trivially, watching TV programming has become increasingly divorced from the TV set (traditionally, the base unit of measurement in a panel.) Today, consumers can (and many do) watch TV and other video content on computers, gaming consoles, mobile phones, and tablets. Fragmentation exists not just across a broader range of content choices, but also across devices and platforms (Over the Air, cable/satellite, streaming platforms, streaming apps, etc.).

Viewership Migration From Linear to Streaming Continues



Source: Nielsen Gauge (2024)

Change #2: The New Measurement Toolkit

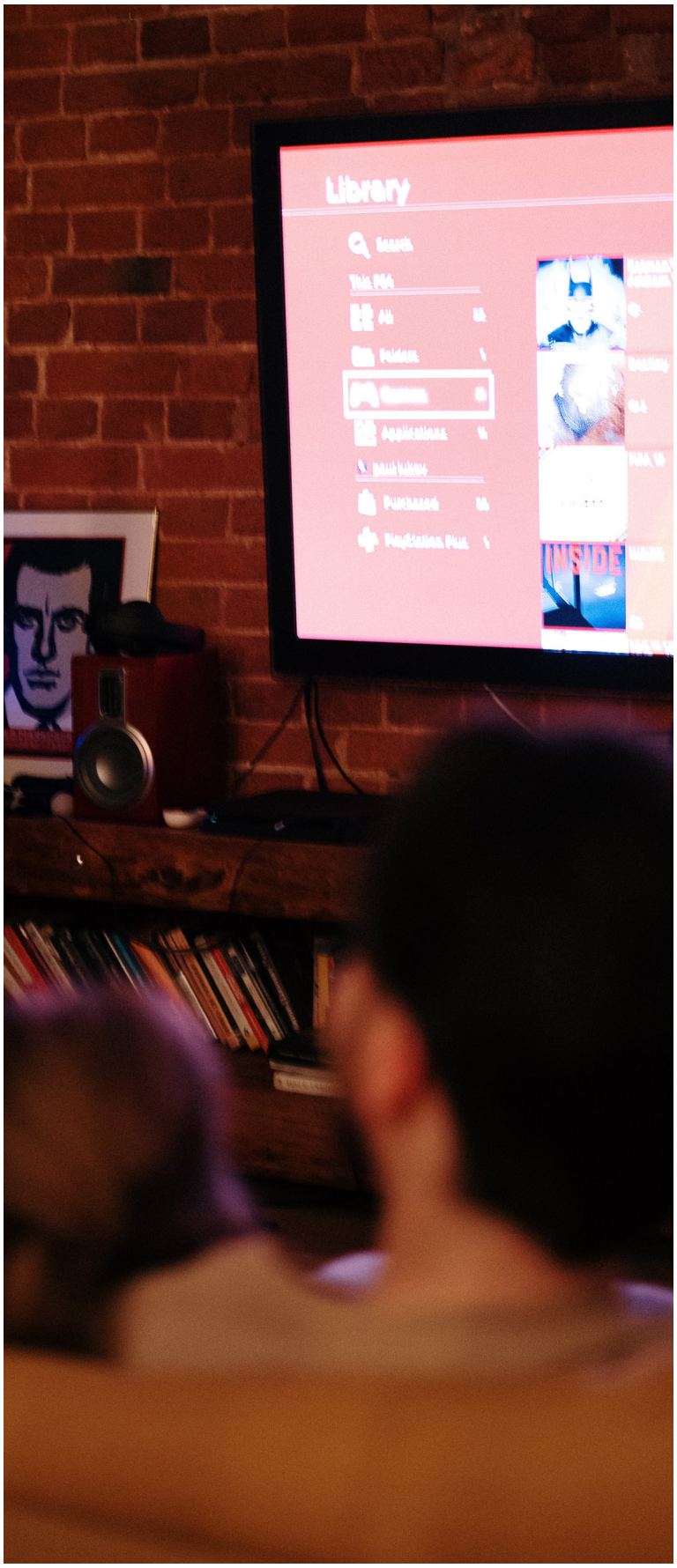
Perhaps the most profound change in the history of audience measurement has been the evolution of the practice over the last decade or so, from one of sampling to one of data science. The pioneers of audience measurement were generally survey sampling statisticians; today though, the development of audience measurement solutions has become the domain of the data scientist.

Put simply, many of the traditional drivers of measurement quality in a sample-based system (sample frame, sample selection, response rate) are less relevant for solutions based on big TV data sets and data science.

Table 1: Panel-based versus Big Data Measurement Solutions

	Sample-Based (Panel)	Data Science-Based
Design	Sample frame representative of the population	Footprint based on accumulation of available data assets to provide signal
Scale	Hundreds to tens of thousands of persons, HHs or devices	Millions of persons, HHs or devices; in some cases, a complete census of all data (e.g. ad server logs)
Selection	Selection of a random stratified sample from the frame	Typically, inclusion of all available reporting HHs/devices
Recruitment	Elicit active consent from selected sample via mail, telephone, Internet, and in-person	Unnecessary; consumers included unless they opt out of measurement in advance
Demographic, other characteristics	Collected directly from panel members (self-reported) with in-person survey assessment	Appended/derived/assigned from one or more third party identity providers
Spine	In-tab persons/HHs/devices	Identity graphs
Measurement instrument	Specifically designed for the task at hand (e.g. a TV meter to collect TV viewing)	Typically, the actual device delivering the content (STB; Smart TV). Can also be a digital return path asset (e.g. a pixel)
Coverage	Given a quality sample frame and design, primarily limited by non-response	Generally comprehensive within footprint; excludes elements not in footprint. Also excludes persons refusing consent to share data.
Data collection	Collected directly from participants via passive, active techniques (people meter button pushing)	Automatic using ACR, RPD, or digital census data. No respondent burden, but also no person level identification
Processing	Data cleansing, weighted and projected to universe	Data cleansing, weighted and projected to universe
Sampling error	Potentially significant with audience fragmentation, especially in long tail networks or among smaller targets	Not strictly applicable as these aren't samples per se; however, reliability may be quantified via replication studies. Such volatility is minimal compared to sample-based solutions.

1. A changing market environment for TV measurement in the US



For vendors, the new toolkit for measurement solutions includes seven main components:

Component #1: Set-top Box Return Path Data: Note that as cable and satellite penetration has declined in the US from roughly 85% circa 2010, to around 50% in 2024, the efficacy of Set-top Box data alone as a measurement asset has declined.

Component #2: Smart TV data: Major smart TV manufacturers (OEMs) include Automatic Content Recognition (ACR) technology in their sets, enabling the identification of what is on the glass via creating an audio or video fingerprint, and matching the fingerprints to content and/or ad libraries.

Component #3: Digital census data: Digital census data, collected either from pixels or via server log integration, enables the measurement of a census of impressions for digital, addressable ads (and offers the promise of supporting a measurement of a census of digital content consumption). As more video consumption shifts to streaming, digital census data will become an increasingly important component of cross-platform video measurement. However, such measurement requires the

participation of the programmer or the advertiser, so the notion of a census is limited to participating parties. All the data, from some of the players.

Component #4: Streaming platform data: While no such data is currently generally available, the possibility exists to create a system of streaming platform measurement analogous to the way Set-top Box data has been used to measure linear TV viewing. STB data provides the atom of measurement for linear TV: start time, stop time, and channel tuned. Similarly, Streaming providers like Roku, Amazon Fire, Google Chromecast, and even Smart TV manufacturers could conceivably make available data containing the atom of streaming measurement: start time, stop time, and app tuned. Such data could, if commercially available, enable the creation of measurement systems that can track streaming across services/apps– with the same sort of adjustments and projections deployed in projection from Set-top Box data for linear TV. Also, it is worth noting that "the JIC" promises to make streaming data for both content and campaigns available to certified measurement providers, although this will be limited to programmers participating in the JIC.

Component #5: Identity: The advent of big data assets and the requirement to combine and deploy them in a privacy-compliant manner has led to the emergence of the concept of identity as an essential component of currency audience measurement. identity is the spine through which big data assets are integrated.

Component #6: Clean Rooms: Clean rooms are a technology wherein data from disparate sources may be combined into reports on an ad hoc basis, with the owners of the data retaining control over who has access, for how long, and at what level of granularity. Typically when clean rooms are deployed, an identity spine is used within the clean room to facilitate data integration.

Component #7: Calibration: some data source, typically a panel, to serve as a "truth set" or training set for training or calibrating data from big data assets.

Big data assets linked through identity in clean rooms is the new model for currency audience measurement. This model is profoundly different from the traditional, sample-based approach to audience measurement. Inevitably, a new set of evaluative criteria must emerge to guide data users to the best solutions.

Change #3: Migration to Big Data-Based Metrics

Given changes in ad and content distribution and viewing technology, the changes in measurement tools available, and the unique measurement challenges these changes introduce, we have identified six critical methodological challenges facing big data-based currency providers. These methodological challenges have significant implications for the measurement outputs provided by vendors.

