Coalition for Innovative Media Measurement

Television Data Match Rate & Match Bias Study

For People Who Are Exploring Big Data Integrations For Advanced Advertising Solutions

What to expect, what you might see, what you might want to do about it

May, 2023

JANUS STRATEGY & INSIGHTS

pre-meditatedmedia

SEQUENT Partners

INTRODUCTION

This report is the final deliverable for a study commissioned by CIMM in late 2021, designed to explore variations in match rates across a range of identity resolution (IDR) providers - and the reasons for these variations.

The study had a relatively simple objective – investigate the extent to which post-match television viewing profiles differed from the original viewing profile. And determine what caused the differences – was it match rate? Small samples? Skews in the identity graphs that link the data? Other causes?

The study was undertaken in 2021-2022 and involved a comparative analysis of 5 identity resolution (IDR) providers who worked with TV data from an ACR and a STB provider, household/consumer demographic cluster data from Claritas, and behavioral outcomes data (visits to casual and quick serve restaurants) from PlaceIQ (now Precisely). The IDR companies conducted 3 matches - their graph to TV data, TV data to HH demographic cluster data and finally, TV data to behavioral outcomes. The study explored the differences in pre-match and post-match data.

A glossary of terms appears in the appendix of this report.

FINDINGS IN BRIEF

Scale: The number of matched households remaining after the match is generally sufficient for typical advertising use cases like targeting and attribution. This is due to big data and the size of today's advanced advertising television datasets.

Bias: But, it turns out that match rates are indeed important, but they're not the whole story. Equally important are the material biases that are introduced in the matching process – biases that clearly impact television activation and evaluation:

- Compared to the pre-matched television source data, the matching process produces a viewer profile that skews older, higher income, towards heavier TV viewers (particularly in Prime and Sports) and visitors of different restaurant brands.
- These biases are driven inherently by skews in the identity graphs and match processes, which is unsurprising as the identity graphs are not national probability samples – they are just extremely large datasets.

Results: Data matching is a generally successful process, though the performance of individual identity providers varies. Some do quite well. Experience with the various datasets and/or cross-walks established which improves the quality of the match process.

INDUSTRY PRIORITIES

It’s important to understand that data matching is a highly complex process, requiring active investigation, testing and evaluation -- small differences matter. TV datasets from MVPD and ACR sources have their own geographic/demographic skews and the match process may exaggerate differences in the data sources used during the process.

Given this, buyers should take note:

- Nationally projectable databases are required for national TV planning and ROI/ROAS measurement, so weighting post-match data to correct biases is important.
- Post-match data needs to be carefully evaluated, to make sure that it sufficiently approximates the source data.

We also recommend that the industry should develop acceptable processes for weighting and should look to develop standard reports with key match rates, profiles and sample sizes.

CONTENTS

BACKGROUND

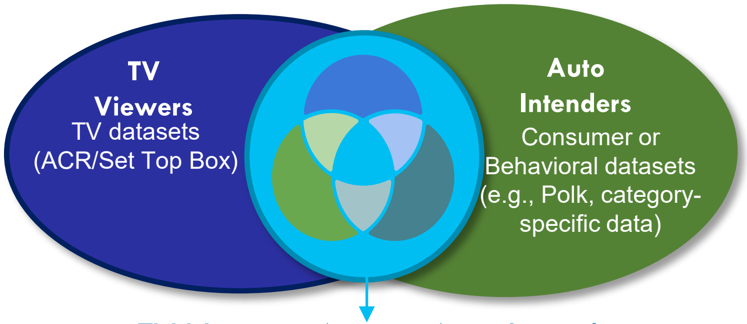
THE REASON WE'RE HERE – TYPICAL USE CASE

Identity resolution* is crucial to advanced television targeting and attribution

- Approaches depend on successfully matching household TV viewing data to target audience characteristics while retaining the scale and integrity of viewing and demographic or behavioral characteristics
- No matches are perfect. We sought an understanding of the viability of match outcomes by IDR providers

Typical Use Case

TV Data Matching for Advanced Audiences



TV Viewing Among Auto Intenders

Matched dataset

*Described In Glossary

BACKGROUND

IDENTITY RESOLUTION AND IDR GRAPHS: THE STARTING POINT IN DATA MATCHING

- IDR providers take different paths to developing IDR graphs*, depending on their foundational data sources
- Some start with the physical household (postal address, Zip11) and connect:
    - HH devices associated with the HH (e.g. CTVs)
    - Persons (or personas) associated with the HH
    - Personal devices associated with the HH (e.g. mobile phones)
- Others start with devices or email addresses associated with persons and build up to households
- Some maintain separate HH and persons graphs and link them
- Most of the providers deal in deterministic data, some probabilistic associations
- This results in a range of proprietary and unique identity graphs – no two are identical
- There is growing recognition that match keys (emails, IP address, street address) may differ in accuracy and quality. However, this is not covered in this study; the data sets were accepted at face value

*Described In Glossary

BACKGROUND

THE CRAFT OF BUILDING AN IDR GRAPH

- Assemble all possible identifiers that can be linked to a HH or individual
  Also called match keys
  HH: name, address, Zip11, device lat/long, IP address, phone number,
  Persons: hashed email, MAID, phone numbers, etc.
- IDR providers consult multiple data sources to check the likelihood of each match key association for quality assurance
- Firewalls, clean-rooms, personification, pseudonymization, anonymization, are routinely used to maintain consumer privacy

Building an IDR graph* and matching TV datasets is not a static undertaking. Graphs are routinely in flux and identity resolution companies are constantly maintaining graph integrity. The process is not simple and projects tend to be custom

*Described In Glossary

BACKGROUND

STUDY OBJECTIVES

- Investigate the extent to which post-match television viewing profiles differ from the original viewing profile
- Reveal how TV data match results vary across identity resolution providers and what that means for media decisions, decomposing three critical stages:

1. TV Data To The Identity Resolution Providers' IDR Graphs

Is there sufficient TV sample size for target audience reporting after the match?

Are matched TV viewing estimates adequate for planning reach of a TV campaign?

Common match – used in in media planning and evaluation

2. Appending HH Demographics To Viewing Data

Is there sufficient TV sample size for target audience reporting after the match?

Are there demographic shifts in viewing data after the match?

Are there skews in demographic representativeness vs. total US or regional geography?

Common match – used in in data enrichment

3. Appending Consumer Behavior Outcomes To The Viewing Data

Is there sufficient TV sample size for target audience reporting after the match?

Are there shifts in share of restaurant visits after the match?

Are there shifts in regional skews?

Common match – used in outcome measurement

BACKGROUND

STUDY PARTICIPANTS* - THANK YOU!

