



Introduction

Ensuring data quality means paying attention to the details at every phase of every market research project. At Zanthus, we focus on the finer points so our clients can confidently make the big decisions.

Here, we offer a full look at our data quality best practices for quantitative studies. And we invite your comments and feedback on our approach, as we continually work to improve our methods.

Study Design

When designing a custom quantitative study, numerous considerations are taken into account so that clients' research objectives are met with the highest standards of quality. Sampling approach, screening criteria, survey questions and analysis methods are designed so that:

- Clients will have actionable information. We find out up-front what actions they are actually able to take as the result of survey findings.
- Survey questions measure what we want them to measure (or, in research parlance, maximize reliability and validity).
- We can conduct appropriate analyses and statistical tests on the resulting data.

This involves putting our expertise to work, doing our homework, and working closely with our clients to generate hypotheses to test. While it's beyond the scope of this paper to discuss these factors in-depth, these considerations are paramount to laying the foundation for quality research results.

Sample Design

Probability vs. non-probability samples. There are two main methods of sampling: probability and non-probability. While this topic is not always discussed in the context of data quality, selecting the right sampling method and properly managing it ultimately affects the ability of the final data set to accurately represent the target audience. Therefore, we consider sample design to be another key foundational element of data quality.

While probability sampling—typically obtained through random digit-dialed (RDD) telephone surveys—is considered the gold standard for obtaining results that can be generalized to the wider population, it is not often used due to cost and other practical considerations. Plus, RDD surveys are not without their own issues for a number of reasons, including the large percentage of households that do not have a landline telephone (in the U.S., about one in four). As a result, non-probability sampling via web survey has become the norm for many market research practitioners. But this method brings its own set of challenges.

The following discussion outlines the considerations involved when selecting a sampling approach, plus our methods for maximizing representativeness when non-probability samples are used.

To start, Table 1 summarizes the advantages, disadvantages and recommended applications for each sampling approach.

Table 1. Sampling Approaches

Approach	Administration Methods	Advantages	Disadvantages	Recommended For
<p>Probability sampling</p> <p>Each respondent has a known chance of random selection.</p>	<p>Primarily random-digit telephone dialing.</p> <p>Limited use in web surveys, typically involves address-based sampling with mailed invitation, or random selection from client lists.</p>	<p>Best way to project findings to larger population.</p>	<p>Can be expensive, time-consuming, especially for narrowly-defined audiences.</p> <p>Can lead to coverage errors (for example, wireless-only households may be omitted), non-response bias and low engagement, as willingness to take phone surveys declines.</p>	<p>Studies providing inputs to high-level strategic business decisions.</p>
<p>Non-probability sampling</p> <p>Respondents elect to participate through panels, communities or other methods, including online “intercepts.”</p>	<p>Most web surveys.</p> <p>In-person intercepts.</p> <p>Telephone studies dialed from opt-in or other non-random lists based on interests or other factors.</p>	<p>Respondents increasingly prefer online surveys over phone surveys, reducing response bias and enhancing engagement.</p> <p>Survey may be more relevant to respondent’s interests.</p>	<p>Like probability sampling, can suffer from coverage errors, non-response bias and low engagement.</p> <p>May be difficult to ensure representativeness due to inability to randomly select from population.</p>	<p>Studies that will be quota sampled or weighted based on probability sample proportions.</p> <p>Studies targeting online, technically comfortable audiences.</p>

Non-probability sampling representativeness. When use of probability samples is not feasible, as with most web surveys, we use sample that is drawn to match key known demographic or other characteristics of the target audience, based on our prior research or from a reputable secondary source such as U.S. Census or the Pew Internet & American Life project. For consumer studies, these target variables may include demographics such as: gender, age, education, household income, state or place of residence, and presence of children under 18 years old in the home. It can also be useful to control for attitudinal variables where the distribution is known in the target population, such as attitude about adopting new technology, or social values. In fact, many sample providers and others in the research sampling industry are conducting research on so-called “stabilization” variables that can be used to better achieve sample representativeness in the future.