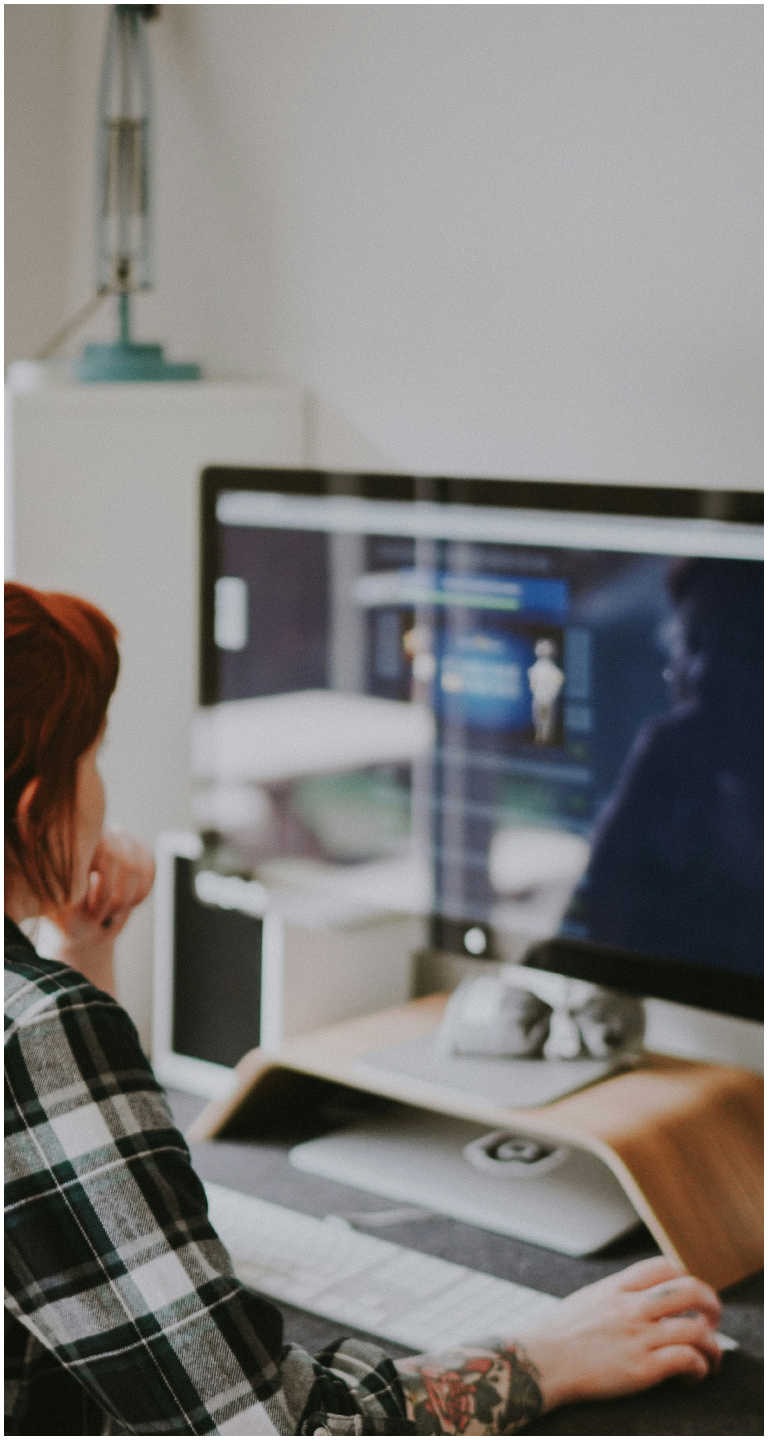
We believe that it is instructive to explore these challenges from the perspectives of both the measurement vendors and their customers, predominantly agencies and publishers.

One important observation stemming from our analysis is that measurement customers, confronted by different currency choices with different results, are strongly motivated to support cross-industry improvements (e.g. standards, common assets, convergent methodologies) that minimize the variations in the reported measurement estimates from different vendors.

Understandably, different approaches, often leveraging very different data sets and methodologies, will often produce very different results, resulting in considerable uncertainties for customers, especially when the measurement results are being used to support important investment (and other business) decisions. These uncertainties drive data users to seek ways to minimize divergence across currency sources. Conversely, measurement providers are motivated to pursue excellence in seeking truth and are not incentivized to strive for convergence. (Indeed, anti-trust concerns create a natural impediment to the pursuit of convergence across competitors in the space.)

However, to the extent that the measurement providers strive to eliminate sources of error and to address these methodological challenges, some convergence seems likely, if not entirely inevitable, given that error is a contributor to deviation from the truth.

2. Critical methodological challenges facing big data measurement solutions



Currency users tend to assess measurement solutions through a different lens, looking for projectable, consistent, and stable metrics suitable for business needs and investment decision.

Challenge #1: Assessing the impact of identity

The measurement vendor perspective

The concept of identity was unnecessary in a sample-based measurement solution. Specific people were recruited, and the measurement vendor knew precisely who they were. Techniques like fusion were developed to integrate data from different databases, which involved appending records from respondents in a donor data set to records for individual persons in a host database. But even with fusion, identity was simply the actual persons in the host dataset.

The advent of big TV data assets and the requirement to combine and deploy them in a privacy-compliant manner has

led to the emergence of identity as an essential component of currency-grade audience measurement solutions. Today, identity is one of the most important variables determining the quality of the measurement solution. Identity is the base, the spine of all big data solutions in marketing, advertising, and media measurement. Identity facilitates the unlocking of the power and utility of big data assets.

However, identity spines are not perfect, and these imperfections can lead to a range of methodological challenges - for example:

- Match rates: Not every household, person, or device in a big data footprint will resolve to a match within a given identity spine.
- Errors in the actual match of devices to households (i.e., wrong household).
- Potential error (and magnitude thereof) in mapping devices to households.
- Potential error (and magnitude thereof) in identifying the persons in the household (i.e. completeness and accuracy of the household roster), and in mapping persons to households when person-level signals (e.g. HEMs) exist.
- Potential error (and magnitude thereof) in demographic assignment at the person and household level (since demography comes from identity providers).
- Accurate accounting for kids and teens (typically, identity providers maintain rosters of adult household members with precise ages and genders; but presence of children is treated as a household characteristic, with presence indicated within age band, and without gender.)
- Differential coverage of identity providers by race/ethnicity (and new privacy laws at the state level restricting the ability to provide such data.) This makes representation of diverse populations particularly challenging; weighting may be remedial, but it cannot replace missing data (i.e. 9dentity partners no longer providing race/ethnicity data.)

Over time, the data quality challenges facing identity providers have been intensely scrutinized. These are non-trivial issues. For example, when some Smart TV or Set-Top Box households in the currency provider's footprint fail to make reporting in-tab sample due to failure to resolve in a match, the measurement provider must decide how best to report those households. If the unmatched households behave differently than the matched households, this can introduce bias into the reported measurement figures.

However, perhaps the greatest quality issue relating to identity is the use of spines to provide demographics for households and the persons within these households. The majority of

cross-platform video is still transacted based on demographic targets, but systematic errors and biases in demographic assignment can lead to misallocation of spend against a demographic target. This is a relatively new challenge, as sample-based data relies on self-reported demographic assignments, which are generally assumed to be accurate. If, for example, an identity provider has false positives or false negatives for the characteristic of Hispanic, resulting data projections may mischaracterize and misreport viewing in Hispanic households.

In addition, currency measurement providers weight and project based on demographic assignments in order to ensure proper representation by demographic characteristics. Errors in demographic assignments can result in misallocation of weights, which will skew results.

Recent studies have identified significant error rates running across demographic assignments for different identity providers, with identity providers less likely to successfully match multi-cultural audiences than the general market.[1]

It is also important to note that privacy concerns are increasingly likely to impact the ability of third parties to provide accurate demographic data for identities. A dozen states have enacted privacy laws that go beyond Personally Identifiable Information (PII) to encompass Sensitive Personal Information (SPI), which includes data on race and ethnicity, limiting the availability of these characteristics for demographic assignment on identity graphs. Identity providers also have to treat data about children and teens differently, which can limit their ability to resolve viewing. The presence of children and teens by age band is treated as a household characteristic, whereas demographic and other data on specific persons is carried for persons 18+. This means that data on kids and teens is less robust than data on persons over 18. Identity providers built on credit card data also have limited information about younger viewers, who generally don't have credit access.

Although all currency measurement providers work hard to improve the quality of their identity graphs, some of these issues are difficult to resolve. At a minimum, data users need to interrogate the identity graphs of their providers, to assess potential biases and holes.

[1] Studies using Truthset data have served to quantify that the error rate by demographics across individual identity providers is non-trivial. Of note is the Sequent Partners White Paper for the ANA's Alliance for Inclusive and Multicultural Marketing, Addressing Biases in Multicultural and Inclusive Identity Data, from 2021. Sequent found that members of diverse populations were less likely to resolve in a match with identity providers than the general market. In addition, a 2023 study by CIMM and Truthset revealed the extent to which error can arise in individual identity graphs.

The customer perspective

For measurement customers, the key priorities for identity are representativeness and consistency. Identity data provides the makeup of the viewers in each data source and helps to identify any duplication between data sources (e.g. targeting data vs. viewing data). While measurement vendors are focused on determining how best to assemble and combine various data sources and how to project to the full market, their customers generally have two priorities:

1. First, each deterministic data footprint could potentially provide target matching opportunities. The larger the data footprint, the larger the number of potential matches will be. If the deterministic data medium came from a vendor with addressable media selling capability (like Roku, Samsung), this provides sellers and buyers a census like environment within that footprint as these addressable media types are not available elsewhere. However, these data and media products will be limited by their data coverage and scale.

2. Second, measurement customers will require guidance on how target matching and composition will perform in the measurement areas not covered by the vendor's deterministic data sources. Some of the determination would need to be modeled to estimate potential viewership and targeting audience's composition.

By determining best practices to profile total and targeting viewers within and out of data footprint, users can determine how the combined medium performs holistically and if the output is representative of the targeting expectation. Obviously, the deterministic data in general would provide greater degree of 1-to-1 matching capability and better targeting accuracy. Measurement vendors that have larger resources to provide more deterministic capabilities compared to relying on modeled missing pieces would be viewed more favorably.

From a targeting perspective, the two main approaches are traditional age/sex demographics and advanced audiences. While each measurement vendor has the capability to produce both types of metrics, their results tend to vary greatly due to their different data and approaches. To better validate the targeting process, traditional age/sex demographics often get some level of validation from third-party data sources on a macro level (e.g. US Census) or micro level (e.g. panel survey questionnaire).

Even with some validation in place, results will vary but could confirm some general common-sense insights (e.g. ESPN is a very male skewing network on one end of the spectrum while Lifetime is on the other end of the spectrum).

2. Critical methodological challenges facing big data measurement solutions

This level of consistency is very important for sellers to know what they are selling is within their delivery expectations. Advanced targeting presents a different challenge. The data that the advertiser wants to use to support targeting against their intended audience could come from their own customer database or CRM system. While agencies could match this data with one of the measurement vendors for planning and activation, the match rate and scale would play a significant role. For traditional linear TV networks, many of their legacy selling model still depends on individual program level performance. A CRM based target is often difficult to gauge without historical trend data to help with estimating program level audience size and skew. This inability to expect a certain level of consistency hurts sellers to accurately assess their transactional performance. In addition, sharing the CRM data with publishers without privacy protection can yield significant security concerns.

The distribution of video content varies across TV and digital platforms. For digital activation, marketers can activate a select set of audiences on various digital platforms. If there's an ID match to activate, then a transaction

can be fulfilled, and the target audience is reached. In a linear TV environment, networks need to estimate the likely size and characteristics of the viewing audience in advance of the actual airing, to determine the appropriate pricing and inventory information. Previously aired programs are typically used as a proxy to match against the target identity set to create an estimated target composition and size for pre-buying transactions.

Once the ad unit is purchased and aired, the stewardship process can prove challenging. The purchased program where the ad unit ran will need to produce the actual target audience to compare to the estimated selling inventory audience. For traditional age/sex demo, historical trends can provide a consistent view on what the expected variances it would be. For advanced audiences, this means another round of identity matching with potentially different estimating targeting and viewing data source, leading to greater degree of stability compared to traditional age/sex demo estimating process. While some advanced targets are based on client's CRM data, some commonly used advanced targets are based on 3rd party metadata source like JD Powers or Experian.