CADENT

EPSILON

experian

neustar
A TransUnion® Company

THROTLE

InfoSum, Transunion and LiveRamp also contributed, but were not part of the match study

Viewing	TiVo SAMBA TV
Behavior	PlaceIQ
HH Demos	claritas
Data processors	ARF

***NOTE: TV data sources and IDR providers are masked throughout this report since the study was designed to elicit generalizable learning**

BACKGROUND

DATA SETS IN THE MATCH STUDY

NATIONAL TV DATA

About 2 M Households from two sources (TV1 – TV2)

Match keys included IP addresses, hashed emails and mobile IDs (IDR companies used any or all identifiers to align TV data to households)

CLARITAS

127M Households

68 cluster segmentation and 11 lifestyle aggregate segments

National

PLACE IQ

- **22.7M Households that Visited Casual, Fast-Casual and Quick Serve Restaurants**
- Provided mobile ad IDs of people who visited these restaurants in the test period

Study Time Period
10/17/21 – 10/29/21

**NOTE: TV data sources are masked throughout this report since the study was designed to elicit generalizable learning. Claritas and Place IQ are identified because they are the sole data source of their type in this study.*

A COUPLE OF IMPORTANT CAVEATS

This study was lengthy and exceedingly complex:

- Securing the datasets and cooperation of IDR providers took a long time and there were legal entanglements , which is fairly typical
- Providers often undertake only one match, e.g., TV data to their IDR graph

There were no constraints put on IDR providers relative to how they should match the datasets, which households to include, which identifiers to use. These decisions were left to their discretion and reflected the real-world

Some identity resolution providers utilized pre-existing cross-walks, or data exchange standards, across different datasets … but others didn’t

- Since matches tend to be custom and a mix of art and science, the actual match rates reported in this study may differ from what any individual matching exercise might yield

CAVEAT
MATCH RATE VERSUS MATCH ACCURACY

Match rates are an important dimension in data matching because they show how many records the two datasets share in common

- For targeting, size/scale of the matched dataset is a key consideration
- For audience currency, accuracy is more important

However, regardless of match rate, not all matches are correct

- Matching data with match keys that have expired will result in an incorrect match, e.g., IP addresses that has been reassigned to another HH or old email addresses no longer in use

This study assessed the nature of the bias that resulted from data matches, not specifically match accuracy

- Match accuracy* requires an independent validation data set which was not available

*Described In Glossary

Match Rate Results

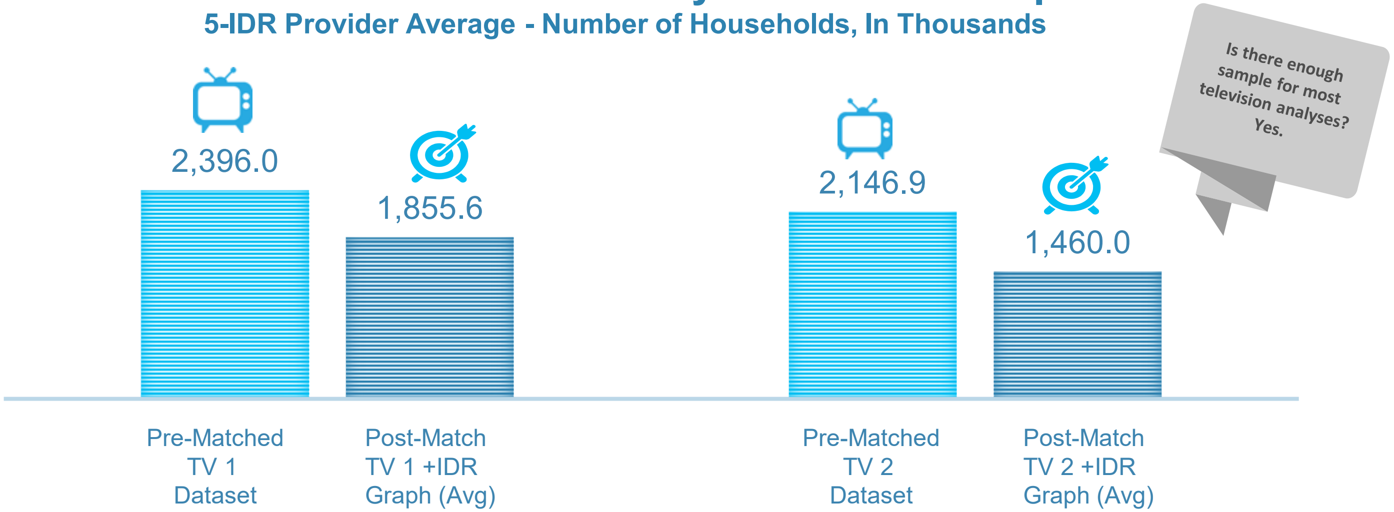
- Linking TV Data To The Identity Resolution Providers' IDR Graphs

TELEVISION DATA MATCHING QUESTIONS

- Is there enough scale in the post-match sample for standard targeting and attribution use cases?
- Are match rates robust?
- Do match rates vary by identity resolution provider?

Household Sample Sizes Remain Robust After Match of TV Data To Identity Resolution Graphs

5-IDR Provider Average - Number of Households, In Thousands



Generally, post-match household counts are robust for advanced television analyses. They are lower than the initial TV datasets, which may impact some analyses of viewing, but they are generally strong.

OVERALL MATCH RATES: TELEVISION DATA TO IDENTITY RESOLUTION PROVIDER GRAPHS

TV Provider 1

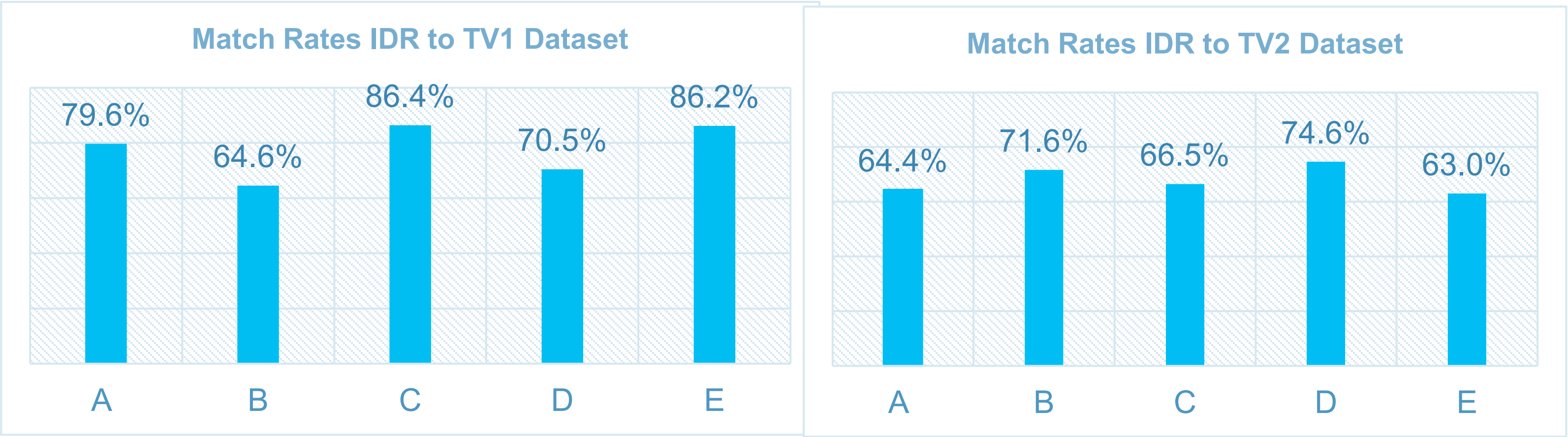
Match between IDR graphs and TV 1 dataset was 77.5%, on average

72.8% Average Match of TV data to IDR Provider Graphs

TV Provider 2

Match between IDR graphs and TV dataset was 68.0%, on average

Match Rates Between IDR Graphs and TV Data Varied from 64% to 86%



IDR Provider Match Rates

IDR – TV Data Match Rate Observations

- Sample sizes remain robust for advanced television analyses after matching TV data to IDR provider graphs
- Provider match performance varies: matches depend on the prevalence of the same identifiers existing in both IDR graphs and TV datasets -- IP address, hashed emails or mobile Ad IDs
- Higher match rates may be a result of proficiency with specific identifiers and pre-existing cross-walks which were used, but not mandated, between the TV data firms and IDR providers
- Matching is a complex operation and experience matters