In some cases, traditional sample sources may not adequately reflect the demographic and/or attitudinal characteristics of the target audience. For this and other reasons, online sample providers now offer web-sourced sample—sometimes referred to as “real-time web intercepts” as alternatives (or adjuncts) to traditional panel sample.

Currently we recommend using web-sourced sample sources when it's useful to reach particular target audiences that are less prone to joining survey panels, such as college students. A downside of this approach is the inability to validate respondent identity without extra time and cost. And, the process of validation for web-sourced sample may in fact introduce another type of error by rejecting those who are difficult to validate against third-party databases, especially young adults who have limited public records and credit histories. As the availability and breadth of this sample source evolves, these limitations may become less of a concern, but for now we recommend using web-sourced sample on a limited basis to fill in any gaps in available panel sample.

Response bias management. For non-probability samples, once potential respondents are invited to participate in our surveys, we enforce measures that allow us to correct for any response bias by demographic or other characteristics. If the target audience's characteristics are well-known, we can set quotas on completed interviews ("completes") to ensure the make-up of the resulting data set reflects the population we have in mind.

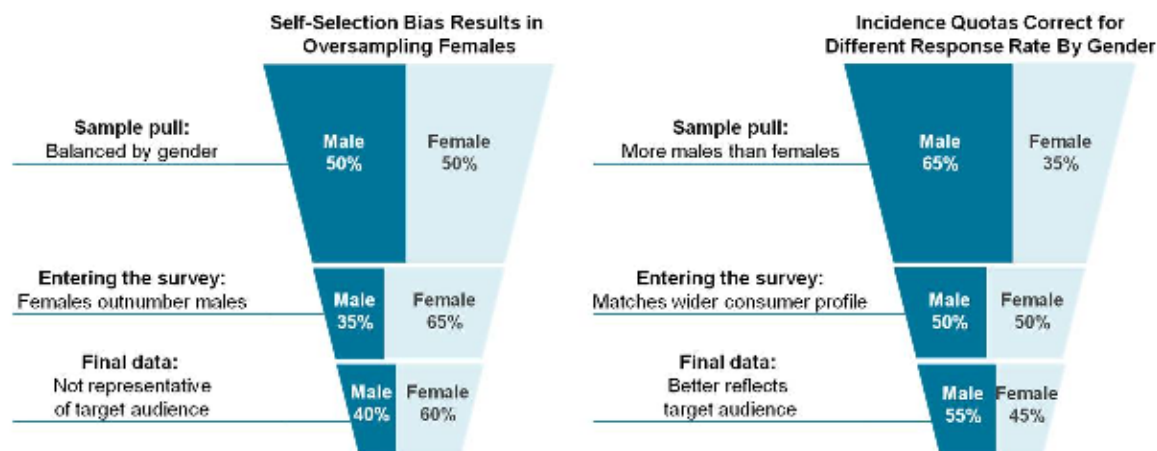
Typically however, the target audience's profile is not known in advance. In those cases, we monitor the characteristics of those entering the survey via incidence (or "click") quotas and make corrections in our sample blend by those characteristics as needed.

For example, let's assume that we don't know the gender proportion within our target audience (say, users of a particular product). That means we can't set a quota on completed interviews by gender. And, let's assume that females are likely to outnumber males in terms of invitation click-through rates. If we evenly balance the sample "pulled" (invited to the survey) by gender to mirror the general consumer population, the make-up of our resulting data set will be biased towards females because of the higher click-through rate of females.

To correct for response bias by gender, we can instead invite males in greater numbers than females to ensure that the group entering the survey (as monitored by the incidence quotas) best matches the wider consumer profile. In so doing, the gender proportion obtained once respondents are screened better reflects the actual proportion in the target audience.

See Figure 1 for an illustration of this methodology.

Figure 1. Using Incidence Quotas to Correct for Response Bias



Unique respondents. We ensure that all sample used is de-duplicated at the sample source before surveys are administered. Once respondents have entered the survey program, we use a unique code assigned to each individual on the panel (or other list source) to prevent individuals from taking the survey more than one time. We also use computer “cookies” to allow only one completed survey from a particular computer. This is especially helpful for studies where multiple sample sources are used. Using cookies also allows respondents to leave the survey and return where they left off.