These targets often provide psychographic data like "truck purchase intender" and can be easily subscribed along with the currency measurement data as an add-on. However, just because the targeting data came from a single source, each measurement vendor could provide different levels to represent the target audience within their own unique footprint with varying differences. All subscribed parties from buyers and sellers as well as marketing clients could access this data and share the same viewership data patterns .

Beyond challenges in striving for representativeness and consistency in the identity related space, there are additional outside challenges prior to the steps of evaluating identity fidelity. If the advanced target originated from client sensitive data like client app usage information, there will potentially be a lot more legal restrictions on the parties that could access the data. For pharma clients, there might be even more government regulations to prevent the use and sharing of such data for targeting. Additional privacy regulations could also provide more challenges. The question then is what will the advanced target be used for and how representative is the output?

Even if the targeting data is approved for use, will it be used as an aggregated level which could provide viewership propensity level insights? Or will the data be granular enough for programmatic 1-to-1 matching activation? Privacy is a hot topic for consumer and many World governments are taking different approaches to regulate user privacy. Currently in the US, there is strong support in Congress on a bipartisan basis to outlaw behavioral targeting.

2. Critical methodological challenges facing big data measurement solutions

If that passes, not only will it create a lot of uncertainty in the digital space but how will it impact advanced targeting for advanced television approaches.

The increased usage of internet connected devices like Smart TV and OTT boxes creates more passive digital signals for measurement and the shift of usage level has not slowed down. Unlike a panel approach which relies on respondent active participation, simply having a device like a Set-top Box turned on and running provides an event level electronic signature at various data transmission points to determine viewership:

- For content assignment, the time and date stamp along with transmission source on the digital event can identify the content associated which can be aggregated into the overall viewership metric.
- For audience assignment, the device level identification along with metadata like IP address can provide household level information which in turns helps with audience assignment and viewership.

The biggest challenge in this area is finding a level of consistency across vendors to accurately represent the targeting data viewership profile.

While there are multiple sources of truth from different TV measurement vendors using their own unique identity process, traditional digital measurement (with their streamline delivery process) provides a more consistent metric that can be used for various attribution evaluation. If the TV viewership data becomes more consistent and representative, they can join traditional digital media in the attribution process and potentially connected to other marketers KPI including sales and responses.

Challenge #2: Addressing footprint coverage bias

The measurement vendor perspective

Measurement vendors face three critical challenges in relation to footprint coverage:

1. **Coverage of projectible and representative population:** To what extent do the footprints omit segments of the population and how different is the behavior of these segments from the segments that are captured?

2. **Footprint biases:** How can adjustments and bias correction be introduced to correct for the design biases in the big data footprint? (That is, how to account for the differences in behavior between household types covered in the footprint, versus those segments systematically excluded or underrepresented by design.)

3. **Geographic coverage:** How representative is the big data footprint geographically?

Coverage of projectable and representative population

Trying to estimate viewing of linear TV channels using data from Set-Top Boxes and Smart TVs create significant methodological challenges, inherent to the big TV data footprint. Potential data gaps include:

- **Over the Air viewing:** Although Set-Top Box data does include broadcast viewership, Set-Top Box households tend to spend less time watching broadcast channels than households that only have access to broadcast or Over the Air (OTA) channels (i.e. households that watch broadcast TV but do not subscribe to an MVPD service). Measurement vendors need to adjust their viewing estimates to account for the differential rate of broadcast viewing in STB households. Smart TVs do capture OTA viewing, but the universe of OTA households with Smart TVs may not be representative of the entire universe of OTA households.

- **No-Wi-Fi households:** Both Set-Top Box and Smart TV data require an internet connection to transmit data back to the MVPD or manufacturer. Households without Wi-Fi will be excluded from Smart TV and STB footprints.

- **Differential Demographic composition by TV household type:** Different demographic groups, and in particularly diverse or multi-cultural viewers, are differentially represented in Smart TV, Set-Top Box, OTA, and broadband-only footprints. Properly representing[2] their viewing behaviors can be challenging.

[2] Shimmel and Broussard for CIMM, "Best Practices in Combining Smart TV and Set Top Box Data", 2021

- **Digital (off-set) viewing:** Big data footprints consisting of Set-top Box and Smart TV data are limited to measuring viewing on TV sets (or, in some cases, via devices connected to the TV). However, it is now well established that a significant share of viewing occurs off-set, on phones, tablets, computers and other devices.

Footprint biases

In a sample-based methodology, the sample frame is generally constructed to encompass or be representative of the universe of interest. Of course, much can go awry between frame design (the pool of households or persons originally selected to be in the sample) and in-tab sample (the pool who ultimately end up providing data), but at least the opportunity exists to design a sample to be comprehensive and representative. With big data assets, the measurement vendor has no control over what is provided and has to determine the characteristics of the universe.

In many cases, big TV data assets are biased by design: Set-top Box data excludes households without cable or satellite and tend to be highly

geographically clustered; Smart TV data excludes households without Internet access and households that have not connected their Smart TV to the Internet. Also, consumers have the right to opt out of data sharing in both Smart TV and Set-top Box footprints, resulting in another potential source for bias, Of course, it is also difficult to imagine households that are concerned about their privacy signing on for a metered panel.These biases in what are effectively convenience samples are not dealbreakers though. The benefits of big data make the juice worth the squeeze. The biggest challenge is projecting beyond the footprint.

Measurement companies devote much of their time, attention and resources to developing (often very sophisticated) methodologies, often using advanced techniques and technologies (e.g. data science, machine learning) to correct for biases in the data they do have and to develop these projections.

One strategy for addressing footprint biases is to use like to represent like, using the available data to develop estimates for viewing on devices or TV platforms that are not represented in the available big TV data sets. For example, one cable MVPD's data could be used to represent only that MVPD, or it could be used to estimate viewing on other cable MVPD footprints. If the latter, some technique is necessary to account for differences across MVPDs in making the cross-MVPD projections.

Another strategy is to use different overlapping footprints to understand the gaps in each data set and to statistically adjust accordingly. For example, a company with a footprintencompassing both Smart TV and Set-top Box data could use the known, identifiable overlap between the two data sets to understand viewing in other universes.

A third strategy is to use weighting, calibration, or both to adjust for footprint composition issues. Several measurement companies are now using the ARF's DASH study to provide weighting targets for technographics, the presence of or access to different devices or technologies in the household (e.g. Set-Top Boxes, cable providers, Smart TVs, use of streaming services, etc.). Big TV data footprints can also be calibrated based on panel data. Measurement companies can also use survey data to understand the differences in viewing between footprint and non-footprint households and to adjust accordingly.

It must be noted that devices, households and services that do not produce data are by design omitted from big data footprints, and this is a non-trivial issue. Over-the-Air (OTA) households are not represented at all in Set-top Box data and are not entirely represented in Smart TV data sets, as OTA-only households are less likely to have internet access. OTA households without internet access remain a hole in any big data footprint.

No-Wi-Fi OTA households are a relatively small slice of the universe, and there are techniques to assure proper representation of OTA viewing through weighting. However, given the disproportionate share of broadcast station and broadcast network viewing accounted for by no-Wi-Fi OTA households, it is essential for measurement companies to adopt a rigorous methodological approach to assure representation.

ATSC 3.0 or Next Generation (NextGen) TV introduces the prospect of return-path-data to the viewership of over the air TV, but only for connected devices, and is likely to become a valuable source of big data specifically for broadcast stations and networks. However, penetration will take time to grow, and the use of this data will present new complications, given likely overlaps with other datasets.

Geographic coverage

Smart TV footprints and STB footprints might not be evenly dispersed across TV markets, county types, and regions. This is a particular issue with respect to cable MVPDs, the footprints of which are "clumpy" (i.e., present in the markets where the MVPD has an infrastructure, but absent in others.) Satellite STB data is less "clumpy" – satellite providers have customers in all TV markets and virtually every county in the US – but these footprints tend to skew toward C and D counties.

The customer perspective

For customer, biases in the available data are a critical issue. Data diversity and completeness to achieve consistency are important considerations.

Measured data is inheritably biased. One way to address this is to leverage diverse data sources to balance and address the overall bias issues. As audiences are using more diverse ways to access their preferred video content, measurement solutions need to reflect this diversity in the range of data sources they use.

Different data sources have different identity makeups, showcasing their data bias and skews. For STB data, there are thousands of smaller cable operators spread across different US regions, often serving as the sole provider of traditional cable TV in a given area. Even with many consolidations over the years, there are regional skews based on their underground wiring infrastructures.

These skews create differences in their demographic makeup as well as viewership. Beyond traditional wired cable companies, customers could also get traditional cable packages from satellite operations like DirecTV/ DISH or one of the telecommunication companies like Verizon Fios or AT&T U-verse. While satellite companies offer national distribution, their customers are more skewed towards suburban counties since most urban environments

like high rise apartments face installation challenges. An example of a wired cable (or MVPD) bias, AT&T U-verse has a large presence in the Southeast states where their audiences are passionate about College Football. If only U-verse were used as a data source, their overrepresentation of those regions in people and viewership would not be a good proxy for another area of the country or the whole country. Even when more diverse data sets are added to offset this bias, there will always be some micro local level which is outside of statistical significance. In the vendor section, over-the-air viewers were mentioned, representing a sizable viewing audience with its own distinctive viewing behaviors, making it difficult to model using other TV datasets. If vendors choose not to address this issue, their estimates for viewing across the full marketplace will be incorrect.

Data bias also extends beyond viewership data. The way identity-based targeting data like Experian are collected from a specific approach and source which can also create inherited bias. Credit bureau-based data like TransUnion tends to skew suburban while mobile-based data like Neustar skews more urban. Even the US Census data is considered biased towards minorities and undocumented populations due to their measurement approach. Undocumented people had traditionally avoid Census measurement which is often viewed as a data source to inform deportation decision. As a result, lower participation yields an incomplete Census picture that is often used as a main standard for universe measurement.

When marketing clients are matching their consumer identities to one of these media consumption data sources for targeting, they are often facing the challenges of low match rate due to various factors including privacy opt-out, media data not having full US coverage, or legal restrictions on digital customer tracking. These customer identities, especially with pharma clients, also carry significant legal responsibilities which make them challenging to share at scale for transactional needs. The data broadness with an eye for representation helps mitigate many biases. While it may never eliminate all bias, the acknowledgement and understanding of what's missing helps fill in the gap.

A data source with a combination of Set-top boxes (cable and satellite) across many regions, Smart TV across various manufacturers and price points, and digital footprint to cover streaming would be a good start. Calibrating them with another source of truth (which could also be bias) helps with the representativeness of the data. When the future ATSC 3.0 standard is rolled out, Smart TV can add some much needed Over-the-air coverage. Until then, be mindful of the completeness of data.

