

PHRONESISPARTNERS

Enhancing Survey Data Quality

In Healthcare Primary Research



Contents

Foreword	01
The data quality imperative in healthcare research	02
Challenges in maintaining survey data integrity	04
Advanced strategies for ensuring high-quality data	07
Conclusion: Elevating integrity for strategic impact	16
About Phronesis Partners	17

Foreword

Healthcare and life sciences companies increasingly rely on primary research collected through surveys and in-depth interviews to understand the perceptions and preferences of physicians, hospital administrators, payers, and other medical professionals. Insights from these key stakeholders play a crucial role in shaping decisions related to pharmaceutical marketing, medical device development, and health policy.

However, as the complexity of research environments grows, and risks such as survey fraud and disengaged respondents continue to rise, the need for robust data quality management becomes even more critical.

In this report, we take a closer look at these contemporary challenges and provide advanced, actionable strategies to ensure that primary research consistently produces accurate, reliable, and impactful results.



Biraj Ganguly
Executive Vice President,
Healthcare & Life Sciences
Phronesis Partners

The data quality imperative in healthcare research

In healthcare and life sciences, data quality is paramount because the stakes are exceptionally high.

In healthcare and life sciences, data quality is paramount because the stakes are exceptionally high. Primary research guides decisions such as which drug candidate to invest in, how to position medical devices, or what reimbursement strategy a health plan should pursue. Unlike secondary data (clinical trial results or insurance claims), primary data directly captures perceptions and experiences from human respondents, revealing why physicians prefer certain therapies or how hospital executives view new diagnostic tools.

High-quality survey data means information is accurate, consistent, and representative of the target population. For a pharmaceutical company surveying oncologists about treatment preferences, quality data ensures respondents truly are oncologists in the relevant specialty, understand the questions, and provide thoughtful answers. For a medical technology firm conducting interviews with hospital procurement officers, quality means each interview reflects genuine opinions of qualified decision-makers without interviewer bias or respondent confusion.

Poor data quality can have serious consequences. Inaccurate or fabricated responses may suggest market demand that doesn't exist or, more dangerously, hide safety concerns that should inform healthcare delivery. Even subtle issues like disengaged respondents can distort findings. For instance, if several participants speed through a survey without reading questions, their random answers can obscure genuine differences in opinion between segments, potentially leading researchers to conclude "no significant preference" exists when it actually does. A pharmaceutical company might subsequently invest millions in a messaging campaign that fails because it was based on inaccurate attitude data.

Regulatory bodies and industry associations increasingly focus on data integrity in research. Organizations like Intellus Worldwide (the industry association for healthcare insights) and codes of conduct such as EphMRA¹ emphasize maintaining high standards of data accuracy, respondent privacy, and ethical conduct. Professional organizations like the American Association for Public Opinion Research (AAPOR) provide comprehensive standards and best practices for maintaining survey research integrity and methodological transparency².



Probability Lab
Which sum comes up most often when you roll a pair of dice?
Roll the dice 100 times and record the results in the table below.
3. Determine the relative frequency of each sum.

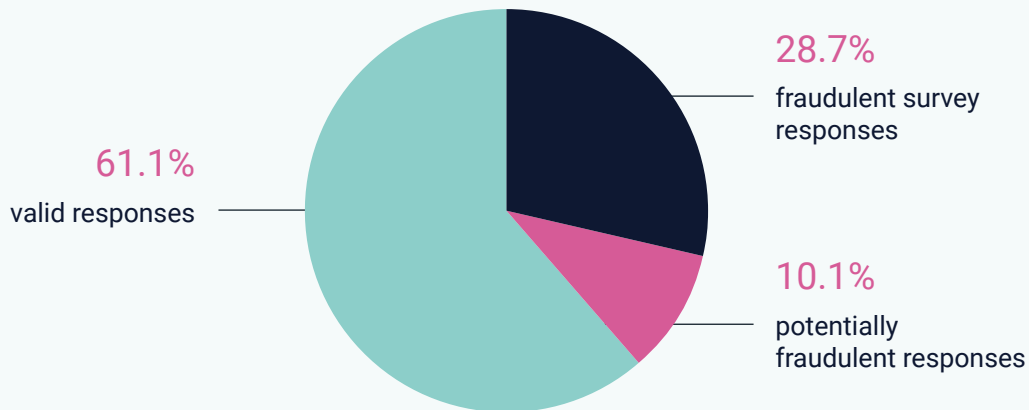


Challenges in maintaining survey data integrity

Identifying and recruiting the right respondents

Reaching busy professionals such as specialized surgeons, pharmacists, or insurance directors is difficult. Ensuring respondents truly fit the target profile is even harder. In online surveys, there's a significant risk of fraudulent respondents misrepresenting themselves to earn incentives. In telephone recruitment, gatekeepers or support staff might attempt to participate in place of the intended expert.