Match Rate Results

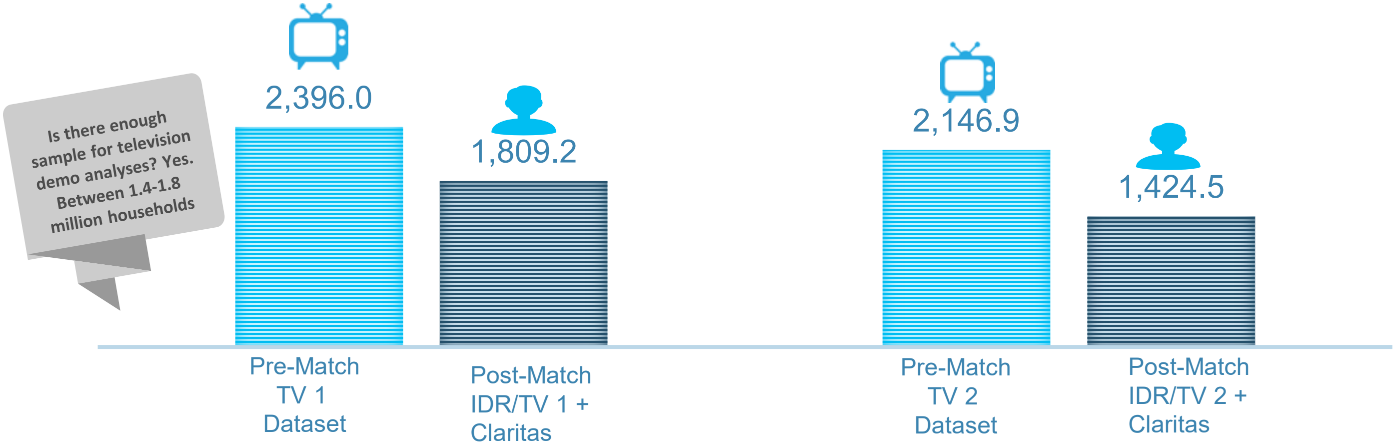
Appending HH Demographics To The Viewing Data

MATCHING TV DATA TO DEMOGRAPHIC DATA - THOUGHTS

- Claritas data is comprehensive; near census-level
- By matching IDR graphs to Claritas, it is possible to see the upscale, traditional skews of the IDR graphs, even before TV data is matched
  - a function of the subscriber, user data, loyalty cards, public records, and other information sources that fuel the graphs – incidence of younger, lower income people tends to be lower in these sources
- Therefore, matching TV datasets with the IDR graphs skews more upscale since Smart TV/STB households are also slightly higher income

Household Sample Sizes Remain Robust After Match of TV Data To Identity Resolution Graphs and Claritas Demographic Data

5-IDR Provider Average - Number of Households, In Thousands



MATCH RATE: APPENDING DEMOGRAPHICS TO VIEWING DATA

TV Provider 1

Match between TV 1 dataset, IDR graphs and Claritas demo data was **75.5%**, on average

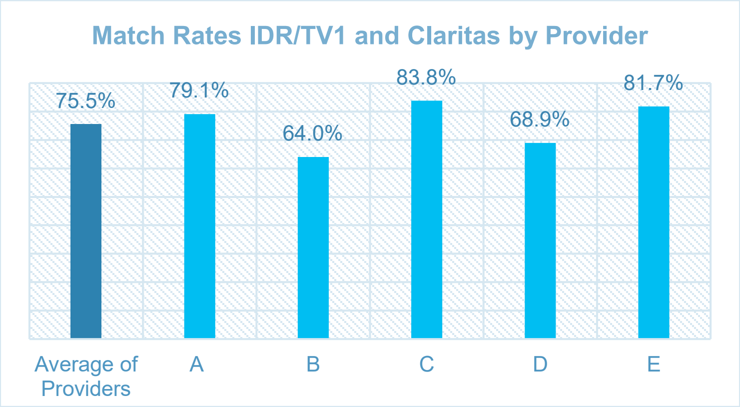
71.3%
Average Match of TV data , IDR Provider Graphs and Claritas Demo data

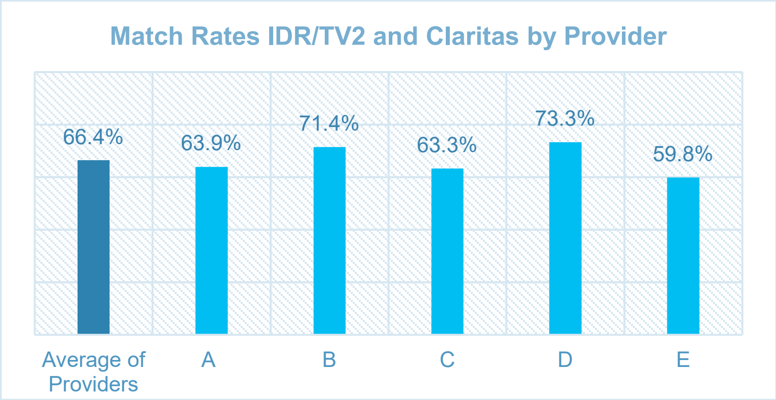
TV Provider 2

Match between TV 2 dataset, IDR graphs and Claritas demo data and was **66.4%**, on average

Strong match rates reflect the census-like nature of Claritas' demographic profiling data

Match Rates Between IDR Provider Graphs, TV Data and Demographic Data Varied from 63% to 84%





TV data sources to Demographic matches are generally strong, but there is little consistency between the two TV datasets by IDR provider. This suggests that familiarity with match keys/identifiers used to link television data as well as the presence of an established crosswalk can make a difference and increase match rate

CLARITAS DEMO DATA REVEALS INHERENT SKEW IN IDENTITY GRAPHS

Claritas is a near-census dataset and provided a clean look at IDR match profiles. Almost all IDR providers skew to more upscale, more established households and under-represent less established, lower income households. Lower income, younger households are more transient and less settled, therefore less likely to be included in datasets that fuel IDR graphs. Most extreme in Providers C and D.

IDR Companies Match to Claritas Demographic Data

				Post-Match Index to Claritas Lifestage Group % US Households					
% of Total	Claritas Lifestage Group	Median Age	Median HHI	Average	IDR A	IDR B	IDR C	IDR D	IDR E
8.95%	Affluent Empty Nests	47.1	$137,551	116	110	107	124	124	95
11.39%	Accumulated Wealth	38.8	$110,950	117	107	109	126	124	96
10.80%	Conservative Classics	42.4	$87,167	105	102	101	110	106	97
7.89%	Young Accumulators	39.3	$78,927	114	106	108	119	123	99
8.87%	Midlife Success	39.2	$73,911	83	86	91	89	64	95
8.67%	Mainstream Families	35.5	$64,991	110	104	105	108	123	103
6.93%	Sustaining Families	31.6	$40,730	95	102	96	86	97	99
12.97%	Cautious Couples	44.7	$38,795	103	104	102	96	111	101
8.9%	Young Achievers	35.5	$24,813	82	88	91	79	69	101
7.03%	Sustaining Seniors	45.2	$22,835	89	96	96	74	92	109
10.0%	Striving Singles	33.4	$21,870	75	89	88	72	53	108

Indices 120 and above or 80 and below are in light blue

CLARITAS DEMO DATA REVEALS NHERENT SKEW IN IDENTITY GRAPHS

These skews become even more apparent when groups are sorted by income and grouped to compare the highest income group to the lowest income group