Challenge #3: Onboarding, Cleansing and Combining of Big Data Assets

Data cleaning and editing is necessary in all currency systems; however, big data assets are typically not specifically designed for the purpose of audience measurement, and different sources provide data in different formats. Alignment and preparation of big data for specific measurement and reporting is a larger task than performing the same functions on smaller, panel data.

In addition, for currency providers using multiple Smart TV manufacturers or using at least one Smart TV manufacturer and Set-top Box data, the different footprints must be combined and deduplicated. For example, a household may have a Samsung set, an LG set and a Vizio set, each connected to a Comcast box. The measurement provider must have methods to ascertain that these devices are in a single household, and to avoid double-counting (i.e. tuning captured from both the set and its associated Set-Top Box).

The measurement vendor perspective

Measurement vendors must focus on the following areas with regard to onboarding, cleaning, and combining of big data assets:

4. How are data sets from different partners in different formats made uniform and combined into a single unified footprint?

5. What QA, editing, and data checks are necessary to remove as much error as possible in translating raw data into usable projections?

6. In cases where devices may end up in the vendor footprint from multiple data partners (e.g. Smart TVs and Set-Top Box es from the same household), how is the device data combined and deduplicated?

7. How disruptive is the inevitable addition or subtraction of a data partner to reported estimates?

8. How are data outages addressed? For example, a provider with a footprint comprised of MVPDs A, B and C, might underreport viewership if MVPD A goes dark; similarly, they might overstate viewership if MVPD E goes dark, unless their projection system takes this into account.

What are the practical implications of these challenges:

1. **Data integration across footprint partners:** For currency providers using multiple Smart TV manufacturers or using at least one Smart TV manufacturer and Set-top Box data, the different footprints must be combined and deduplicated. For example, a household may have an LG set and a Vizio set, each connected to a Comcast box. The measurement provider must have methods to ascertain that these devices are in a single household, and to avoid double-counting (i.e. tuning captured from both the set and its associated Set-Top Box). The incidence of a single household with multiple MVPDs (say subscribing to both cable and satellite) is sufficiently rare that it need not be addressed.

2. **Projection methodology accounting for potential addition/ subtraction of data sources:** As noted above, when a currency built from MVPD and Smart TV data adds or loses a partner to the footprint, the data inevitably shifts. Questions around the long-term availability of Inscape/Vizio data for use in currency solutions serves to underscore that currency measurement providers are well-advised to develop systems that take partner churn into account. Systems and methodologies should be robust enough to account for the loss or addition of a new footprint partner with minimum disruption.

3. **Key Considerations in onboarding:** Data cleaning and editing is necessary in all currency systems; however, big data assets are typically not specifically designed for the purpose of audience measurement, and different sources provide data in different formats. Alignment and preparation of big data for specific measurement and reporting is a larger task than performing the same functions on smaller, panel data. Key considerations in onboarding data include:

    - **Data structure:** There is not a single standard or template for viewing data; different data partners provide databases in different formats. Also, Set-top Box data is inherently different from Smart TV data (the former comes with channel tuned, and schedule data enables translation from channel to the more granular program; whereas Smart TV data has content identified via fingerprint match, and the content must be matched back to channel/source).

    - **Missing records:** In addition, big data assets require cleaning to adjust for missing records, incomplete records, duplicate records, and other types of data issues.

- **Deduplication:** If a measurement provider combines Smart TV and Set-top Box data, or combines multiple Smart TV footprints, the data must be combined into households and deduplicated.

While data users may tend to believe that if all providers used the same data assets, the results would match, the measurement companies tended to stress that when it comes to turning big data into actionable audience measurement, the devil is very much in the details. Put simply, when you do bring in disparate data sources, how do you put them together?

Different data assets have different structures, data, and problems. So, in a big data world, we can't do statistical inference from a sample, so methodology is just as important as data quality and data provenance. Data ubiquity doesn't negate the need for methodological excellence; if anything, it makes such excellence even more critical.

One measurement company executive stressed the need to continue pushing data providers to produce the best data possible. Given that their data is a component part of services that are or will be in the MRC process (or are already accredited), it is essential that big data providers do as much as possible to take appropriate data standards into account in preparing the data for distribution.

4. **Standards and Guidelines:** The MRC's Multi-Channel Digital Video Data Capture, Accumulation, and Processing Guidelines, published in 2012, remain in play, and the user of big data currencies might want to review these. Thus far there aren't standards specifically for Smart TV data capture/accumulation/processing, although many of the specifications in the aforementioned standards are relevant; and several subsequent sets of MRC standards have relevant content (e.g. section 4, Cross-Media Measurement Standards (Phase I: Video). There are also no published standards yet for best practices in combining Smart TV and Set-top Box data, although CIMM published a paper on Best Practices in Combining Smart TV and Set-top Box Data in 2021.

The customer perspective

Customers are looking for viewership metrics that is embraced by all the parties (sellers, buyers, marketing clients, procurement) using the data. The user community understands there will be differences in the reported number amongst the vendors but all of them should have correlated historical trends on total viewership level.

After vendors have identified their data makeup and combined them with bias mitigation, this is where we see the first part of how the data is projected on a totality basis. There's no specific targeting in this section and the focus

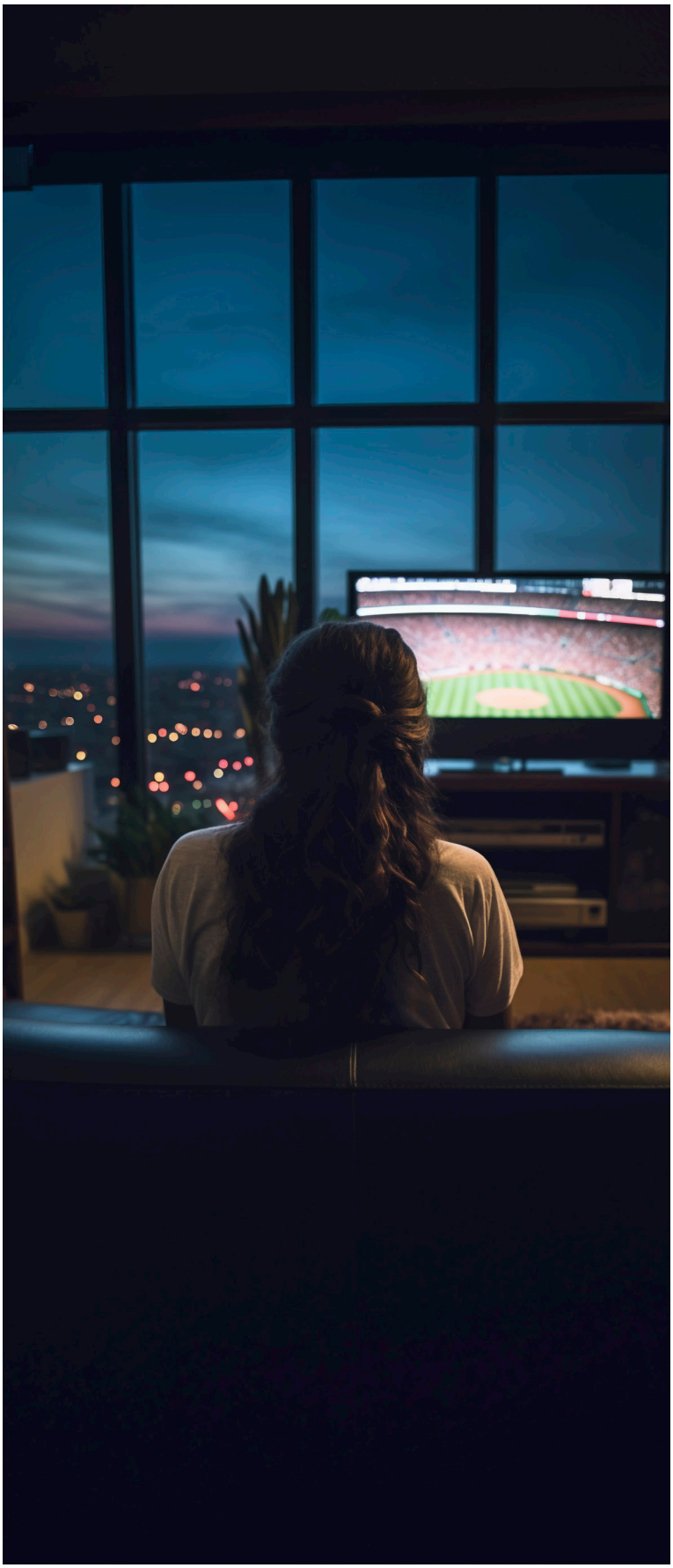
is on the total household or persons level to inform the total volumetric of TV viewership. This is the first opportunity to examine which data sources are reported and how they compare to others. With each viewership metric, there are many contextual layers and viewership definitions. Will the metric reflect content vs ad measurement? Live vs delayed? What are the basic Impression level definitions – average commercial minutes (C3/C7) or exact seconds of ad viewership? What does the measurement actually cover in terms of distribution? With a sizable audience of cord-cutters using vMVPDs, will these customers' linear TV content viewership be covered?

The answer to some of these questions could make a big difference in how data is reported. For example, regionally distributed programs that are sold nationally are often difficult to reaggregate from the bottom up. Syndication programs might be easy to identify in general but combining different airing times in different geographic locations could prove challenging. Similarly, NFL programs on Sunday mornings are grouped into 3 distinct major broadcasting windows between CBS and Fox that are sold at 3 different unit costs.

While one can compare between different measurement services using total day network level, it is more insightful if buying packages are defined with specific dayparts and program filters. For example, separating primetime news, entertainment, and sports programs at ease is necessary for TV networks to identify viewership levels and audience profile.

Keep in mind this is the first step applying a calculation methodology to estimate the topline total audience for each TV network. If that approach changes due to projection techniques, data access, and data editing rules, this could make comparison between vendors very difficult and frustrating. For example, STB editing rules differ greatly between all the companies. Each has their own unique perspectives which yields different results.

While each measurement company produces different levels of viewership for the same content inventory, the period-over-period viewership trend should be more in line between the companies. If one vendor is showing a larger erosion than the others, users must examine the cause of this. Ideally long-term review of over five years of viewership trend will provide the necessary benchmark for assessing consistency. Most emerging measurement companies currently do not carry more than two years of data or, if they do, their datasets are often non-comparable, due to significant changes in methodology and/or the addition of new data sets. This lack of stability negatively impacts projectability, resulting in low level of quality assessment.