If unqualified or fake participants slip through, the data becomes instantly unreliable. A 2019 study by Ballard et al.³ found that 28.7% of submitted web-based survey responses were fraudulent and another 10.1% were potentially fraudulent. This underscores how pervasive the problem has become, especially in web-based health research.



2019 study by Ballard et al. on fraud detection in web-based surveys

Respondent engagement and fatigue

Healthcare experts often face survey fatigue. Physicians might receive invitations to numerous studies each week. If a survey is too long or repetitive, even genuine respondents might start giving cursory answers or drop out. Disengaged respondents exhibit behaviors like:

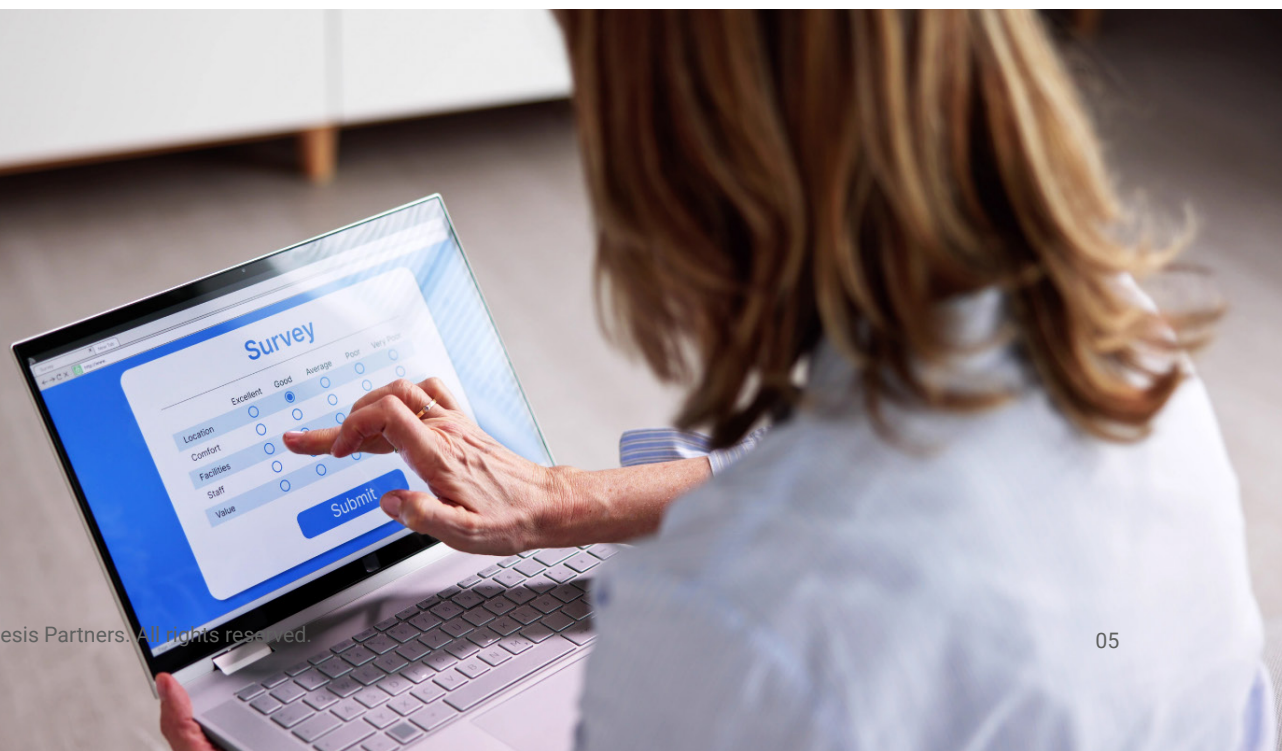
- Speeding through questions
- Straightlining (selecting the same answer for every item in a grid)
- Giving nonsensical open-ended responses
- Satisficing (providing merely acceptable rather than carefully considered answers)

In self-administered surveys, participants control their pace, and some will “expedite” participation to finish quickly. This lack of engagement directly jeopardizes data quality.