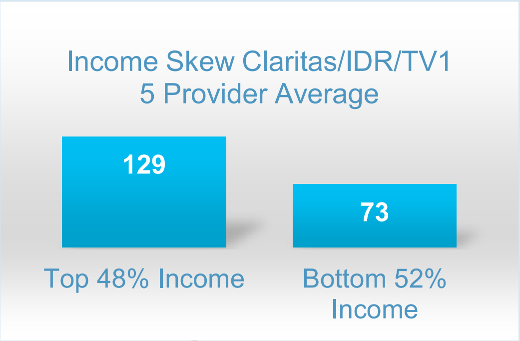
- The top 20%, with median income of $122,655, is overstated by 12%; the bottom 14%, with median income of $22,251, is understated by 14%
- But for two of the IDR providers, the higher income group overstatement is +24% to +25% and the lower income group understatement is -27% to -31%.
- Its clear that the data matching process can bias a data set's demographic profile, sometimes a little, sometimes a lot, and results vary by IDR provider

IDR Companies ID Graph Match to Claritas Demographic Data

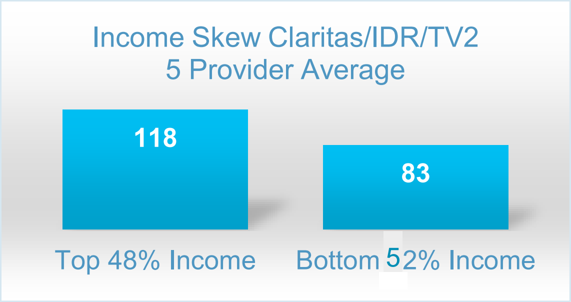
			Index of IDR Providers' Post-Match % US Households Indexed to Claritas					
% of Total	Claritas Lifestage Group	Median HHI	Average	IDR A	IDR B	IDR C	IDR D	IDR E
20%	Highest Income Fifth	$122,655	112	108	108	125	124	95
17%	Lowest Income Fifth	$22,251	86	92	91	73	69	108

Indices 120 and above or 80 and below are in light blue

MATCHING TV DATA TO DEMOGRAPHIC DATA REVEALS UPSCALE BIAS



Increases the proportion of upscale HH in the matched sample



» Not surprisingly, since IDR graphs skew upscale, the match of TV data and demo data results in a higher concentration of upscale households (20-30% more, on average) than in the original IDR/TV data match

UPSCALE BIAS SEEN IN ALL IDR PROVIDER MATCHES

TV Data 1

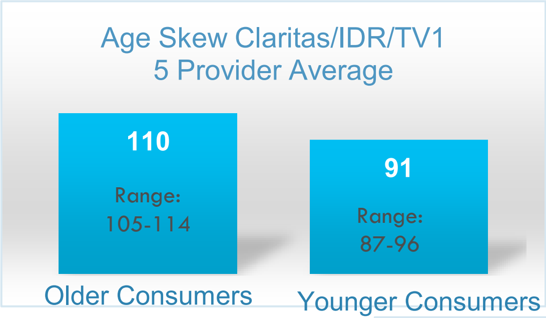
	Index to Total US Lifestage Household Population				
	IDR A	IDR B	IDR C	IDR D	IDR E
Top 48% Income	127	134	131	125	127
Bottom 52% Income	75	69	71	77	75

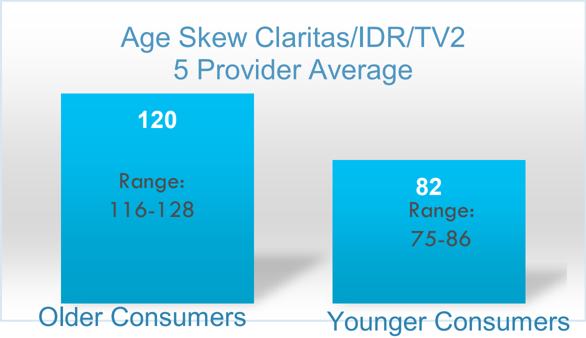
TV Data 2

	Index to Total US Lifestage Household Population				
	IDR A	IDR B	IDR C	IDR D	IDR E
Top 48% Income	116	117	125	115	118
Bottom 52% Income	85	84	77	86	84

- Extremely similar results - all five IDR provider TV data matches produced a more upscale audience than expected based on the Claritas data
- Inclusion of less established and lower income homes is a challenge for all IDR providers
  - The imbalance should be addressed with weighting, which will correct the under-and over-representation of particular groups in the matched dataset

MATCHING TV DATA TO DEMOGRAPHIC DATA PRODUCES OLDER SKEW





- TV databases matched to Claritas data also contain more older consumers than younger consumers
  - All providers report similar results, though there is a bit of a range
- This is a function of younger adults tending not to participate in common identity resolution data-gathering systems, like car loans or home mortgages, at a high rate
- This imbalance should also be addressed with weighting

Older – Top 47% of adults by age, Younger- Bottom 53% of adults by age. Range represents the highest and lowest individual provider results, indexed to original Claritas data

IDR/TV/Claritas Match Rate Observations

- Sample sizes remain robust for advanced television analyses after matching TV data to IDR provider graphs and Claritas demographic data
- Resulting matched samples skew more upscale and older than original television datasets, which can impact television viewing profiles and impression counts against viewers by income and age
- All providers report similar overstatements, though there is a range of results
- These skews should be addressed through weighting.

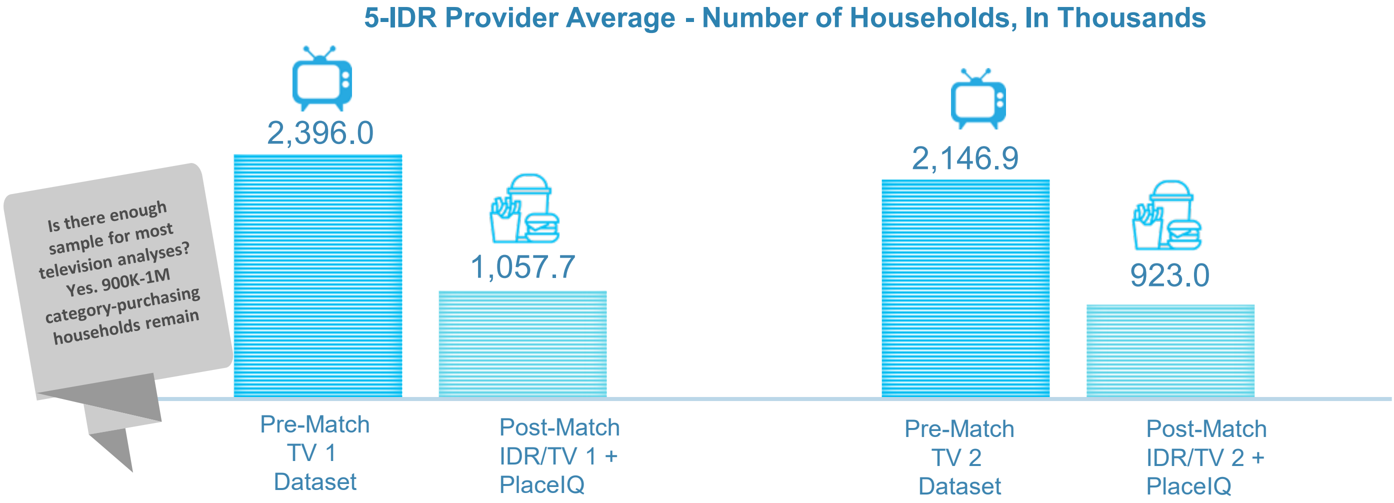
Match Rate Results

Appending Consumer Behavior Outcomes
To The Viewing Data
(Restaurant Visits)

BEHAVIORAL OUTCOME DATA - RESTAURANT

- Location-generated restaurant visitation data provides a good look at the impact of integrating behavioral data into television data
  - These matches are based exclusively on Mobile Ad IDs, not IP addresses or hashed email identifiers as in the other datasets
- Importantly, only a subset of the population visited Casual, Fast-Casual or QSR restaurants in the test period, so sample and match rates are smaller than the TV or demographic data matches

Household Sample Sizes Remain Robust After Match of TV Data To Identity Resolution Graphs And Restaurant Visit Data



Sample sizes are lower than the initial TV dataset samples because only about half the population visited casual or fast food restaurants during the test period. This may materially impact analyses of viewing and behavioral attributes.