Challenge #4: Metadata

The measurement vendor perspective

Measurement vendors need to address two key considerations in relation to metadata:

1. How good are the start and stop times for programs as aired (as opposed to as scheduled)? Does the vendor incorporate actual as-run start and stop times before reporting? How accurate are these?

2. How good a job does the vendor do in properly identifying ad occurrence– including brand, time of airing?

Metadata – data associated with and describing data and other assets (ads, programming) – is an essential component of audience measurement data creation. “Then there is the perpetual problem of metadata,” said one measurement executive, “to my mind an unsolved problem.”There are two general types of metadata in play: content identification and taxonomy; and ad identification and taxonomy. There is also data on genre and other types of descriptors of content that are useful in deploying measurement data

2. Critical methodological challenges facing big data measurement solutions

(e.g., a media planner might wish to recommend programs in one or more specific genres).

Within the scope of both types of metadata is time of airing. For example, a TV network may show a football game scheduled to air from 4PM to 7PM; but the game might actually run to 7:25. If the measurement provider reports program ratings based on scheduled air times, they will get the audiences wrong for the game, and for the rest of that evening's program schedule (i.e. a show that actually ran from 7:25 to 8:25 will end up with the ratings from the minutes of 7:00 to 8:00.). So as-run schedules are essential to creation of accurate ratings. These can come from the programmers themselves, or from a third party. For ads, while average program commercial minute (ACM) remains the transactional metric, big data supports the use of exact commercial second reporting, making precise detection of time of airing vital to campaign reporting (and even in an average program commercial minute environment, the start and stop times of programs as aired must be accurate to get ACM right.)

While there have been numerous initiatives to standardize taxonomy for both content and

ads, an industry consensus has not yet emerged. As a result, the sources of data on ad occurrence, schedule time, program identification, and network/channel/source identification are variables that can affect accuracy, and that drive some of the differences across services.

Measurement providers may rely on first parties (programmers, ad logs) or third parties (schedule providers, ad occurrence providers); or they may maintain their own libraries of both content and commercial occurrence.

The customer perspective

For customers, program inventory availability completeness and cross platform readiness are critical issues, impacted by the availability of metadata.

After getting the larger holistic view of what each of the measurement companies produced, there will be expected differences between them due to incremental and supplemental data sources, cleansing and aggregation techniques. How these approaches are applied could determine the degree of variability.

Earlier in the paper, we identified several areas where big data could have difficulties in estimating missing elements. The use of additional data sources could help fill those gaps but not all of the companies will use the same data sources or even attempt to account for them. The issues of Out-of-Home and Co-Viewing often come up as incremental audiences that should be estimated for TV networks to monetize them.

Since most Americans have more than a few combinations of devices, a cross platform device or media approach is needed to isolate the unique audience for targeting purposes. Knowing how each audience is exposed to a message could help tailor the effectiveness of a media campaign. Someone with a cable STB and a Smart TV could provide duplicate signals when content is viewed. Deciphering different devices and how they fit into an overall consumption pattern in a cross-platform environment is one of the most difficult challenges in the media industry today.

Even when that unique audience is identified with their media consumption habits, getting all their attributes for targeting purposes is also another challenge. Traditionally, age sex demo was the common viewership metric TV networks can sell in a package to advertisers. Some agencies might tailor their plans using third party research studies to determine audience skews to make media plans more efficiently and effectively targeted.

As more marketers have more attributable information about their consumers, there is a greater emphasis to directly target these audiences similar to the digital media approach without subjecting to the traditional demo metrics. Data from credit services

and various identity solution companies have created more opportunities for the media industry to learn more about all US consumers especially when they could be targeted digitally. While this approach was often utilized in digital marketing, their ease of use to access advanced targeting has made it attractive to use in traditional TV environments.

In today's environment, TV programmers still require their programs to be measured with as much completeness as possible. Without representation, networks cannot sell their inventory to buyers at scale. Syndication TV or regional distribution programs like the NFL Sunday programs often have program assignment issues by big data providers. Their ability to correctly aggregate viewership on a programming level as well as other viewership sources (like OOH, Co-Viewing, and vMVPD) will be a key assessment consideration set.

Challenge #5: Integrating Linear and Digital (Streaming)

Generally accepted best practice for streaming measurement is to use some form of census collection; either pixel-based, or via server log integration. This requires the compliance of the streamer, the advertiser or their agency; currently there are no big data assets available that provide a horizontal look across all streaming platforms, which could be used to project audiences for both content and ads.

The measurement vendor perspective

When integrating linear and digital (streaming) measurements, vendors face five important challenges:

1. How much coverage does the vendor have in the streaming space? What are the gaps?
2. How are linear and streaming audiences combined and deduplicated?
3. If the vendor intakes a census data feed of streaming data from one or more sources, what QA methods are in place for cleaning and processing the data, and for assuring the integrity of the data delivery?
4. To what extent does measurement and reporting cover viewing both on TV and on other digital devices? What assumptions are made about viewing on TV sets versus on mobile and other devices? What are the rules for counting impressions for the different devices?
5. If streaming data is collected and projected based on a methodology besides census-level data, all the other QA issues associated with linear TV projection and measurement apply.

In video cross-platform measurement today, the elephant in the room is streaming. According to the July Nielsen Gauge, streaming content viewership on TV now accounts for over 40% of viewership[3]. There is no streaming data in Set-top Box data and an insufficient amount in Smart TV data, and a lot much streamed TV is consumed over devices besides TVs. As a result, it has become generally accepted that the state-of-the-art technique for measuring streaming– for both content and ads– is through a census-based technique (either from pixels or server logs.)

While Smart TV data could provide at least a partial solution to streaming measurement, a preponderance of streaming comes from providers who require via license that the Smart TV manufacturer not run ACR while the native app is engaged. This does not prevent the use of ACR when the app is running on an external OTT device (e.g., a Roku device linked to a Vizio TV set). However, even if a Smart TV provider picks up streaming from external devices, they would need to fingerprint the original content libraries of streamers in order to recognize the content.

If ads in a specific campaign are fingerprinted, either by the smart TV manufacturer or by the measurement company, then ad detection is possible even without identification of the content in which the ad airs.The availability of and need for a census of streaming data creates a paradoxical situation. Theoretically, all the data is available without gaps or coverage issues. In practice though, the reliance on data owners for such data creates a

[3] https://www.nielsen.com/data-center/the-gauge/

measurement checkerboard, wherein different currency providers will have access to different streaming data assets. So, measurement providers must confront an environment where they have ALL the data, for SOME of the streamers.

Thus far, census streaming content data has not been made available for syndicated reporting. The JIC promises to remedy this by creating a data asset comprising a census of content and campaign data, made available to certified providers (hopefully for syndicated content reporting). However, thus far this data asset is slated to include only JIC participants (a material improvement in the availability of content data, but not addressing the issue of the checkerboard).

One measurement executive stressed the fact that streaming will, for the foreseeable future, be based on OPD (Other People's Data). Meaning that there isn't currently a streaming analog to the use of Set-top Box or Smart TV data for measuring linear TV; streaming is measured at the pleasure of the streamer. This same exec notes that privacy concerns could lead to an environment in which streaming data will be processed via publisher-specific privacy-

compliant intake via clean rooms, wherein the vendor does the integration. (While we do not address the topic of clean rooms herein, clean rooms are an important tool in building cross-platform currency solutions. Similarly, the WFA/ANA Virtual ID is another tool for developing cross-platform reach and frequency).

This will mean that while all providers (or, in the case of the JIC data, all certified providers) might have access to the same data, the creation of overlap metrics requiring parsing exposures into reach and frequency will be dependent on the quality of the vendor's identity graph, and their methodology for data integration (whether or not the integration takes place in a clean room environment).

In general, the measurement providers somewhat resignedly accept that access to streaming data will be incomplete. One cited a "best data available" strategy; census-based where available; panel-based if available where census data is not; and projection and modeling beyond that.

Strategies to address this generally include using some sort of panel for calibration to account for exposure to streaming where census data is not available. These panels could be virtual panels, constructed from identities (or synthetic identities), with data science and AI used to fill in streaming behavior for virtual panelists to streamers who do not make data available.

The customer perspective

Within this space, the key parameters are integration coherence and identity match rate.

For the future of media measurement, coverage to account for streaming and digital video viewership is vital for the industry to account for all media behavior holistically. While some streaming platforms created measurement silos within their footprint and provided very limited data sharing, the industry needs to increase their effort to account for many of these emerging viewership options as they could be potential marketing investment venues.

Currently, none of the big data sources (OEM or STB) provides native app streaming viewership data at scale side-by-side with linear TV content viewership. Part of this is due to legal limitations native apps impose on OEM devices to carry the app on their platform. Some of the measurement companies combine campaign level digital ad tagging approaches with their linear TV coverage to produce a cross-platform solution. However, this approach merely focuses on post campaign delivery on ad exposure which does not provide holistic pre-campaign level insights for investment strategy based on content.

For the industry to embrace digital delivery from multi-currency companies, they need to provide content level

viewership as a proxy of ad exposure. Understanding the scale of viewership side-by-side with linear TV viewership would help marketers understand the relationship between the platforms and determine strategies to solve for their marketing mix. In the meantime, the industry is left with metrics that cover linear TV content only which is only half of today's TV consumption pattern.

Since digital vendors have the most direct path to account for viewership on their platforms, getting their 1st party data incorporated would greatly enhance digital viewership in those areas. Amazon Thursday Night Football currently provides their 1st party data to Nielsen and iSpot as part of their 3rd party measurement agreement. More measurement companies and digital streaming vendors should work together to account for all viewership sources to provide the marketplace with a comprehensive picture of TV viewership activities. Meanwhile, the industry should provide standard guidelines to ensure some level of consistencies in streaming viewership reporting. When combining streaming and linear TV viewership, unique viewers across the two platforms need to be carefully identified as they have a big impact on reach

and frequency calculations. Failure to properly integrate the two forms of medium could potentially inflate reach and understate frequency.

Challenge #6: Processes and Methods for Addressing Coverage Gaps and Shortfalls

With naturalistic big data assets, you tend to get what you get. Set-top Box footprints, for example, contain neither Over the Air nor stream-only households; Smart TV footprints contain both, but in a different proportion than in the universe.

Diverse populations, particularly including Latinos and African-Americans, tend to have a lower incidence of representation in big data viewing assets. This is driven by both differential rates of penetration of Smart TVs and Set-Top Boxes by race and ethnicity; and, by differential match rates by race and ethnicity in identity and demography sources (See the ANA AIMM paper referenced above).

ince big data-based currencies are derived from device-level data (Smart TVs, Set-Top Box es), typically viewership data has been compiled with the household as the unit of projection Persons-level data is typically created by methods that have come to be called personification, and typically involve some algorithmic assignment of viewership in a household to the known persons in that household. Sometimes panels are used as a training set for this algorithm. Persons-level behavior is not directly observed from within the projected footprint, requiring modeled inference. This technique may or may not be better than direct collection via people meters in panels, but it certainly is different, with far less methods-based resear ch behind it.