Complex topics and questionnaire design

Healthcare surveys often involve complex medical or scientific topics. Poorly worded questions, misuse of terminology, or overly technical language can confuse respondents, leading to incorrect answers. For instance, asking physicians to “Rate your familiarity with the latest INS-237 formulation” might cause confusion, whereas

“How well do you understand the mechanism of action of Drug X?” could yield clearer responses. Conversely, oversimplifying questions to avoid jargon might yield responses that lack nuance. Crafting questionnaires that are both precise and clear to professional audiences requires careful balance.



Multiple data collection modes

Research projects often use mixed methods, perhaps an online survey followed by telephone interviews, or multi-country studies using different collection modes across regions. Ensuring consistency across modes is challenging. Questions asked by an interviewer might be interpreted differently than the same questions on a computer screen.

This is particularly relevant in global pharmaceutical studies, where researchers might interview respondents in emerging markets by phone while collecting data from developed markets online. Without careful alignment, mode effects can compromise data comparability.

Data fraud and professional respondents

The lure of incentives can attract malicious behavior. Survey fraud has become increasingly sophisticated in online research. Individuals may create fake identities (posing as physicians or payers) to take surveys multiple times, or even deploy automated bots for completion.

Even with legitimate healthcare professional panels, there are “professional respondents” who rush through surveys to maximize honoraria, contributing low-quality data. This challenge is particularly acute in healthcare research, where incentives tend to be higher to reflect respondents’ professional value.

Privacy and compliance concerns

Healthcare professionals are often guarded about sharing information, especially regarding patient data or competitive intelligence. If a survey lacks clear compliance with privacy regulations (like HIPAA or GDPR) or fails to provide confidentiality assurances, respondents might withhold honest answers or drop out completely.

For example, oncologists might hesitate to discuss off-label prescribing practices without explicit confidentiality guarantees. Surveys on sensitive topics (such as prescribing influenced by incentives) may face deliberate under-reporting due to social desirability bias. Ensuring anonymity and ethical data handling directly impacts response quality.

Advanced strategies for ensuring high-quality data

Rigorous participant recruitment and verification

The foundation of quality data is quality respondents. No amount of cleaning or analysis can salvage data collected from the wrong people.

Screening and double-screening

Implement robust screening questionnaires that go beyond demographics or job titles. Include questions only genuine target professionals could answer accurately. For instance, when surveying cardiologists, ask about the approximate number of angioplasties performed monthly or familiarity with actual drugs like Entresto versus a fictitious medication like “Cardexol ” (which legitimate cardiologists would not recognize).

For critical studies, consider implementing a double-screen process:

- Initial online screener to assess basic qualifications
- Verification phone call or email validation for qualified respondents
- Cross-checking against professional directories or LinkedIn profiles

Though this adds time and cost, it dramatically improves sample integrity.

Leverage professional recruitment for niche audiences

For high-level or niche stakeholders (such as formulary committee directors or rare disease specialists), panel sources may be limited. Professional recruiters or subject matter expert networks can help through referrals, professional associations, or direct outreach with personalized communication that establishes trust and verifies the person’s role.

While more costly, this approach ensures hard-to-reach respondents are exactly who they claim to be, and often yields more engaged participants due to the personal recruitment approach.

Manage incentives wisely

In healthcare research, honoraria typically exceed consumer survey incentives, reflecting the value of respondents' time. Research on survey incentive strategies⁴ demonstrates that while higher incentives generally produce higher response rates in physician surveys, beyond a certain point it follows with diminishing returns. For example, in one of our studies comparing \$125 versus \$150 incentives found only a 2% increase in response rate for the higher amount, despite a 20% increase in cost.

If rewards are excessive (e.g., over \$200 for a 20-minute physician survey), they may attract fraud; if too low (under \$50 for specialists), genuine experts may decline or rush through.

Aim for fair-market honoraria (based on market, audience profile, and client's specified FMV), and clearly communicate that payment depends on meaningful participation, not mere completion. Some studies explicitly state that responses will be reviewed for quality before payment is processed, which can dissuade those solely interested in quick compensation.

Intelligent survey design and pre-testing

The design of the survey instrument plays a pivotal role in data quality. Well-designed questionnaires minimize respondent error, maintain engagement, and incorporate quality checks.