Are there enough for Advanced TV, analyses, though? Yes, household counts are robust.

MATCH RATES: APPENDING CONSUMER BEHAVIOR OUTCOMES TO VIEWING DATA

TV Provider 1

Match between IDR graphs, TV data, and PlaceIQ restaurant visit data was **44.1%**, on average

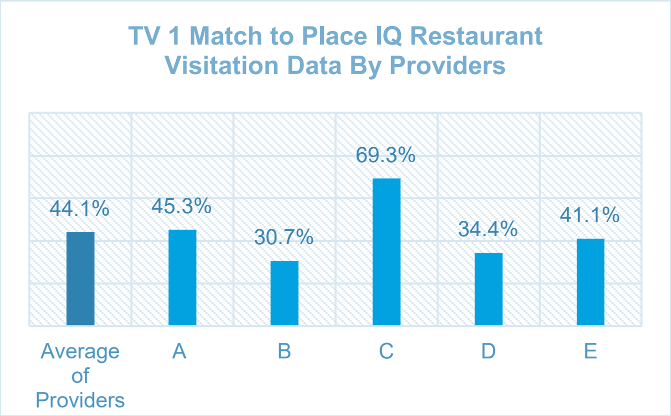
43.6%
Average Match of TV data to IDR Provider Graphs and PlaceIQ Restaurant Visit data

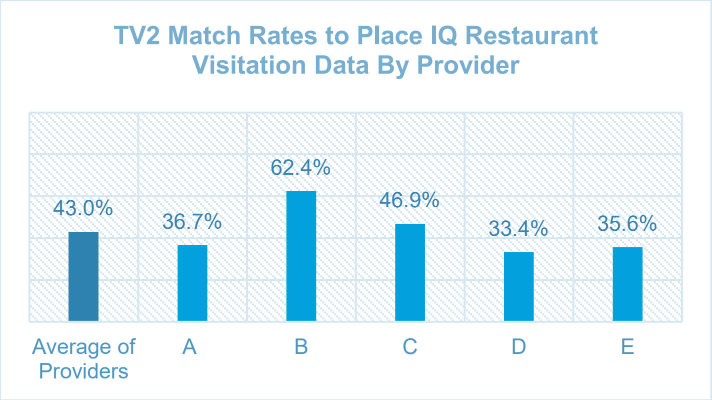
TV Provider 2

Match between IDR graph, TV data, and PlaceIQ restaurant visit data was **43.0%** on average

PlaceIQ file only reflects the mobile ad IDs associated with Casual, Fast Casual and Fast Food Visits between 10/17/21 – 10/29/21

Match Rates Between IDR Provider Graphs, TV Data and Behavioral Outcome Data Varied Significantly from 30% to 69%





Little consistency by IDR provider across the two TV datasets possibly due to variability of mobile ad IDs in the graphs or the presence of established crosswalks for some providers

IMPACT OF THE MATCH PROCESS ON CHAIN-SPECIFIC VISITATION DATA

Since match rates between PlaceIQ and the TV data varied considerable among IDR providers, it is not surprising that there is significant variability in the post-matched data – visits to the Casual, Fast Casual and Fast Food categories as well as to individual brands, by identity resolution provider.

In this section of the study, pre-matched PlaceIQ restaurant visit data is compared to post-match IDR provider graphs and then to post-match TV data …

- To see if the matches correctly reflect the share of visits to that type of restaurant or individual chain
  - For example, did Quick Serve Restaurants maintain the largest share of visits post-match?
  - Did McDonald's receive the largest share of visits in the post-matched data as they did in the pre-matched data?
  - How did the smaller chains fare in the matching process – was their share of visits over- or under-reported?

Skews in the post-matched TV/outcome data could definitely impact a restaurant's advertising performance and ROI

WIDE VARIABILITY SEEN IN RESTAURANT CHAIN VISITS BY IDR PROVIDER

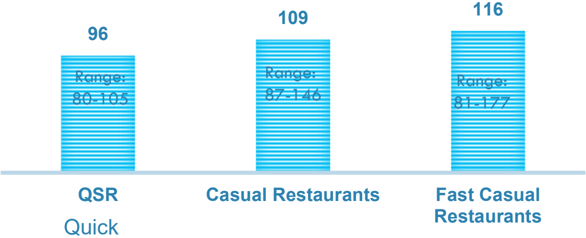
		IDR/Place IQ Match - Before TV Data Brought In				
		IDR A Indexed to Source	IDR B Indexed to Source	IDR C Indexed to Source	IDR D Indexed to Source	IDR E Indexed to Source
Category Visits	QSR Total	81	100	99	100	106
	Casual Restaurants Total	147	102	105	103	87
	Fast Casual Restaurants Total	171	98	100	95	76
Brand Visits	McDonald's	82	100	101	100	116
	Wendy's	121	101	104	101	87
	Burger King	115	102	105	101	89
	Sonic Drive In	99	99	100	106	101
	Pizza Hut	110	101	108	100	87
	Panera Bread	122	101	105	99	81
	Chipotle Mexican Grill	126	95	101	91	76
	Domino's Pizza	103	99	2	97	89
	Panda Express	133	97	101	92	69
	Applebee's	117	103	112	103	83
	Waffle House	123	99	100	106	80
	Papa John's Pizza	105	101	111	94	89
	Little Caesar's Pizza	131	100	106	95	70
	In-N-Out Burger	125	96	95	79	65
	Steak-N-Shake	141	101	104	103	66
	Moe's Southwest Grill	148	100	102	105	63
	Sweetgreen	70	78	114	59	103
	Bonefish Grill	100	103	125	96	94
	California Pizza Kitchen	139	97	103	85	58

Indices above 110 or below 90 in light blue

Consistent overstatement of visits to many chains by Provider A, understatement by Provider E

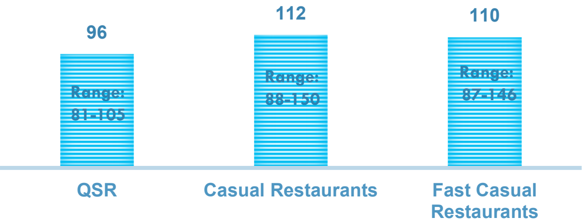
Matching TV Data to Behavioral Outcomes – Slightly More Visits To Casual and Fast Casual Restaurants

IDR/TV1/PlaceIQ 5-Provider Average



Quick
Service
Restaurants

IDR/TV2/PlaceIQ 5-Provider Average



Range represents the highest and lowest provider results, indexed to pre-match Place IQ data

» Matching TV data with PlaceIQ increased the share of visits to casual and fast casual restaurants, compared to the unmatched PlaceIQ visit data
  - Quick Serve Restaurants were relatively flat, though there is a significant range by provider