The measurement vendor perspective

Measurement vendors need to address four measurement challenges to deal with coverage gaps and shortfalls in big TV data sets:

1. **Personification:** How does the vendor personify device-level and household-level data to create persons estimates? Do panel estimates come into play at all, either for training, calibration, or to actually provide the VPVHs? If yes, what issues are associated with the panel that might introduce bias?

2. **Diversity:** How good a job does the vendor do of representing different diverse (and sometimes difficult-to-measure or identity) populations?

3. **Data gaps:** What edit and other data adjustment rules are in place to account for missing data and other anomalies? How does the vendor treat the phenomenon of Box-on/Set-off in Set-top Box data? How does the vendor treat partial set coverage within households in a Smart TV footprint?

4. **Out-of-home viewing:** How does the vendor measure out-of-home viewing, and how is this viewing combined and deduplicated with in-home? Is guest viewing within a household treated as in-home or out-of-home viewing? What defines viewership in out-of-home measurement?

Coverage and design challenges create various issues for big data-based measurement providers and, indeed, other measurement vendors:

1. **Personification:** Device-based measurement of TV viewing has required personification in the US since the very first use of TV set meters. Prior to the development of people meters, household viewing was personified based on a separate, set-based paper diary sample wherein each household member entered their viewing for that set. Currently, there is much debate about the relative efficacy of personification techniques; these appear to accrue primarily in the variance of personified persons ratings from big data currencies compared to expectations developed based on historical panel-based people meter data.

In order to report persons-based as well as household-based viewing, big data providers muse develop a technique for attribution of viewing to specific persons within a viewing household. The measurement provider will have available viewership data within each individual household in the footprint, and a roster or set of identities of household members, with demographics. The task of personification of a big data asset, as performed today, typically involves a probabilistic assignment of the likelihood of viewership of each person in the household, to the content viewed on the screen.

There is enough data in a big data footprint that an assignment model may be created simply using the footprint and observing the differences in viewership by household demographic composition (the incidence in the footprint of a household with kids 2-11 watching a given kids show informs the likelihood that if that show is on screen and a kid 2-11 is in the household, that kid is watching the show.)

Another approach is to create a person-level viewer assignment algorithm, but to train the algorithm based on panel data. In such a scenario, questions arise about the quality and characteristics of the panel, and about how the measurement provider addresses these questions.

Finally, personification may be performed by applying panel-based demographic VPVH to a big data footprint. Ideally this would be done at the individual program/airing level, but viewership fragmentation and panel sample size essentially renders this impossible without collapses (e.g. by genre, daypart, and household characteristics; or over time.)

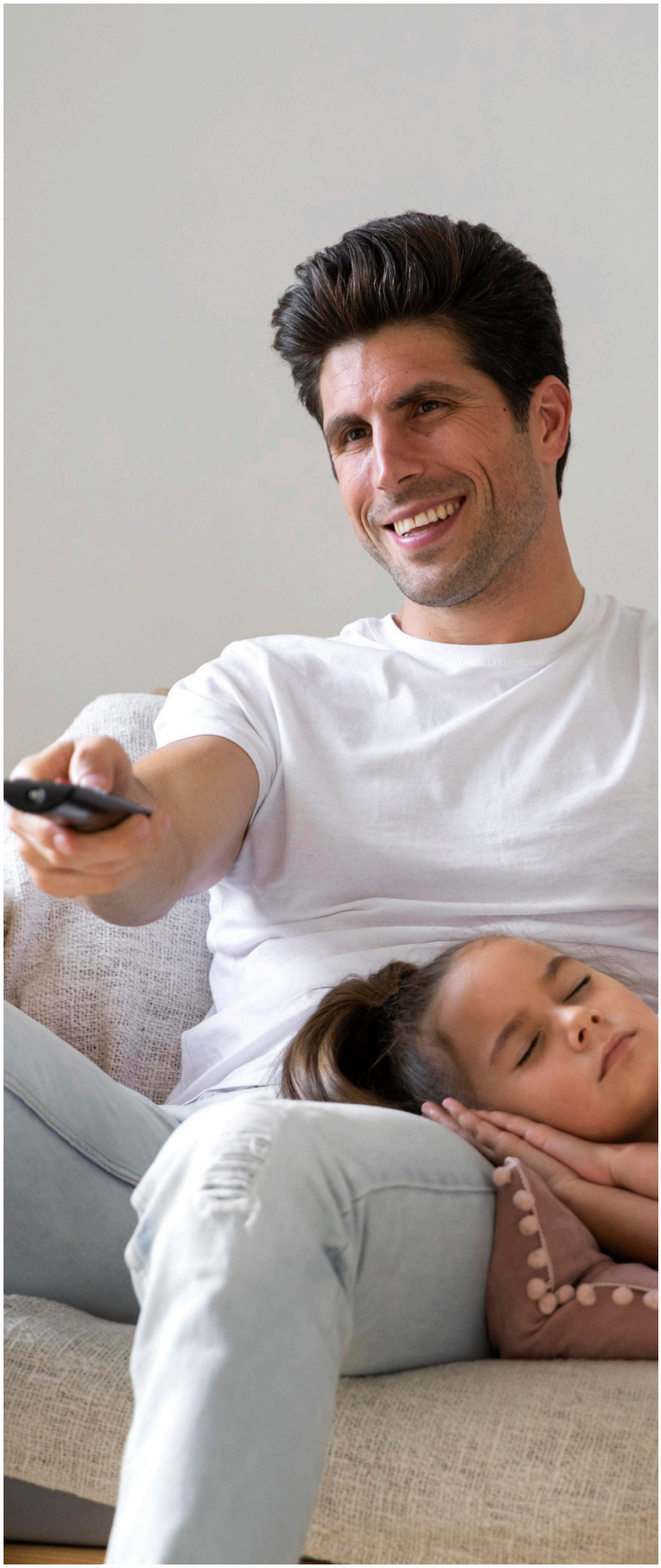
The advent of big data and identity have led to much science having been deployed toward personification. While variance from panel-based VPVHs may be disruptive, such variance may well represent progress. A clear theme across both data providers and data users was the importance of understanding and scrutinizing personification methods in order to use persons data for trading with confidence.

2. **Diversity:** Fragmentation in viewing isn't the only way fragmentation affects measurement. Fragmentation– or, more appropriately, diversity– among the measured population makes measurement more challenging.

Sample-based methodologies have always required special measures to accurately represent African-American and Hispanic populations, and 18-24 year-olds. Big-Data-based currencies have similar but different issues that require different solutions. For example, some diverse audiences may be more or less likely to have Smart TVs or Set-Top Boxes.

The ANA's AIMM group has demonstrated that different races and ethnicities have differential match rates within identity providers. Privacy legislation is making demographic data

increasingly difficult to procure for identity resolution. Different races and ethnicities have differential use of Over the Air TV, which as noted poses a challenge to big data currency providers, As audiences to specific programs get smaller, different content appeals differentially to different diverse population segments, making it essential to accurately represent all populations in the reported data– by race, ethnicity, language preference, sexuality, and even gender (as per one measurement executive, all legacy systems support only two genders.)



Probably the core challenge confronting vendors today as regards diversity is the increased requirement to accurately represent a more and more granular view of society, while at the same time privacy concerns are making the data to enable this, increasingly scarce. Almost certainly, data science, modeling, and AI will become essential tools in assuring that measurement keeps pace with the diversity of the American viewing audience. It will become important to understand the implications of this shift.

3. **Data gaps:** Big TV data sets may suffer from various gaps, which the measurement vendor will need to resolve – including:

- **STB:** Box-on Set-off: For Set-top Box data, currency providers must account for the situation wherein the TV is turned off, but the box is not. In these cases, the Set-top Box will continue to log viewing. Some methodology is necessary to truncate recorded viewing to account for this; otherwise, tuning will be dramatically overstated. Box-on/Set-off essentially requires an edit to create an end time. Methodologies to create these end times may be developed based on an analysis of matched pairs of TVs and Set-Top Box es (i.e. the set and the box associated with it); or by using a pool of Set-Top Box es that can be trusted to be off if the set is off.

- **Smart TV:** partial coverage: Smart TV footprints are typically either full or partial households of sets (a household with 3 Comcast boxes may have a Vizio, an LG, and a Samsung set.). Thus, some adjustment or projection technique is required to project the sets in the footprint to the entire set population. Without such an adjustment or accounting, estimates projected from Smart TV data will dramatically undercount viewing.

There are different potential solutions to these challenges. One is to use universe estimates for sets per household by demography, and to weight the Smart TV footprint accordingly. Another is to use the overlap of Smart TVs and STBs to develop an algorithm for appending viewing to Smart TV households in order to replicate what total viewing in these households would look like. It is important for data users to understand how currency providers adjust to account for this phenomenon.

4. **Out of home viewing:** Measurement of the Out of Home (OOH) viewing audience has become increasingly important due to the prominence of live sports in the TV ratings landscape, and the extent to which live sports audiences accrue out of the home (primarily in bars and restaurants.) Measurement of the OOH audience is currently done with techniques incremental to in-home audience measurement. It is beyond the purview of this work to delve into the respective merits of these approaches. Data users should note that since OOH is measured separate from the big data footprint, it is important to understand how the measurement provider combines or integrates OOH and in-home viewing.

The customer perspective

For customers, accountability and reliability are critical considerations. After initially reviewing the metric differences on a macro level, a targeted viewership in a select network will show more metric divergence between vendors. Even if the same program has the same HH level projected viewership, the person's level for demo or advanced audiences will also differ based on the meta data applied to derive the size and composition of the audience. In addition, local level reporting could face vast differences based on the data bias on processing approach.