Keep surveys purposeful and concise

With busy healthcare professionals, shorter, focused surveys yield better data than sprawling ones. Every question should connect to a research objective. To achieve this:

- List all research objectives before drafting questions
- Remove questions not clearly linked to objectives
- Eliminate redundant items
- Consider splitting very long surveys into separate studies

When respondents feel their time is respected, they remain more thoughtful throughout. Additionally, concise surveys reduce fatigue effects that lead to satisficing.

Use clear and professional language

Ensure question wording is clear, neutral, and appropriately tailored to the audience's expertise level. In life sciences research, using correct terminology builds credibility. For example, oncologists expect drug names and clinical terms to be spelled correctly and used in proper context. Each question should be tested for clarity:

- Will all specialists interpret this consistently?
- Does terminology match what practitioners use daily?
- Are there regional or practice setting differences in interpretation?

If ambiguity exists, rephrase or add definitions (perhaps as hover-over explanations in online surveys).

Incorporate logical flow and sections

Structure surveys with related topics grouped together in a logical progression. Begin with easier questions to engage respondents, then move to more complex ones. Grouping by theme (current practices, opinions on treatments, demographics) helps orient respondents and reduces cognitive burden.

Use skip patterns to bypass irrelevant sections based on previous answers. For instance, if a respondent doesn't prescribe a certain drug, skip detailed questions about that medication. This shortens the perceived length for respondents and prevents data contamination from people answering questions that don't apply to them.

Embed quality control questions

Advanced questionnaires include "trap" or attention-check questions to gauge if respondents are paying attention. Quality check types include:

- **Attention traps:** Within a matrix about treatment preferences, include "If you are reading this carefully, please select 'Strongly Agree'."
- **Consistency checks:** Ask the same factual question in different ways at different points. For example, ask a physician's patient volume directly, and later ask for a percentage breakdown that should sum to approximately the same number.

- **Red-herring options:** Include a deliberately implausible option in a list, such as "Quantum Resonance Therapy" in a list of standard cancer treatments.

Use these sparingly, as savvy respondents might find excessive traps patronizing. A few well-placed checks help differentiate attentive participants from careless ones.

Pre-test the survey instrument

Piloting surveys is crucial but often overlooked. Have a small group of representative participants take the survey before full launch.

During pre-testing:

- Observe completion time compared to expectations
- Note questions causing hesitation or confusion
- Check if any questions are frequently skipped
- Look for overuse of “Other” options, suggesting incomplete answer lists

Cognitive debriefing, which involves asking pilot participants to explain how they interpreted questions, ensures consistent understanding. For example, when testing a diabetes management survey, ask physicians to explain terms like “treatment intensification” or “glycemic control” to confirm uniform interpretation.

The goal is to eliminate ambiguity or bias before real data collection begins, optimizing the instrument for quality responses.

Real-time quality control during data collection

Once data collection begins, whether via online survey or CATI interviews, implement real-time quality monitoring rather than waiting until completion to discover problems.

Digital fingerprinting and bot detection

In online surveys, technology can help prevent and detect fraudulent entries in real time. Digital fingerprinting captures information about a respondent’s device and browser (without collecting personally identifiable information) to identify duplicate participation or known fake accounts. Advanced authentication methods for web-based research⁵ include electronic verification strategies, fraud detection protocols, and social media validation methods.

- IP addresses and geolocation
- Proxy/VPN usage that might hide true location
- Browser fingerprints and cookies
- Hidden trap questions (sometimes called “honeypot questions”) that are invisible to human respondents but might be completed by automated programs

If a survey receives multiple entries from the same IP range or shows suspicious patterns (like numerous completions in unrealistically short timeframes), these systems can automatically flag or block those cases. Given the high fraud rates reported in studies like Ballard et al. (28.7% fraudulent submissions), these technical defenses have become essential.

Real-time monitoring dashboards

For larger studies, set up live dashboards to monitor incoming data and key metrics, including:

- Number of completes versus targets
- Drop-out rates per question
- Average completion time overall and by section
- Quality flag triggers (speeding, straightlining, etc.)
- Sample composition versus quotas

If a particular question shows unusually high

drop-out rates, you might pause fieldwork to investigate whether the question is confusing or the response options are malfunctioning. Similarly, if average survey duration is 5 minutes when expected to be 15, this may indicate respondents are speeding through or skipping content.