» These skews are likely to impact planning decisions and ROI/ROAS measurement

» Should be addressed with weighting

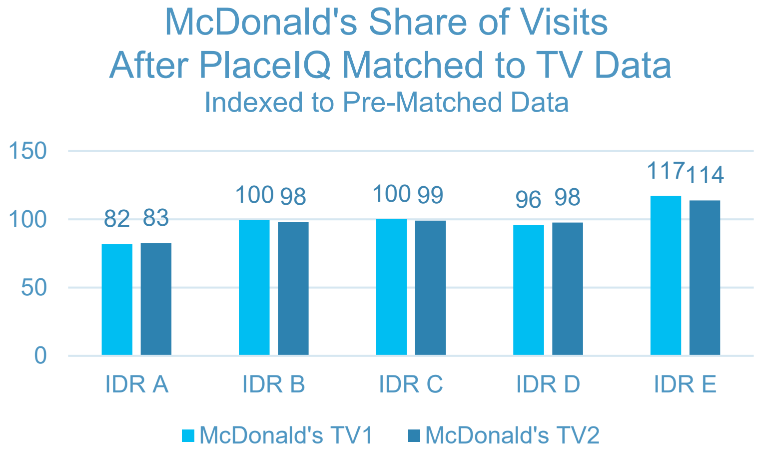
IMPACT OF MATCH ON INDIVIDUAL BRAND OUTCOMES

VISITS TO MCDONALD'S IMPACTED DIFFERENTIALLY BY PROVIDER

POST MATCH SHARE OF RESTAURANT VISITS, PLACE IQ AND TV DATA

McDonald's share of all restaurant visits between 10/17/21 and 10/29/21 was 47%

McDonald's:
Largest restaurant chain and geographically dispersed in population centers



In this match, share of visits to McDonald's is consistent in post-matched data– biggest differences is between the providers

IMPACT OF MATCH ON INDIVIDUAL BRAND OUTCOMES

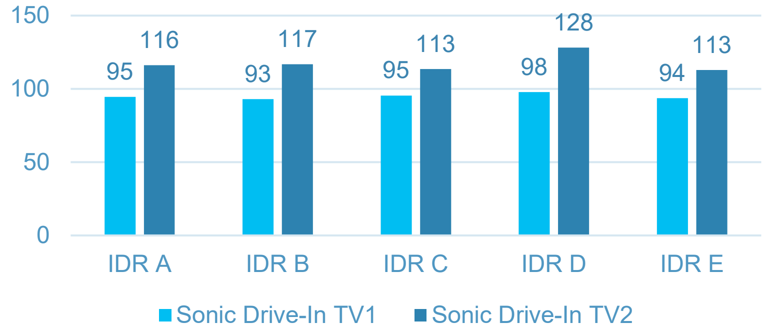
VISITS TO SONIC DRIVE-IN SIGNIFICANTLY IMPACTED BY THE TV DATA MATCH PROCESS

POST MATCH SHARE OF RESTAURANT VISITS, PLACE IQ AND TV DATA

Sonic's pre-match share of all restaurant visits between 10/17/21 and 10/29/21 was 9%

Sonic
Relatively large but
Regionally
concentrated

Sonic Post-Match Share of Visits
PlaceIQ Data Matched to TV Data



In this match, share of visits to Sonic is quite different when location data is match to the TV datasets – smaller differences seen across the providers; all show more visits to Sonic in the TV2 matches

IDR/TV/PlaceIQ Match Rate Observations

- Sample sizes, generally, remain robust for advanced television applications after matching TV data to IDR graphs and PlaceIQ category data
- Matching TV and visitation data results in more Casual/Fast Casual visits versus the original data - all providers report similar overstatements, though there is a range of results
- Since none of these datasets are nationally representative, differences in match rates appear when brands have a distinct regional footprint
- These geographic skews should be addressed through weighting

Match

Impact Results

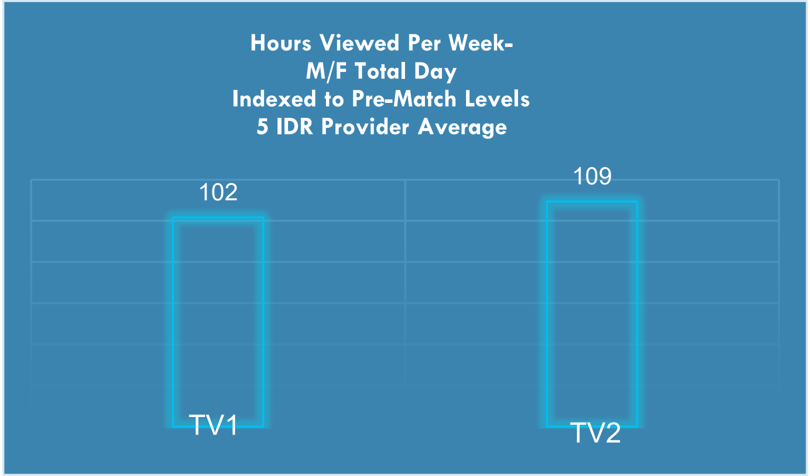
Effects of Matching on TV HH Viewing Profiles, Hours of Viewing and Daypart/Genre Skews

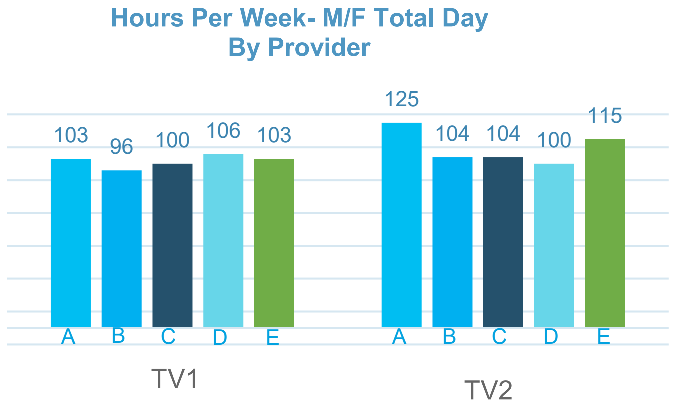
IMPACT OF MATCH PROCESS ON HOUSEHOLD TELEVISION VIEWING PROFILES

The study was designed to reveal the extent to which post-match television data viewer profiles differed from the original viewer profile, e.g.,

- How does the skew towards older, upper income consumers in the IDR graphs impact television viewing levels? Are the post-matched households heavier viewers in terms of hours of viewing? Are some dayparts watched more than others? Does the distribution of Heavy/Medium and Light viewing households remain the same post-match?
- How much variability is there by IDR provider in these viewing hours?