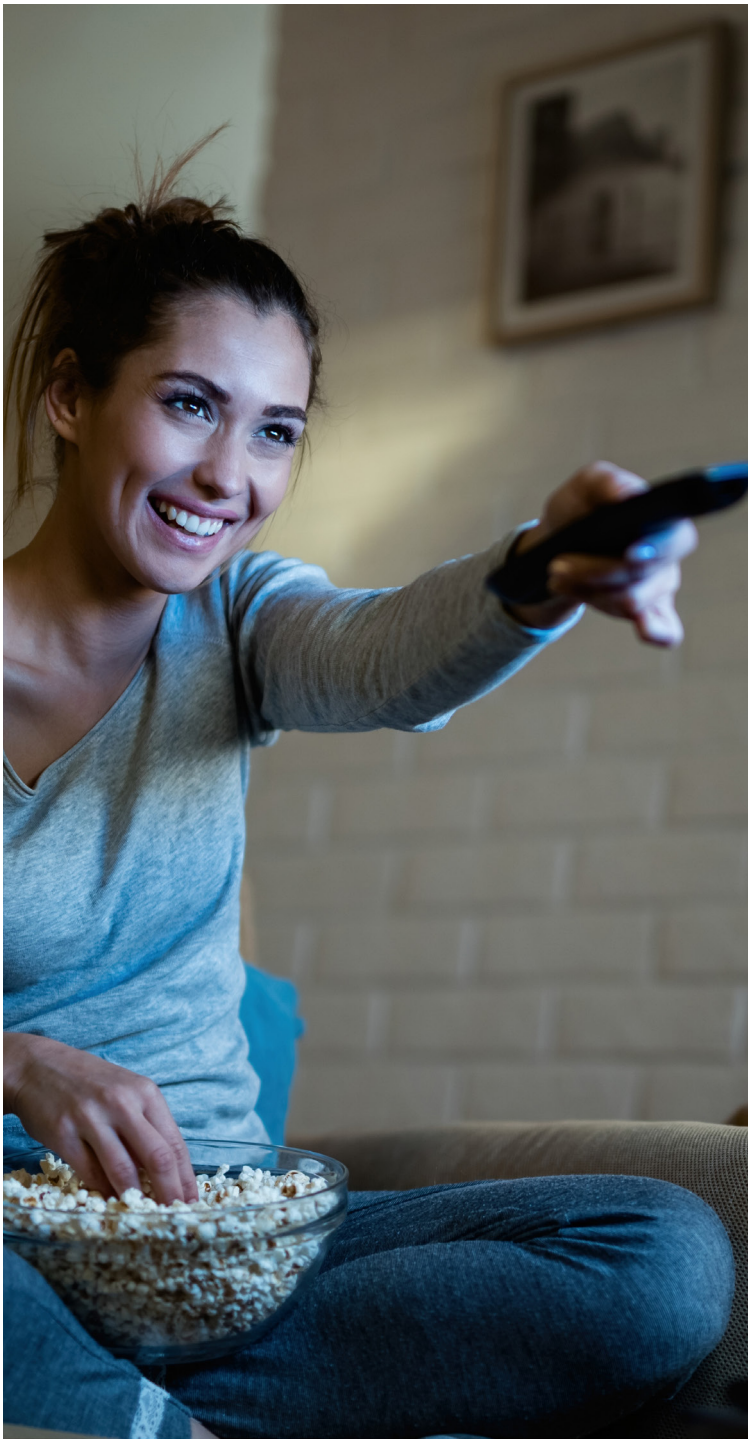
If a vendor is using a significantly smaller panel to perform personification assignment, the metric projection may not provide the wide enough coverage to report the variance expected. The lack of diversity in the data source would skew some demographic reporting and make them less insightful.

When TV networks begin selling their inventory, they need to have trust in the way data is projected with stability. Typically network inventory provides an estimated audience that is often above an expected actualized delivery, leaving room for makegoods and not overselling audiences that can't be monetized. The degree of projected and actualized audience needs to be manageable to create stability. If the actualized audience produces a high degree of variance, it will make projection for a monetizable audience much more difficult.

Most linear TV deals are based on legacy demo based CPMs that are derived from unit costs and estimated demo viewership. Shifting to another metric, whether it's the same demo but from a different measurement vendor or moving towards an advanced CRM target audience, requires a CPM value adjustment. The unit cost would remain the same, but the guaranteed audience difference would have a great impact on how the CPMs are calculated. If it is a like-to-like conversion, marketing clients would get the same number of units and pay the same costs which results in a zero-sum scenario.

With each data source addition, there were often changes in the resulting metrics and the same data addition impact varies by vendor. Some vendors applied these data changes and showed minimal differences across the board while some showed large volumetric changes upwards of 20% on average. For users to have reliance on the data, there needs to be some consistency with methodology and resulting output. This produces the much-needed projection reliability on current data as well as future estimates for buying and selling. From a user-perspective, any elements that have an impact on output data needs to be addressed to ensure viability.

3. Moving forward – the scope for collective solutions



Clearly, individual measurement vendors will already be taking steps to address these and other challenges. However, we believe that there are also practical steps that could be taking, collectively and collaboratively, to address these challenges. Here are some recommendations:

Recommendation #1: The Future of Identity: Given concerns about privacy, loss of signal (i.e. cookie depreciation), and laws restricting access to and use of consumer data at scale, the industry may be confronted with a need to migrate to a construct of synthetic IDs (such as, e.g., the WFA/ANA Virtual ID construct, as developed by Google.) Will AI play a role in populating a universe of synthetic IDs with demographic, technographic, and other consumer traits? Given our collective reliance on big data, and the profound importance of identity in making big data useful, perhaps the ad tech/ video advertising industry needs a coordinated effort to explore and define what the concept of Identity looks like in 5 years.

3. Moving forward – the scope for collective solutions

Recommendation #2: Scoring and Validation of Identities: Given the importance of identity to big data-based solutions, and the extent to which the quality of an identity graph can have a profound impact on reported data, the industry might want to consider some systematic assessment of the sources of error in identity resolution and in demographic assignment from identity vendors. While it may be tempting to address identity-driven error by mandating a single industry identity source, in practice (1) many publishers, programmers and brands already maintain their own proprietary identity graphs; (2) eliminating alternatives doesn't magically make the remaining choice right. The goal should be to strive for elimination of error in identity graphs and demographic assignment; not to strive for the elimination of identity providers themselves.

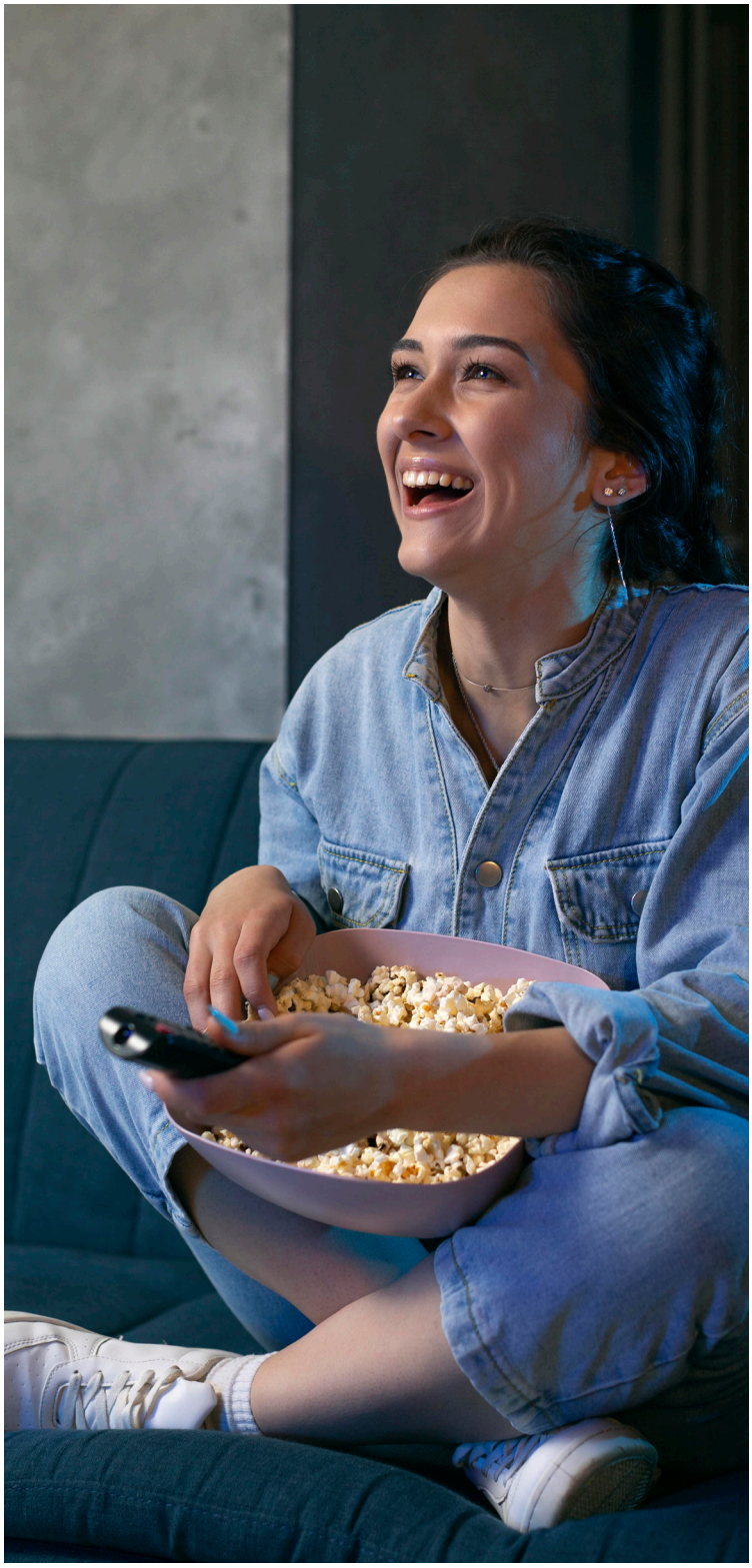
Recommendation #3: Personification research: We know that, in general, algorithmic persons viewing assignment in a big data footprint tends to yield different persons ratings than the legacy people meter panel. The fact of these differences has contributed to the delay in migration to alternative currencies for age/gender (as opposed to advanced target) buys. In addition, differences in persons ratings (and in demographic VPVHs) exist across the big data providers, who use different techniques for personification. Some validation work on the efficacy of different personification techniques (including people meter button pushing) could go a long way toward making buyers and sellers comfortable using alternative currencies for demographic transactions.

Recommendation #4: Centralized metadata: The streaming data asset under development by the JIC, and the various conversations about the creation of an industry panel, demonstrate the efficacy of creating shared industry assets. A single industry-accepted source and taxonomy for metadata about content, ads, and as-run schedules would help to eliminate the variances different metadata can create and would facilitate comparing and combining data across providers. And it is almost certainly more efficient for the programmers to share as-run schedules once, than four or more times.

Recommendation #5: Creation of Codified Standards: Currently the MRC Set-top Box standards are 12 years old; and there are no such standards explicitly for ACR-based Smart TV data. There are no published standards for integrating Smart TV and Set-top Box data. While there are numerous sets of relevant standards and guidelines that pertain to cross-platform currency measurement, there are several important areas without published standards. These include identity; deriving demography in a big data footprint; best practices for using clean rooms; and standards for measuring program audiences across platforms in a hybrid linear/on-demand environment. As content and campaign measurement have bifurcated, the emphasis on standards creation has, logically, been with regard to campaign measurement.