Treat the first 10-15% of data collection as a sentinel sample. Reviewing it in real time allows you to catch anomalies early and make adjustments before collecting the full dataset.

Employ “speed bumps” and warnings

Implement speed checks in online surveys to identify overly fast completions. If a physician progresses through pages much faster than average, trigger a warning message: “You seem to be answering very quickly. Please ensure you’re giving each question careful consideration for accurate results.” These “speed bump” messages remind respondents that quality matters and that their attention is being monitored.

Similarly, failing an attention check question can trigger a warning or automatically disqualify the respondent, depending on your protocol. In CATI interviews, train interviewers to recognize signs of disengagement (like consistently brief, hasty answers) and offer to reschedule at a better time rather than collecting poor data.

Interviewer oversight and training

For telephone or qualitative research, quality control focuses on supervising interviewers. Ensure all interviewers receive thorough training on the questionnaire or discussion guide, including the intent behind each question.

Supervisors should monitor samples of live calls or recordings to verify interviewers are:

- Following the script without significant rewording
- Not leading respondents toward certain answers

- Probing appropriately when answers are vague
- Recording responses accurately

Provide immediate feedback if interviewers are found ad-libbing excessively or failing to probe where needed. Some CATI centers employ technology allowing supervisors to see data being entered in real time to identify unusual patterns.

Hold daily debriefs with interviewers during fieldwork to address any difficulties with questions, which might indicate a need for clarification across the interviewing team.

Post-collection data validation and cleaning

Even with careful design and active monitoring, no study is completely free of problematic data. Post-collection validation is crucial for identifying and handling remaining issues before analysis.

Statistical identification of outliers and inconsistencies

Modern data analysis software can efficiently scan for anomalous responses. For quantitative surveys, use algorithms to flag outliers at both respondent and question levels. For example:

- If most physicians report seeing 50-100 patients weekly but one entered “1,000”, investigate this outlier
- Calculate completion time and flag respondents below the 10th percentile as potential “speeders”
- Check for straightlining patterns in grid questions using variance metrics
- Examine logical consistency across related questions

These statistical and logical checks can generate quality scores or lists of suspect cases for further review.

Response quality scoring systems

Develop composite scoring systems for respondent quality by assigning points for each quality issue:

- Speeding through the survey (e.g., <50% of median completion time)
- Straightlining on matrix questions
- Failed attention check questions
- Inconsistent answers to related questions
- Poor quality open-ended responses
- Suspicious geolocation or digital fingerprint

Respondents exceeding a threshold score might be classified as “unreliable.” Establish these rules before reviewing results to avoid selection bias. Having an objective scoring method provides an audit trail showing data was removed for defined reasons, preserving study integrity.

Item-level cleaning vs. respondent-level

Sometimes respondents provide mostly good data with just a few problematic answers. Rigidly removing every respondent with any issue could unnecessarily reduce sample size. A nuanced approach determines when to clean at the item level versus dropping entire respondents.

For example, if a physician's survey is excellent except for one impossible value, you might set that specific answer to missing data while keeping

their other responses. Conversely, if a respondent failed multiple attention checks and provided contradictory answers throughout, removing the entire case is appropriate.

Document all cleaning rules clearly: "We removed any respondent who failed 2 or more of the following 5 quality criteria..." This transparency helps explain sample size changes if questioned later.

Manual review by domain experts

After automated cleaning, have subject-matter experts review flagged cases, especially in healthcare research where context matters. For instance, a respondent's answer pattern might seem unusual until an expert recognizes it reflects an unconventional but valid practice setting (e.g., a pediatric endocrinologist using different protocols than adult specialists).

Expert review can also identify duplicate respondents, different entries with suspiciously similar answer patterns might represent the same person using different email addresses. This three-phase approach (design checks, algorithmic detection, expert review) ensures good data isn't unnecessarily discarded while bad data is properly excluded.

Documentation and transparency

Maintain comprehensive logs of all data cleaning steps:

- Total responses collected
- Number excluded for specific quality reasons
- Item-level corrections made
- Final analysis base size

This documentation serves both integrity and analytical purposes, as analysts need to know if base sizes changed for certain questions. When presenting results, briefly mention quality assurance measures: "We applied rigorous data validation, which involved removing respondents failing quality benchmarks, leaving N=200 valid interviews for analysis." This assures stakeholders that insights are built on solid data foundations.

