BEST PRACTICES IN DATA TRANSPARENCY

AIMM encourages all data providers to be transparent about their multicultural data quality, coverage, and accuracy metrics. We call upon every data provider – those who directly classify consumers by cultural identity and those who rely on a third-party source for classifying their data – to join the effort for greater disclosure and transparency.

QUESTION	ISSUE	BEST PRACTICES
Are Data Sources Consistent and Appropriate?	There is a wide range of data available from different sources: probability panels/surveys, public records, transactions, searches, social activity, physical visits, cookies, mobile event data, or proprietary algorithms. It is important to know and understand how the underlying data was obtained and how accurate it is.	• Disclose the specific sources of underlying raw data and third-party data sources. • Provide details about the nature of the source data. • If third-party data is matched to native data, disclose match rate.
Are Segment Descriptions Accurate and Understandable?	It is important for data providers and users to understand the true nature of source data. It is also important to understand the intended use of the data. "Hispanic new car intenders" may be the use case, but recent visitors to Spanish-language auto websites may be the provider's actual data source. Conflating the two would be misleading.	• Clearly and accurately describe the method of assignment of multicultural identity. • If applicable, describe the role of name, address/location, online/ offline behaviors. • Be prepared to document the composition of the segment and be open to external assessments of label "claims."
Does the Source Data Provide Good Coverage of All Segments?	Correct representation of the total U.S. population of each multicultural segment is essential. Bias can be introduced due to low incidences of consumers in certain sub-groups. For example, not all consumers or households appear in standard data sources such as credit card holders or retailer loyalty card programs, and there is no reason to believe those who participate are similar to those who don't. Multicultural consumers who are not in the data sources may be less acculturated, leading to bias and inaccuracy.	• Disclose any known gaps/biases in the data. • Demonstrate that the incidences of assigned multicultural identity align with trusted representative and reliable data sources. For example, the U.S. Census profiles the multicultural segments by Census District, Urban/Suburban/ Rural, key high-density metro areas, household income, size, presence of children, etc., which can be used to benchmark the representativeness of the data.

QUESTION	ISSUE	BEST PRACTICES
Is the Data Representative?	Due to the inherent biases and coverage gaps in third-party data for multicultural consumers, it is important to understand how the data is cleaned and calibrated against representative panels/surveys to ensure proper racial/ethnic representation.	• Disclose processes for reducing bias such as filtering, stratification, or quota sampling from the source data. • Disclose when missing source data is imputed and how it is done. • Disclose other data hygiene and quality assurance processes.
Is the Data Timely and Consistent?	Although someone's multicultural identity is unlikely to change over time, the match key by which they are identified can change. For example, multicultural consumers move at higher rates than the general market, so location data must be refreshed regularly. Multicultural data should be refreshed regularly to account for changes in mobility and household composition – at a minimum, updated and validated annually.	• Disclose when and how data refreshes are made. • Provide both the average time since last refresh of the dataset and the frequency distribution of identities in the data segment by refresh latency. • Disclose validation study results.

BEST PRACTICES IN IDENTITY ASSIGNMENT ACCURACY AND VALIDATION STUDIES

AIMM strongly encourages data providers to validate and disclose their accuracy and coverage metrics on a regular basis. There are a number of ways of validating segments and datasets.

AIMM strongly supports benchmarking against self-reported data. AIMM and Media Rating Council believe that for marketing purposes, self-reported identity and language preference data should not be dismissed. Regardless of what a consumer's birth certificate or DNA may say, it's how they self-identify that makes products and their messages more or less relevant and appealing. Increasing social mobility, blended households, and diminished geographic homogeneity further support the need to depend upon self-report. Visual verification, as provided by in-person interviews, can add another layer of certainty.

Self-reported cultural identity is a high-order deterministic measure. Probabilistic measures can also be useful for many use cases but should be validated against self-report.

QUESTION	ISSUE	BEST PRACTICES
How Should Multicultural Data be Validated?	There are different ways providers define race/ethnicity, such as first name, surname, country of origin, English proficiency, U.S. Census definitions, neighborhood, as well as expert AI systems and algorithms. Benchmarking has shown substantial differences in data coverage and accuracy generated by the different methods. All methods should be validated routinely.	**Four approaches are considered best practices for validating multicultural identity assignments:** 1. **Cross syndicated source verification** (e.g., MRI-Simmons with self-identified individuals) 2. **"Truth" dataset comparison** (e.g., client first-party data with known, self-identified individuals and attributes from a representative source) 3. **For modeled segments, comparison to holdout samples of self-identified individuals** 4. **Audit from independent third-party sources** (e.g., Neutronian, Truthset, or providers that can validate with self-report intercept studies, such as Jolt or Lucid) **In all cases, the standard of accuracy is self-report.**
How Accurate Should Multicultural Data Be?	The validation study will reveal how good the data is, but how good is good enough? The need for accuracy and coverage varies with the use case. Benchmarking has shown that it is reasonable, for broadly defined cultural identities, to expect accuracy of at least 67 percent. With this in mind, AIMM recommends a minimum accuracy rate of 67 percent. Higher is better. We expect this low bar to be raised over time as industry practices improve.	To be considered a Hispanic, African-American, Asian-American, or other multicultural segment, at least 67 percent of records in the segment must be accurate and verified as that target. 67 percent is the minimum concentration of multicultural consumers/records within a segment necessary to be called that particular segment.
Can the Accuracy and Coverage of Modeled Audiences Be Validated?	A marketer's need for reach often requires that a data-based target audience segment be extended through modeling. Validation studies will reveal accuracy and coverage trade-offs between probabilistic and deterministic approaches.	• Providers should disclose details about the underlying base data, how the match process works, and match rates. • Validation studies should reveal the coverage and accuracy of modeled segments.

QUESTION	ISSUE	BEST PRACTICES
Does Data Matching Reduce Cultural Identity Accuracy and Coverage?	Cultural identity data is frequently matched to other consumer data for targeting purposes. Inevitably, data is lost in the process. Unless the data loss is random with respect to cultural identity, it will also reduce accuracy and introduce bias into the resultant matched data.	• Providers should disclose data resolution, linkage, and appending processes. • The accuracy and coverage of the resulting data should be validated.

STANDARDIZED VALIDATION STUDIES DISCLOSURE

AIMM recommends that data providers routinely validate and disclose information about multicultural data accuracy and coverage. This proposed "True Transparency" grid standardizes communication of validation study approaches and results. We encourage all marketers and data providers to use the grid and consider its application beyond multicultural data. All advanced advertising targets should have validated metrics.

Proposed AIMM Standardized "True Transparency" Grid for Segment Validation

Target	Validation Method	Validation Study Date	Validation Source Data (Benchmark)	Percent Coverage	Percent Accuracy
Segment 1					
Segment 2					
...					