4. Future considerations for users

The six methodological challenges outlined above have implications from a vendor and user perspectives. Each measurement vendor has their own unique approach to the marketplace and not all of them are primarily focused on both content and ad viewership. Some of them have created a full funnel attribution process in which linear tv viewership is only a component in a larger marketing plan. For digital or streaming viewership, there's a general understanding that measurement should be census-based (server logs or pixeled), which leads to more of an alignment on what the actual viewership levels are. For linear TV viewership, the user community is balancing measurement "truthfulness" and adaptability for everyday business needs, including integration into planning, buying, posting, and attribution systems.



The majority of linear TV transactions are still based on a national unit adjacency to content, regardless of distribution model. Unit pricing is assigned to each inventory unit, and an estimated level of viewership

is used to grade for guarantees. Unit costs remain constant regardless of currency; but CPMs based on the guaranteed audience viewership will vary moving forward, as a function of audience estimate source. In other words, a typical 30 second ad unit on a given program will cost X dollars. CPM, or Cost-per-Thousands, will vary based on the different dominator from different audience source. Since this is a very standard rate level calculation, this relationship impact will be the same for both traditional demographic and advanced audiences. Moving to advanced targets could drive even more divergence than traditional age/gender targets, in part because data are weighted to demographic universe estimates, but not to advanced target universe estimates. Proposed and guaranteed viewership metrics must both use the same methodology and data source to ensure predictability, stability, and consistency in a currency transaction.

In 2021, MRC suspended Nielsen's accreditation for national TV which coincided with the emergence of several alternate TV measurement companies like iSpot and VideoAmp. The suspension opened the door for buyers and sellers to reconsider their methods of transaction. Most in the industry felt Nielsen was undercounting audiences and doesn't have the capacity to accurately capture long-tail network audiences and lacks the structure to measure advanced audiences coveted by advertisers. For the past few years, several measurement vendors tried to break Nielsen's stronghold on transactional currency adoption by offering more comprehensive TV measurement using large data footprint and the capacity for advanced audiences that align with digital campaigns.

While the MRC has reaccredited Nielsen (and Comscore later) with national TV accreditation, many buyers and sellers already engaged several alternate currency providers with the focus on targeting advanced audiences at scale. To complicate Nielsen's challenges to re-engage with the industry, Nielsen floated the idea of sunsetting the C3/C7 in favor of Individual Commercial Metrics (ICM) which created uncertainty as regards the future of media transactions and the measurement thereof. While ICM has better metric alignment with digital campaigns, the execution in the current state proved to be challenging and problematic. This metric is particularly problematic for sellers as each ad unit within a single could be priced differently due to different viewership in each pod position and pod break. The 1st position of the pod generally has a significantly higher viewership which could lead to tier pricing for a single program and potentially a stewardship nightmare for all units.

As the industry continues to lean towards advanced audiences for a

better non-demo holistic approach across digital and linear TV, there are differences in how each currency provider onboard and process these coveted targeting segments within their platforms. Their methodology differences and varying resulting metrics could provide a challenge to assess which provider has the best practice approach productizing advanced audiences. With more testing and research, each seller and buyer will need to determine the best transactional metric deemed worthy of the currency label. Based on our interviewees on the user side, most respondents valued consistency to ensure the process from planning to posting has limited and manageable variability.

How publishers and agencies (on behalf of marketers) transact is determined between the two parties. If the ad unit is non-guaranteed, the agreement is simply based on the cost of the program unit. If there's an audience guarantee, then a metric is required to grade the viewership performance. The big difference today is the degree of audience choices available, from demographics to advanced audiences, and from exact unit time to average commercial program metric. Since many digital campaigns can readily target advanced audiences, more marketers with linear TV plans are making the shift to align audiences on both platforms but mindful of the cost implications that could be potentially challenging given most procurement goals are still dependent on traditional demographic C3/C7 metrics.

The MRC's role is not to determine the transactional standards, but rather to accredit measurement vendors meeting various sets of published standards. which typically include requirements for transparency. Recently, MRC accredited both Nielsen and Comscore for their national TV service, but in each case with limited scope. More standards (or a refreshing of existing ones) are needed for things like building an identity graph, assigning demographics based on identity, and viewership source data from OEMs and STBs. While MRC accreditation is a standard for industry users to evaluate and potentially adopt new metrics for transaction, it adds a level of confidence knowing industry watchers are performing due diligence into the audit process and questioning different levels of transparency.

To accelerate the adoption of multiple currencies, a Joint Industry Committee (JIC) was formed in 2023 to certify measurement vendors based on their ability to fulfill the operational requirements of buyers and sellers. The media industry should continue to understand the roles of both MRC and the JIC as they certify and audit, and thus facilitate, currency measurement development. Afterwards, each marketer with their own unique audience goals can determine which measurement path best cover their marketing and measurement needs.

Another real industry consideration is the question of the sustainability of these measurement vendors, both operationally and financially. Since most of them require data access from OEMs and STBs to produce their measurement products, their ongoing access to this data is a driver of stability over time. Losing one or more data sources would almost certainly change the data output. When Wal-Mart purchased Vizio in the first quarter of 2024, the continued availability of Inscape data—a component of most of the currency providers—became far less certain.

Potentially Wal-Mart could combine their emerging retail media platform with Vizio's CTV footprint, but it is unclear that such a data asset would be made available to currency providers. While TV panel management is expensive, big data access is also a very costly expense that is hard to sustain. Even when measurement companies have ample data access, storing all of television viewership over multiple years is a non-trivial data maintenance expense. Most users need at least multiple years on most content viewership to assess future viewership. However, some marquee events like the Olympics, the World Cup, and the US Elections would require more

historical data as benchmarks. In addition, data access for identity matching will also face limitations and restrictions.

Within the past year, Nielsen-owned Gracenote terminated other measurement vendors' access to programming metadata, causing a one-time disruption in services and performance (but also resulting in an increased profile for Gracenote competitors.). Ensuring ongoing data access is a non-trivial challenge, as changes to data inputs inevitably change the outputs. Given TV inventory is purchased in advance with a guaranteed level of viewership, metric stability is vital to the transactional process.

Another practical issue that cannot be ignored in assessing the multi-currency marketplace is the question of how user licenses and subscriptions can generate enough revenues for all the measurement companies. For most media research departments at networks and agencies, the bulk of their data expense commitment is with Nielsen products. Most of these contracts are multi-year commitments linked to investment or inventory volume. If transactions of units and dollars can be made using a different viewership grading

system, potentially this would allow for greater flexibility on how to transact. Moving to advanced audiences is an appealing approach which mirrors some of the digital approaches with high potential and can provide greater cross platform synergy. However, many marketing clients and agencies still have operational agreements that require certain procurement levels to ensure optimal purchasing efficiency is achieved. The longer these agreements are in place, the harder it is to agree to alternative measurements. This impedes the ability of alternate currency providers to scale subscription revenues quickly enough to turn an ongoing profit.

From the user perspective, resources are scarce. Changing metrics for the same linear TV inventory will inevitably lead to a pricing restructuring, which is a switching cost with no distinctive business upside. Mining increased value from a currency measurement change requires an investment in data and people resources that data users are loath to take on without clear near-term benefit. This may be why the new currencies are primarily seeing traction in advanced targets, given that advanced target transactions aren't dependent on multiple years of back data.

Another challenge with alternate currency providers is that they haven't accrued viewership data over an extended historical time frame yet. Several of them provide only two years of viewership data, while the planning-to-posting cycle often extend beyond that. In addition, methodology changes could also yield inconsistent data output for trending purposes. The most problematic area in not having extended viewership data history is planning for large viewership events that happen every two or four years, like the Olympics, the World Cup, and the US election.

The multi-currency discussion is still primarily limited today to linear TV content. The better focus would be on streaming-based metrics, and how each vendor integrates linear and streaming data in a singular offering. However, this also raises the concern of parity in reported data availability across streaming companies. If streamer A integrates their first party data in measurement vendor X but streamer B does not, streamer A would inherently have an advantage in investment selection among agencies using vendor X. Conversely, if streaming data from streamer A is only available in measurement vendor X, measurement vendor Y is now at a significant disadvantage.

While linear TV measurement is inherently subject to the data sets and methodologies each provider deploys, the migration to streaming should introduce less variation across vendors if each have access to the same first party data. During the 2023 NFL football season, there were several streaming-

exclusive games in addition to Amazon's Thursday Night Football. There was also a very highly rated streaming-only NFL playoff game on Peacock. These new distribution channels continue to keep viewers tuned to traditional, live linear TV content but using digitally based distribution. Capturing this viewership at scale requires that vendors have data agreements with these streaming platforms, in order to access their 1st party data. While presumably the results of different vendors using such first party data will converge, a new issue arises from the requirement to use first party data—that not every vendor will have everyone's data. Consistency in streaming data source from various content provider into various TV measurement company will be challenging.

This also suggests that, as more viewing is via streaming (and thus inherently measurable through first party data), the task of the currency provider will shift to be providing combined, deduplicated audiences for reach and frequency. Since this will involve mapping first party data to an identity graph, identity will become an even more important component in currency measurement.

As projections inevitably change if different data sources are added or removed, users will require vendors to facilitate the assessment of these changes well in advance (ideally, more than a year of lead time). As incremental data sources are added to the vendor footprint, marginal impact is reduced. The industry should expect some changes, but stability is essential, and it falls upon the currency provider to minimize the impact of changes to data inputs on reported data.

5. A final word

Despite the ongoing media consumption behavior shift to digital based platforms, the bulk of TV/video transactions, especially for linear TV, are still based on legacy demographic measurement and legacy metrics (C3, C7.). This impedes market adoption of alternative measurement currencies, which have seen far greater traction in the Advanced TV space. This is especially true for smaller TV networks or most advertising agencies where there are limited resources to subscribe as well as commit knowledgeable personnel to monitor, analyze, and find incremental use cases for alternative providers. Both buyers and sellers need to better understand the value proposition of both switching currencies; and, of using multiple currencies.

The migration to streaming should lead to some convergence in viewership or consumption levels. But perhaps more profoundly, the role of the currency provider will change. Programmer or ad server data will be available for an increasing portion of a media buy, minimizing he need for currency vendors to produce viewership metric from scratch. The currency business is migrating to one in

which the role of vendors will be to integrate first party data across sources; validate this data; map the data to an identity spine (at the household, device, and/or person level), and create cross-platform reach and frequency.


cimm
Coalition for Innovative
Media Measurement
ARF