Ensuring quality in qualitative research and CATI

While much discussion focuses on quantitative surveys, qualitative research and CATI studies require their own quality assurance strategies.

Recruit qualified and willing participants

For expert interviews, such as conversations with Key Opinion Leaders (KOLs) or payers, verify credentials and assess willingness to discuss topics openly⁷. **During screening, evaluate communication ability with questions like:**

“Please describe your role in three sentences”

“What aspects of [topic] are you most involved with?”

“How would you characterize the current state of [therapeutic area]?”

If someone provides only terse responses during screening, they likely won't deliver rich insights in the actual interview. Ensure diverse perspectives within your sample frame, which includes incorporating both national and regional payers in reimbursement research.

Use skilled moderators/interviewers

Qualitative data quality depends heavily on moderator expertise. Skilled healthcare interviewers know how to probe vague responses effectively: Respondent: “I'm concerned about efficacy.”

Skilled moderator: “Could you elaborate on what aspects of efficacy concern you? Are there specific endpoints or patient populations where you see limitations?”

They clarify questions without leading respondents and gently challenge inconsistent statements by referencing earlier comments. Train CATI interviewers to handle common respondent queries neutrally and consistently.

Conduct mock interviews during training and certify only interviewers who demonstrate competence. Quality in qualitative research means capturing authentic, detailed perspectives, which requires skilled facilitation.

Guide design and note-taking

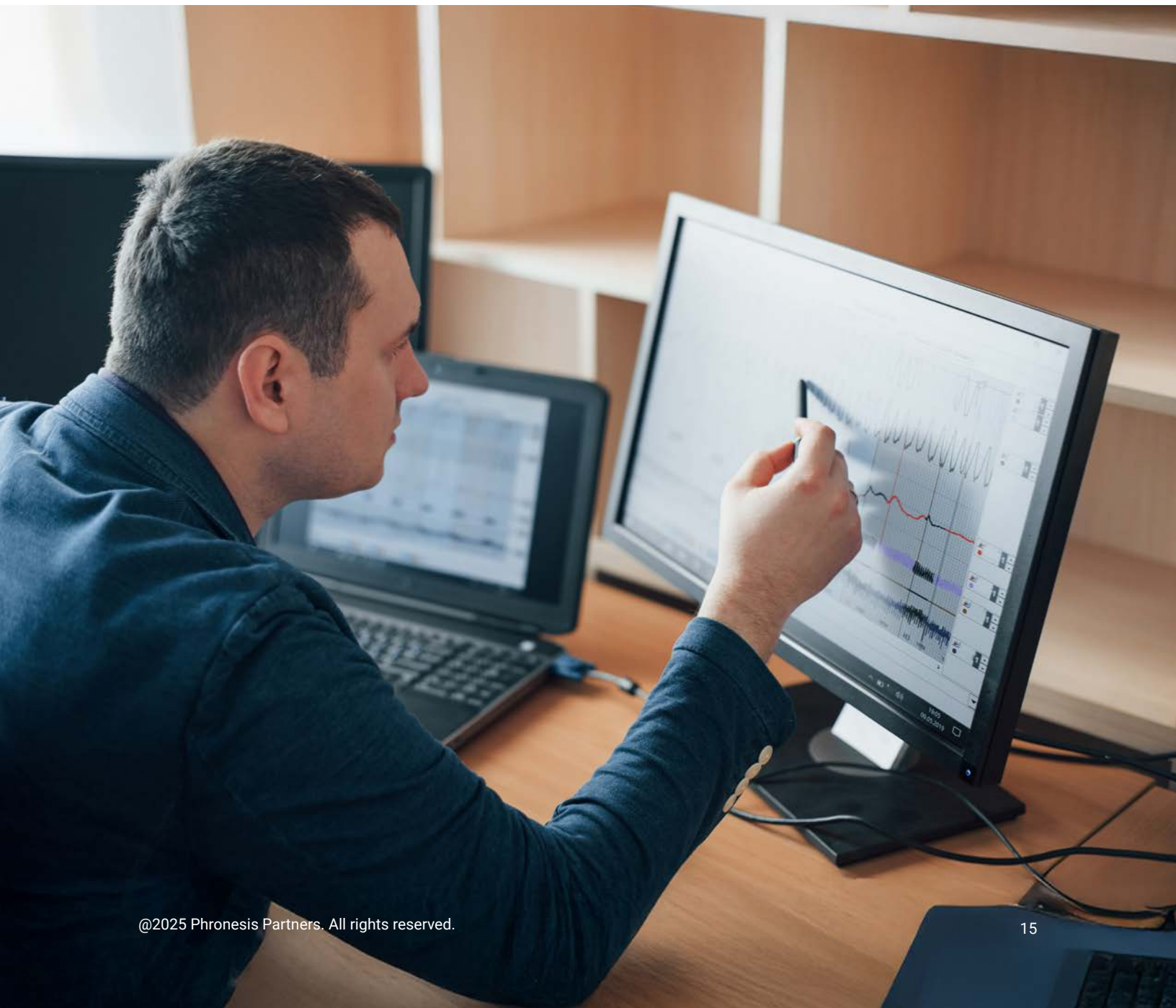
Design discussion guides with the same care as survey instruments. They should flow logically, use appropriate terminology, and avoid biasing responses. While qualitative guides are semi-structured (allowing exploration of emerging topics), they should clearly identify must-cover themes and key probes to ensure consistency across interviews.

