EXECUTIVE SUMMARY

Audience measurement is the basis for half a trillion dollars of advertising spend. It addresses two basic questions about audiences of media properties: How large are they and who are they? These seemingly simple questions have become increasingly more difficult to answer as advertisers look to reach specific targeted audiences, requiring media sellers to measure ever more granular audiences. This short paper reviews current digital audience measurement methods and how they meet the demands of the modern media environment.

THIS REPORT WILL DETAIL

1 DIFFERENCES BETWEEN PANEL-BASED AND DIRECT MEASUREMENT

2 CHALLENGES WITH PANEL-BASED MEASUREMENT

3 QUANTCAST’S UNIQUE METHODOLOGY

KNOW AHEAD. ACT BEFORE.”
THE EVOLUTION OF AUDIENCE MEASUREMENT

Audience measurement was first used in the 1920s to measure radio listenership. Participants in a measurement company’s panel kept diaries of their radio listening, and the collected data was used to estimate the total listenership for a program. That estimation was primarily done via extrapolation. While the data collection techniques changed, this same basic method, extrapolating from a sample, was applied to television in the ’50s, then online in the ’00s.

Only in the last decade, with lower data costs and more computing power, has direct measurement become feasible. As the name implies, direct measurement measures the media property directly via a digital beacon. The publisher incorporates the beacon into their content so it is activated with every content consumption event. The beacon can provide data about every directly measurable attribute, such as visits, geography and platform. While digital traffic measurement has largely transitioned to direct measurement, many measurement services still rely on panel-based extrapolation for audience attributes, such as demographics.
CHALLENGES WITH PANEL-BASED MEASUREMENT

Today, most digital publishers sell advertising with some form of targeting. Advertisers typically ask that their campaigns appear in specific sections or in front of specific target audiences. Panel-based measurement can struggle to accurately represent these more granular audiences, since a panel might comprise only 1% of the population being measured. Here we discuss some of the challenges with panel-based measurement.

1. SAMPLE BIAS

Panel-based methodologies rely on the sample being representative of the population. For instance, if 5% of the population are tennis fans, the panel should also be comprised of 5% tennis fans. When the panel is not representative of the population being measured, it is known as sample bias. One of the significant challenges of panel-based measurement is recruiting a panel that does not result in sample bias. For example, if the panel is recruited using free NASCAR tickets, it would overrepresent the population’s enthusiasm for racing and likely underrepresent the population’s interest in tennis.

2. MEASURING SMALLER AUDIENCES: SECTIONS AND AUDIENCE SEGMENTS

As advertisers and agencies demand more targeted media buys, publishers wish to measure audiences for specific sections, such as the sports pages, or audience segments, such as High-Income Moms or In-App Purchasers. By their nature, panel-based measurement solutions lose accuracy when measuring smaller audiences, because it is hard to ensure correct representation of the smaller audience in the panel.

Example (Fig 1.): Of an Internet population of 100 million, yoursite.com typically receives 200k users, and 20k users visit the sports section. With a panel representing 1% of the population, only 200 users from the panel are likely to visit the yoursite.com sports section. Even if the gender of those 200 users were known, it would be challenging to accurately infer gender for the whole section, and inferring age or income would be even less accurate.
3. MEASURING AUDIENCES WITH MULTIPLE ATTRIBUTES

When advertisers want to reach college-aged females from Texas, they aren’t interested in reaching females, Texans, and college-aged people separately—they want to reach an audience with all three of those attributes. Panel-based measurement quickly loses accuracy as you layer on more attributes, because the representation of that desired audience on the panel is small—the panel will include females, but it will include fewer who are college-aged, and fewer still who are from Texas. Extrapolating against this small sample is likely to produce a poor result.

Example (Fig 2): LargeApp typically has 1MM users, but only 1,500 are Texan, college-aged females. With a panel representing 1% of the population, only 15 panel members meet these criteria, not allowing any reasonably accurate extrapolation.