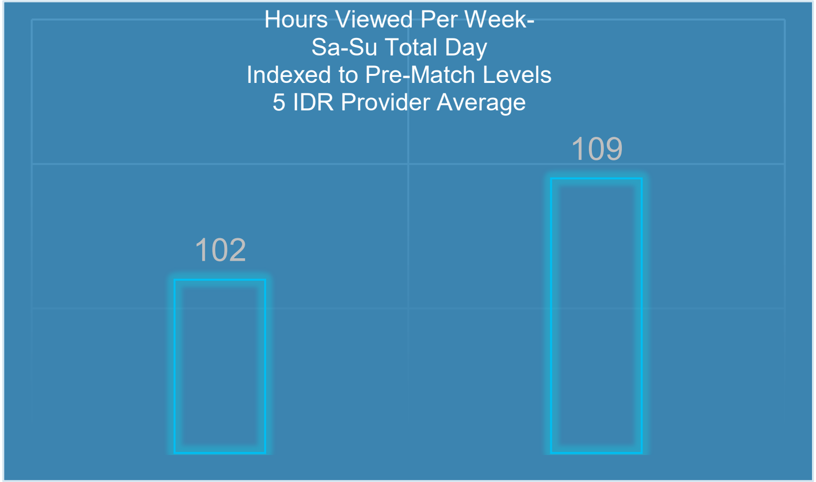
THE IDR/TV DATA MATCH SKEWS TOWARDS HOUSEHOLDS THAT WATCH MORE WEEKDAY TELEVISION COMPARED TO PRE-MATCHED TV DATA

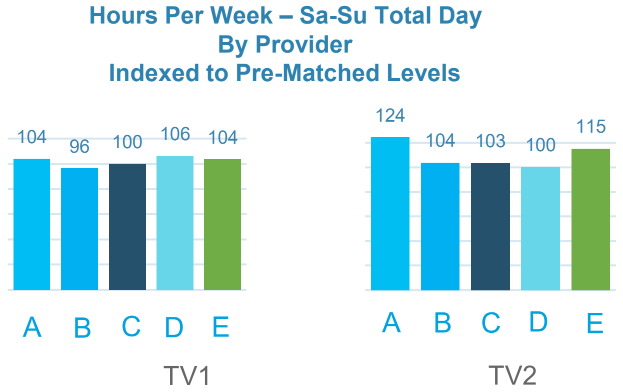




Matching TV data with the IDR graphs created a dataset with increased time spent viewing, particularly for TV2, where reported viewing increased an average of 9% across IDR providers; increase in reported viewership for TV2 dataset was driven mostly by two providers .

SIMILARLY, IDR/TV MATCHES SKEW TOWARDS HOUSEHOLDS THAT WATCH MORE WEEKEND TV COMPARED TO PRE-MATCHED TV DATA





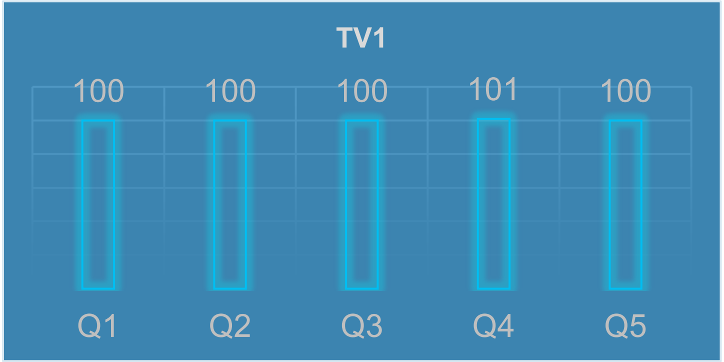
Matching TV data with the IDR graphs boosted weekend time spent viewing, particularly for TV2, where reported viewing increased an average of 9% across IDR providers; increase in reported viewership for TV2 dataset was driven mostly by two providers

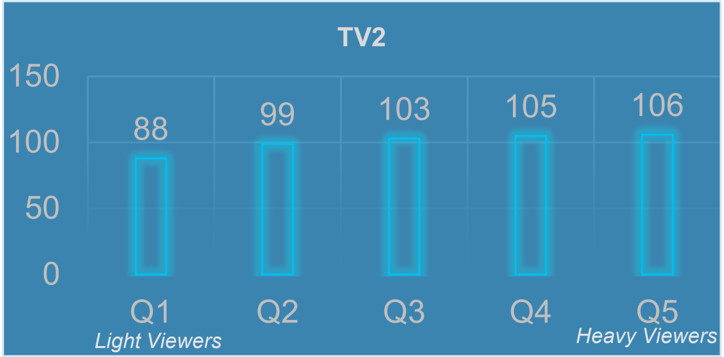
IDR/TV DATA MATCHES SKEWED BY FEWER LIGHT VIEWING HHs (QUINTILE 1) AND MORE HEAVY VIEWING HHs (QUINTILE 5) IN ONE OF THE TV DATASETS, COMPARED TO PRE-MATCHED TV DATA

Impact of IDR/TV Data Matches On HH Viewing Levels

Match Rate by Quintile*, Hours/Week

5 Provider Average Indexed To Overall Match Rate





*Described In Glossary

ACROSS THE BOARD, IDR PROVIDERS' MATCHES TO TV DATASET 2 CONTAIN FEWER LIGHT VIEWING HH AND MORE HEAVIER VIEWING HH, COMPARED TO PRE-MATCHED TV DATA

Impact of IDR/TV Data Matches On HH Viewing Levels

Match Rate by Quintile By Provider

Hours/Week Indexed to Overall Match Rate

TV1

	IDR A	IDR B	IDR C	IDR D	IDR E
Quintile 1	99	97	100	99	101
Quintile 2	100	99	100	99	101
Quintile 3	101	102	100	99	100
Quintile 4	101	103	100	99	100
Quintile 5	99	99	100	103	98

Light

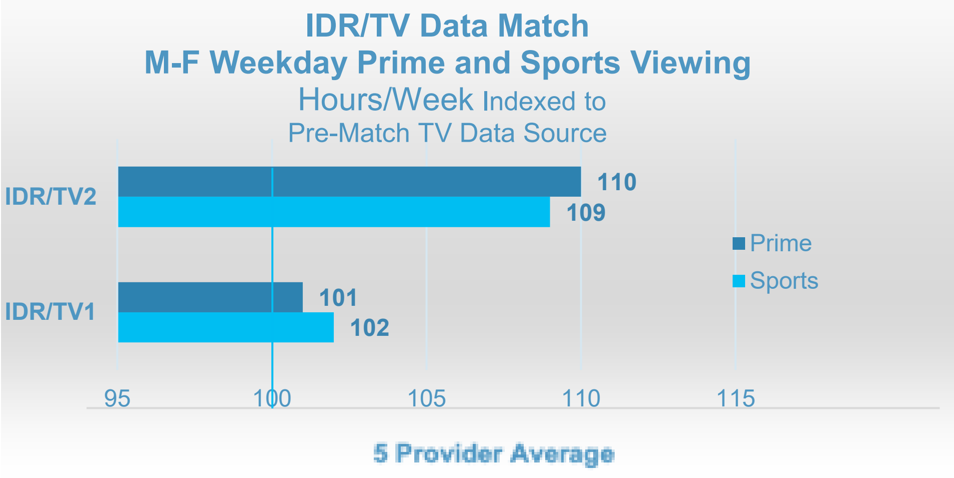
Heavy

TV2

	IDR A	IDR B	IDR C	IDR D	IDR E
Quintile 1	87	87	88	88	87
Quintile 2	99	99	99	99	99
Quintile 3	103	103	103	102	102
Quintile 4	105	105	104	105	105
Quintile 5	106	106	106	106	106