Record all interviews (with permission) and transcribe them rather than relying on memory or hasty notes. High-quality transcripts (ideally reviewed for accuracy by a second person) provide reliable raw data for analysis. For focus groups, employ dedicated note-takers to capture non-verbal cues and group dynamics that transcripts alone might miss.

Validation through triangulation

Strengthen qualitative data integrity through triangulation, using multiple analysts or data sources to validate findings. Have two researchers independently code initial interview transcripts for themes, then compare results. Agreement on key themes increases confidence; disagreement prompts refinement of the coding framework.

Mixed-methods approaches can further enhance validity. For example, follow quantitative surveys with qualitative interviews to explain surprising results or confirm survey interpretation aligns with interview findings. While triangulation primarily addresses insight validity rather than data cleaning, it elevates research integrity by ensuring conclusions aren't artifacts of one method or researcher's perspective.



Conclusion: Elevating integrity for strategic impact

In healthcare and life sciences research, data quality forms the foundation for credible insights and informed decision making.

This paper has outlined comprehensive strategies, from participant recruitment to data analysis, that collectively establish a robust quality assurance framework.

Implementing these strategies involves investing in stronger panel partnerships, sophisticated survey platforms equipped with advanced fraud detection, interviewer training, and rigorous validation procedures. Although these efforts require resources, the costs of inadequate data quality, misguided decisions, compromised research integrity, or repeated studies far outweigh such investments.

As healthcare becomes increasingly digitized and research inquiries more complex, prioritizing data integrity grows even more critical. By adopting these quality-focused practices, research teams can produce robust findings capable of withstanding scrutiny and driving meaningful improvements in healthcare delivery and patient outcomes.

Healthcare decisions informed by primary research directly impact patient lives, from therapy development to care delivery and reimbursement structures. Ensuring data quality at every step provides the confidence needed to make these consequential decisions effectively.

References

¹EphMRA. (2016). Code of Conduct. European Pharmaceutical Market Research Association

²American Association for Public Opinion Research. (2022). AAPOR Standards Best Practices. Retrieved from AAPOR.org

³Ballard, A. M., Cardwell, T., & Young, A. M. (2019). Fraud Detection Protocol for Web-Based Research Among Men Who Have Sex with Men: Development and Descriptive Evaluation. *JMIR Public Health and Surveillance*, 5(1), e12344

⁴Mizes, J. S., Fleece, E. L., & Roos, C. (1984). Incentives for increasing return rates: Magnitude levels, response bias, and format. *Public Opinion Quarterly*, 48(4), 794–800

⁵Young, A. M., Ballard, A. M., & Cooper, H. L. F. (2021). Methods for Authenticating Participants in Fully Web-Based Mobile Health Studies. *JMIR mHealth and uHealth*, 9(8), e28232

About Phronesis Partners

At **Phronesis Partners** data quality is foundational and integral to every aspect of our primary research process. Our commitment extends beyond compliance; it is central to our mission of providing clients with insightful, reliable, and

actionable results. Discover how we can support your research objectives and help you achieve excellence in data quality and strategic impact at www.phronesis-partners.com/solution/healthcare-research.

Let's keep in touch

Read more insights and thought leadership at www.phronesis-partners.com/resources

Contact us at www.phronesis-partners.com/contact

Follow us on social media:



PHRONESISPARTNERS

www.phronesis-partners.com

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.