Validated identity. Country of residence and individual respondent identity for panel-sourced web survey sample are validated against third-party databases by our sample partners before admission to the survey. For projects utilizing real-time web intercepts (via invitations on other web sites to take the survey), we offer respondent identity validation as an optional add-on service to our clients for an additional fee.

Sample partners. We partner exclusively with leading online (web) sample companies that provide responses to the ESOMAR “26 Questions to Help Research Buyers of Online Samples,” a well-regarded guideline for helping researchers determine the suitability of sampling practices for specific research needs. We maintain close contact with our partners to ensure we are continually apprised of all procedures and changes that may affect our projects.

We understand that some of our clients prefer using certain sample or panel vendors due to interest in specific respondent identification validation, de-duplication and/or engagement measurement methods. Because we work with all leading vendors and provide full transparency about sample sources to our clients, these needs are easily accommodated.

Regardless of the sampling approach used for a specific project, all sampling details are clearly communicated to our sampling partners using a standard Zanthus form to ensure full understanding of all specifications and requirements.

Questionnaire Design

Invitation, pre-screening & screener. Our surveys start with a generic introduction and appropriately ladder screening questions to mask the topic of the survey until respondents have qualified. This mitigates respondents' ability to "game," or complete the survey despite lack of qualifying credentials. It also helps to minimize self-selection bias. All survey communications with respondents by our sample partners—including invitations, pre-screeners and screeners—are reviewed and approved in advance by a Zanthus consultant.

Maximizing respondent engagement. Respondents who are the most engaged with the survey are the most likely to provide thoughtful and accurate responses. To enhance engagement, we design understandable and thought-provoking survey instruments by employing the following approaches:

- **Keeping the average survey length to a minimum**—ideally under 20 minutes—to reduce respondent fatigue.
- **Using respondent-appropriate language** that is casual, non-threatening and free of industry jargon. For phone interviews, we include clear interviewer instructions, including pronunciation guides, to ensure consistency across interviews.
- **Conducting well-executed foreign language translations with trusted partners** and validating those translations with the client team.
- **Measuring and benchmarking respondent satisfaction** for each survey as part of a program to continually improve our survey designs.

Identifying low quality responses. While our goal is to design engaging surveys, it's still important to seek out and eliminate any low-quality survey responses when they do occur due to fraudulent or satisficing behavior. Offenders are reported to our sample partners for elimination from future Zanthus surveys.

We utilize some or all of the following methods for identifying suspect data, depending on the study:

- **Overall survey completion time.** Cases below minimum established survey times (using the standard of mean minus two standard deviations, or other appropriate criteria) are removed from the sample.

- **Survey section completion time.** This may be enforced on survey sections that require careful review of text, graphics or video before proceeding. Respondents are only allowed to proceed after the necessary time has elapsed; those attempting to speed through are warned and if they re-offend, dropped.
- **Flatlining or straightlining.** Those giving the same answer to a series of scaled questions are flagged for review and possible removal.
- **“Trap” questions.** We insert fictitious brands or products among other valid choices to catch respondents not paying attention to response options.
- **Contradictory answers.** We flag these by using oppositely-worded attitudinal or behavioral statements that reveal inconsistent responses. For example, we might show the statements, “I find it difficult to use my phone” and “Using my phone is easy,” and compare the answers. Fact-based questions also may be used, such as asking for the respondent’s zip code at both the start and end of the study.
- **Open-ended questions.** We include at least one question in each survey requiring respondents to provide text-based responses in their own words. These are reviewed for quality and consistency.

“Strike” rules for eliminating suspect data based on these methods are established during the questionnaire design process, and enforced during fielding through automated processes when appropriate. Suspect data are removed and replaced. This approach reduces the likelihood that substantial cleaning will be required in the analysis phase.

Programming & Testing

Questionnaires undergo a rigorous review at every stage of development, starting when they are in draft form, through each revision, and finally, after programming. Research consultants and survey programmers collaborate to make sure logic instructions are clear, efficient and properly programmed. Survey programs are reviewed for input by fielding partners during development and just prior to launch.