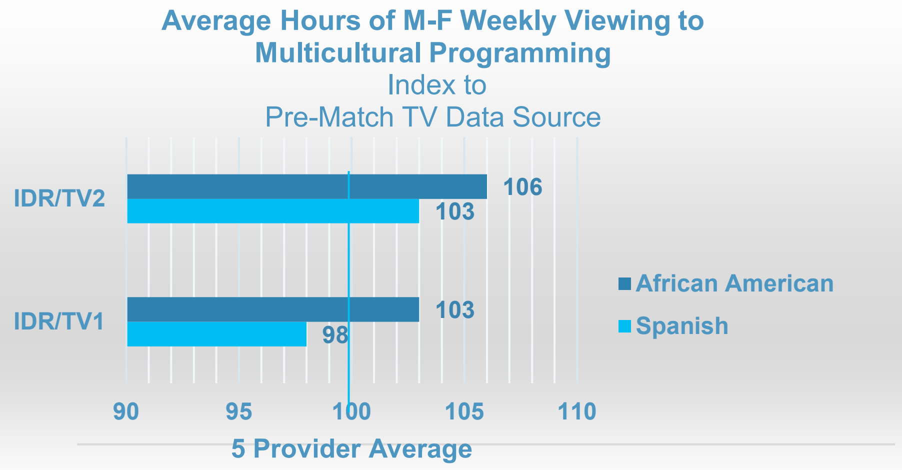
Difficult to know why, exactly, but it's possible geographic skews or match key differences in TV Dataset 2 make it more difficult to match light viewing households. However, inclusion of more heavy viewing households than in the pre-matched data will increase impressions and potentially alter planning decisions

IDR/TV DATASET MATCHES RESULT IN INCREASED HH VIEWING TO PRIMETIME AND SPORTS, RELATIVE TO PRE-MATCHED LEVELS



Inclusion of fewer light viewing HH and more heavier viewing households in the IDR/TV2 match produces more viewing hours in Primetime and to a lesser extent, Sports

IDR/TV DATASET MATCHES RESULT IN VERY SLIGHT INCREASES IN VIEWING TO AFRICAN-AMERICAN AND HISPANIC PROGRAMMING RELATIVE TO PRE-MATCHED LEVELS



Compared to pre-matched television data, matches produce higher viewing hours to African-American and Hispanic programming; especially with TV 2 dataset

WHY BIASES IN IDR/TV DATA MATCHES MATTER

- The differences in matched geographic or demographic viewer profiles produce overstated viewing levels that can lead to buying fewer TV campaign Impressions and would likely dampen the reported ROI performance
- May also impact planning decisions and allocation - TV costs could be 10-20% higher for audiences that aren't there and reported ROI/ROAS could be 20% lower, depending on the IDR provider
- End users of the matched data should be aware of these overstatements and adjust the viewing levels to be closer to what is found in the pre-matched TV viewing data sets

CLOSING THOUGHTS

- The data matching process is highly complex process, requiring active investigation, testing and evaluation
- The media world has focused on match rate, which is important. But match biases will exist and post-match results should be evaluated on both standard and a case-specific bases
- Weighting should be applied, as needed, to pull data biases into alignment
- The desire for scale (reach) in advanced television applications must be tempered by the realization that, at some point, accuracy is compromised. Lack of a validation/truth set prevented an analysis of match accuracy, but there were indications that the identity resolution process faces an accuracy versus coverage trade-off like we see related arenas such as advanced audience segments
- This study compared ad hoc matches among providers. However, every provider assured us that investing in cross-walks between datasets would reduce variability and yield better results

Recommendations

RECOMMENDATIONS
FOR USERS OF MATCHED IDENTITY AND TV DATA

1. It's important to understand that data matching is a highly complex process, requiring active investigation, testing and evaluation; small differences matter

2. TV data has its own skew and the match process exaggerates differences in data sources

3. It's also important to understand that each IDR provider's matched output will differ based on how datasets are collected and used in the household graph

    a) Different identity resolution providers have strengths with different matching variables (i.e. postal address or hashed email address). The IDR provider selection should include finding the identity resolution provider with strengths in the matching variable attached to the databases being integrated

4. Some of the key statistics, like the representativeness of the IDR provider's device graph or results from other TV database appends, could be provided ahead of initiating a project

5. Actively participate with IDR provider during the process of database integration, ensure that they are taking the proper adjustment steps, e.g. weighting

6. Develop QA benchmarks for checking incoming data, e.g. broad daypart viewing, quintiles, internal historical match rates, etc.

RECOMMENDATIONS
FOR PROVIDERS OF MATCHED IDENTITY AND TV DATA

1. Adopt standard report with key match rates and post-match sample sizes to inform buyers of resulting TV data skews and biases

2. Nationally projectable databases are required for national TV planning and ROI/ROAS measurement
   a) Need to consider building processes for weighting to adjust for demographic skews in the IDR provider's device graph

3. After applying weights, IDR provider should evaluate the extent to which this brings TV viewing levels more in-line with pre-matched levels, and make further adjustments if necessary

4. Match rates, sample sizes, skews and weighting must be disclosed in a methodology report

5. Allocate resources for professional development of data science staff for deeper understanding of TV and consumer data

STANDARDIZED REPORT
POST-MATCH SAMPLE SIZES AND MATCH RATE FOR IDR/TV DATA MATCHES

CIMM recommends the industry adopt a transparent, standardized reporting system that provides key facts about the quality and composition of matched data used in audience targeting and measurement.

The intent is to inform users about the relative skews of the post-matched data sources on standard television metrics – as well as their applicability for media strategy and investment decisions – and to address issues of transparency.

An example appears on the next page.

STANDARDIZED REPORT FOR TV DATA MATCHES

		TV Data Source	ID Graph/TV Data Source	Consumer Data & ID Graph TV Data Source	Adjustments
	TV DataSet HH Match %				
	HH Match Count (000)				
		Demographic Composition %			Recommended Weighting %
Income	Top 20%				
	20%				
	Mid 20%				
	20%				
	Low 25%				
		100%	100%	100%	100%
Age: Household Head	<35				
	35-64				
	65+				
		100%	100%	100%	100%
Ethnicity/Race	Asian				
	Hispanic				
	African American				
		100%	100%	100%	100%
Presence of Children	<Age 18				
	<Age 6				
		100%	100%	100%	100%
Region	Northeast				
	East North Central				
	West North Central				
	Southeast				
	Southwest				
	Mountain				
	Pacific				
		100%	100%	100%	100%
TV Viewing	Average HH Wkly Viewing Hours (000)				

Television Data Matches – Audience Skew & Adjustments Report

A standard template for comparing pre-match and post-match TV data sample compositions as well as the weighting that must be done to bring the sample back into balance and reflect US population

Identity providers should report the TV data match rate (%), the resulting match count (000) as well as the household demographic composition percentages and the number of average household weekly viewing hours (000) for full transparency

APPENDIX: Identity Resolution/Matching Glossary

Crosswalk: An integration between disparate datasets, connecting the dots between datasets at scale. Takes advantage of multiple individual and household level match keys present in an identity graph and maps the equivalent fields across various datasets.

Identity Resolution: The process of identifying, matching and merging records that correspond to the same entities - households, or devices - from several databases using personally identifiable information like street address, device IP addresses, etc.

ID graph: The backbone of data matching - a database that stores identifiers that correlate with individual consumers. These identifiers range from usernames to email, device IDs, phone, cookies and offline identifiers like demographics or loyalty card numbers, etc.

Match Rate: The percentage of the overlap portion between two datasets that can be found with a common identifier such as hashed email, device address, mobile Ad IDs or cookies

Match Accuracy: The percent of matches that can be validated against a truth set; the match keys are associated with the same HH, device or person, in the truth set

Viewing Quintiles: Total viewing households sorted by viewing hours, then divided into five equal parts to highlight differences in hours viewed

Thank you!

Howard Shimmel *Howard@JanusStrategyandInsights.com*

Gerard Broussard *gerard@pre-meditatedmedia.com*

Alice K. Sylvester *alice@sequentpartners.com*

Jim Spaeth *jim@sequentpartners.com*