4. AUDIENCE DE-DUPLICATION

The goal of audience measurement is to measure people, not cookies or other identifiers. The challenge for audience measurement services is that people typically use multiple devices and browsers, which generate multiple cookies and identifiers. The measurement service must de-duplicate those cookies and identifiers to arrive back to an accurate count of people. Panel-based audience-measurement services manage this de-duplication by examining the usage characteristics of panel members who visit a property. For instance, if the audience in the panel tends to clear their cookies more often than the general population or owns more mobile devices than average, that would inflate the de-duplication factor and result in undercounting the number of people visiting the property.

QUANTCAST’S UNIQUE METHODOLOGY

Quantcast employs direct measurement to capture traffic and other directly measurable attributes, then determines other audience characteristics, such as demographics, through a technique called statistical modeling. For example, starting with a set of users with a known gender, Quantcast infers the gender of new users based on their similarity to the known users, scored against hundreds of data points. Quantcast can employ this technique because it sees each U.S. online user on average 600 times a month, and those additional data points, while not providing direct demographic data, provide a strong signal about user similarity.
HOW DOES QUANTCAST STATISTICAL MODELING WORK?

1. Start with reference data from registrations and surveys
2. Examine characteristics of reference and measured users (which can include sites visited, apps used, app usage frequency, content categories, device type and many more)
3. Infer attributes for each user based on similarity to reference users
4. Validate inferences against reference data and external census data

<table>
<thead>
<tr>
<th>FACTORS</th>
<th>VISITED SITE X</th>
<th>VISITED SITE Y</th>
<th>USED APP Z</th>
<th>VISITED CATEGORY K</th>
<th>DEVICE TYPE M</th>
<th>KNOWN MALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE ATTRIBUTES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISITOR 1</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>X</td>
</tr>
<tr>
<td>VISITOR 2</td>
<td>●</td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACTORS</th>
<th>VISITED SITE X</th>
<th>VISITED SITE Y</th>
<th>USED APP Z</th>
<th>VISITED CATEGORY K</th>
<th>DEVICE TYPE M</th>
<th>KNOWN COLLEGE GRAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE GRAD ATTRIBUTES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISITOR 1</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>VISITOR 2</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

[REPEAT PROCESS FOR EACH ATTRIBUTE]

Note: example for illustration only; actual models include hundreds of factors
OVERCOMING CHALLENGES FACED BY PANEL-BASED MEASUREMENT

By taking a different approach, Quantcast’s methodology overcomes many of the challenges encountered by panel-based measurement.

Panel-based measurement is sensitive to sample bias because it assumes the panel represents the audience being measured. Quantcast’s methodology is based on an assumption that users who share an attribute, such as gender, behave similarly in a detectable fashion.

Limited by their size, panels can’t accurately represent the smaller, more targeted audiences that advertisers demand. Compared to a panel, Quantcast’s statistical modeling can accurately infer attributes for a relatively larger group of people, enabling accurate measurement of even the smallest audiences.

Finally, panel-based measurement services de-duplicate users based on the attributes of their panel, which is subject to sample bias. Quantcast de-duplicates users by examining multiple data points across all sites and apps, such as time period, observed visit frequency, visit source and property type, so is less subject to the potential bias of a sample.
THE MIGRATION TO DIRECT MEASUREMENT

Until recently, direct measurement has been cost-prohibitive to do at scale. Panel-based measurement has been a cost-effective means to measure a large population, using a small set of sample data.

Today, lower computing and data costs have changed that equation, where direct measurement is not only possible but necessary, given how publishers are increasingly adopting segment-based audience sales and new media platforms.

Quantcast has been providing direct measurement coupled with statistical modeling since 2007 and in that time has refined its collection and modeling techniques to provide consistent, accurate traffic and audience-profile data across any digital media platform, for free.

LEARN MORE ABOUT AUDIENCE MEASUREMENT—CONTACT US AT MEASURE@QUANTCAST.COM.