Zanthus handles most survey projects from beginning to end, with our own in-house programming team. For projects that involve outsourced programming, our in-house staff conducts rigorous testing to assure adherence to our internal standards. We provide clients with the ability to test all surveys prior to launch.

Fielding

Pre-testing. When time and budget allow, Zanthus recommends pretesting the programmed survey with four to six respondents using a think-aloud protocol, where a consultant listens as the respondent voices his or her reactions in real time while taking the survey. This typically extends the research timeline by two to three days. Cost varies depending on audience specifications.

Pretesting is especially important when stimuli are complex or technically challenging, cultural differences could impact understanding, and/or uncertainty about the comprehensiveness and appropriateness of response options or question wording is high.

Soft launch. All surveys are launched with a small audience initially so data can be reviewed the following day to ensure accuracy of the program logic, to determine survey length, and to calculate preliminary incidence of the target audience. After this review is complete, the survey is launched to the larger audience.

Data monitoring during fielding. Open-ended responses and other key questions are monitored throughout the fielding period by the project consultant to ensure quality.

CATI procedures. Phone interviewing supervisors are fully briefed to ensure a thorough understanding of the screener and questionnaire. Then, live monitoring of telephone interviewers is conducted by a Zanthus consultant and CATI field partner supervisor within 24 hours of project start; the client also is invited to participate in a monitoring session.

Monitoring live interviewing serves two important functions. It gives us a chance to observe whether interviewers are introducing bias or misrepresenting any questions, and to identify questions that may not be as effective as expected. Each live monitoring session is followed by a debriefing session with the field supervisor, and if needed, revisions to the questionnaire (and additional testing).

Data Processing

Once fielding has ended, we conduct a final review of survey responses to confirm logical consistency and quality of open-ended responses. We also do a final check at this stage to confirm all respondents are unique. Then, we apply standard decision rules for cleaning numerical data to minimize the impact of outliers. Data tables are then checked against the questionnaire and the tabulation specifications.

At this stage, weighting may be applied to bring demographic or other proportions in-line with known characteristics of the target population.

Analysis & Reporting

We deliver reports with clear, compelling visuals and well-considered conclusions. We don't assume our clients will draw what seem like "obvious" conclusions on their own after reading just the detailed findings. We seek input from clients on all final deliverables, with an emphasis on providing business context to research findings.

All graphics and data in our reports are carefully checked by a second staff member for accuracy, and a senior consultant other than the main researcher reviews each report in its entirety prior to client delivery.

A technical report is made available to all clients (typically in the full report appendix) demonstrating the impact of applying data quality practices to the project. The report includes the individual study's respondent satisfaction score relative to benchmark satisfaction measures, the number of cases removed due to low engagement, and other relevant measures.

Post-Project Wrap-Up

Most of our projects do not involve handling personally-identifiable information (PID) from survey respondents. For projects that do require we process personal information (such as projects utilizing client-supplied customer lists), we have procedures in place to retain only the necessary information for the least amount of time, and to protect it in encrypted, password-protected files.

Conclusion

When all parties—researchers, clients, vendors and respondents—are fully engaged, quality research is the result. Engaging these stakeholders means appreciating each party's unique contribution to the effort, and building in systematic processes and approaches that take advantage of these varied perspectives.

At Zanthus, we continually examine what "doing the right thing" means for each project phase—from design through data collection, reporting and project wrap-up—and ensure that any vendor partners continually uphold the same high standards.



Sign up for Updates

We periodically send updates to inform our clients and partners who want to be inspired and informed by our insights via email. If you would like to receive email updates, please sign up today on our website.

 [Sign up for updates](#)

About Us:

Zanthus is a full-service market research-based consulting firm serving high-tech companies. Headquartered in Portland, Oregon, we are particularly well-known for our uniquely compelling combination of industry and research expertise as well as our commitment to reliable market research methods and analytical techniques.


 www.zanthus.com

Questions?

Feel free to contact us by phone or email:


Mike Lehman, President

 mlehman@zanthus.com

 971/404.0275 ext: 102

Carole Wiedmeyer, Partner

 cwiedmeyer@zanthus.com

 971/404.0275 ext: 122

