Sampling And Collecting Quantitative And Qualitative Data

Posted on May 29, 2018

View this essay online:
https://academic-master.com/sampling-and-collecting-quantitative-and-qualitative-data/

Introduction

Sampling and data collection determine whose experiences enter a study and what evidence can legitimately support its conclusions. A sophisticated statistical analysis cannot correct a sample that systematically excludes important groups, and a large dataset cannot compensate for a measure that does not represent the construct it claims to assess. Research design therefore begins by linking the research question, target population, unit of analysis, sampling strategy, measurement plan, and intended form of inference. Probability sampling is designed around known selection probabilities and is especially useful when researchers want to estimate characteristics of a larger population, while non-probability sampling is often appropriate for exploratory, qualitative, hard-to-reach, or theory-building work (Teddlie & Yu, 2007). Data may then be collected through questionnaires, interviews, observations, records, tests, sensors, or mixed methods. Each choice creates different strengths and possible errors. Credible research requires transparency about these choices rather than assuming that one sampling method or data-collection tool is universally superior.

Defining the Population and Building a Sample

A study should distinguish the target population from the accessible population and the sampling frame actually used to recruit participants. The target population is the group about which the researcher ultimately wishes to draw conclusions, such as nurses in a region, university students in a system, or adults with a specific diagnosis. The accessible population is the portion that can realistically be reached, while the sampling frame is the list or operational mechanism from which eligible units are identified. Coverage error occurs when the frame excludes eligible people, includes ineligible entries, or duplicates individuals. Researchers must also identify the unit of analysis clearly. Responses may be collected from individuals while conclusions concern households, schools, hospitals, or organizations, and confusing the respondent with the analytic unit can lead to invalid interpretation. Eligibility criteria, geography, time period, and key definitions should be established before recruitment, because changing them after seeing the results can introduce bias and make the final sample difficult to interpret.

Probability and Non-Probability Strategies

Probability sampling uses a random selection mechanism and gives each eligible unit a known, nonzero chance of selection, although those chances do not always have to be equal. Simple random sampling gives each unit the same probability, systematic sampling selects a random starting point and then every *k*th unit, and stratified sampling divides the population into meaningful groups before sampling within each stratum. Cluster and multistage designs select naturally occurring groups such as schools, clinics, villages, or geographic areas and can reduce fieldwork costs, although clustering often increases sampling variance because people within a cluster may resemble one another. Non-probability methods serve different purposes. Purposive sampling selects information-rich cases, maximum-variation sampling seeks contrasting experiences, criterion sampling requires a shared characteristic, and convenience sampling recruits those who are easiest to reach. Snowball approaches can help researchers contact hidden populations, but network dependence may influence which participants become visible. The method should follow the research objective rather than a blanket rule about which approach is “best.”

Sample Size, Quantitative Collection, and Measurement

Quantitative sample size should be justified by the primary outcome, expected variability, effect size, desired precision, significance level, statistical power, design effect, subgroup analyses, and anticipated missing data. A larger sample reduces random sampling error under appropriate designs but does not automatically remove selection or measurement bias. Questionnaires are common because they can standardize responses across many participants, yet their quality depends on clear wording, appropriate recall periods, balanced response options, accessible formatting, and a direct connection between each item and the construct being measured. Pilot testing and cognitive interviewing can reveal whether respondents interpret questions as intended. Online surveys may reduce cost and speed data collection, but they can exclude people with limited connectivity, digital literacy, language access, or assistive technology. Response rates are useful but incomplete indicators of quality; the more important question is whether nonrespondents differ systematically from respondents on characteristics related to the study outcome. Administrative records and sensors introduce different problems, including coding error, missingness, and measurements created for purposes other than research.

Qualitative and Mixed-Methods Collection

Qualitative research uses interviews, focus groups, participant observation, documents, diaries, photographs, and other open-ended methods to examine meaning, experience, interaction, and context. Semi-structured interviews balance consistency with flexibility, while focus groups reveal

shared norms and disagreement but may discourage disclosure of sensitive experiences. Observation can compare reported behavior with practice, and documentary evidence can show how institutions formally represent policies or decisions. The researcher is part of the qualitative instrument because identity, assumptions, language, tone, and relationships can influence what participants disclose and how the material is interpreted. Reflexive notes, interviewer training, iterative guides, and negative-case analysis help make that influence visible. Mixed-methods designs combine quantitative and qualitative evidence when the integration serves a clear purpose. A survey may establish a population pattern followed by purposive interviews that explain it, or qualitative findings may identify constructs used to build a later questionnaire (Onwuegbuzie & Collins, 2007). Integration should occur in interpretation rather than leaving two unrelated studies side by side.

Reliability, Validity, Bias, and Missing Data

Reliability concerns consistency, whereas validity concerns whether evidence supports the intended interpretation of a measure or conclusion. Test-retest reliability examines stability over time, inter-rater reliability examines agreement among observers, and internal consistency evaluates whether items intended to form a scale behave coherently. A measure can be highly reliable while consistently measuring the wrong thing, so reliability is necessary in many contexts but not sufficient for validity (Drost, 2011). Validity includes content, construct, criterion, internal, and external dimensions depending on the research question. Missing data can also threaten credibility when participants skip items, leave a study, or cannot be linked across records. Complete-case deletion may bias estimates if missingness is related to the exposure or outcome. Prevention through concise instruments, reminders, flexible administration modes, and careful fieldwork is preferable, while analysis may require weighting, multiple imputation, or sensitivity analysis. Researchers should describe likely directions of bias rather than using generic statements that a sample was simply “small.”

Ethics, Inclusion, and Transparent Reporting

Ethical sampling requires voluntary participation, fair selection, privacy protection, risk minimization, and a meaningful right to stop or skip questions. Recruitment through teachers, supervisors, clinicians, or community leaders can create perceived pressure, so the separation between participation and access to services, grades, employment, or care should be explicit. Sensitive studies may collect information about health, immigration status, illegal activity, trauma, or family conflict, which increases the need for secure storage, minimal collection of identifiers, clear confidentiality limits, and appropriate referral procedures. Inclusive administration also requires accessible materials, translated versions where needed, screen-reader compatibility, readable layouts, and alternatives for participants who cannot use the dominant mode. Transparent reports should identify the sampling frame, recruitment period, inclusion criteria, response and attrition, collection mode, weighting, instrument development, and deviations from protocol (American Association for Public Opinion Research, 2023). Preregistration can distinguish planned confirmatory analyses from later

exploratory findings, while qualitative projects can maintain an audit trail explaining how evolving interpretations influenced later recruitment and data collection.

Conclusion

Sampling is not a preliminary administrative step but a central part of the logic connecting evidence to conclusion. Probability methods support population inference when selection probabilities, coverage, implementation, nonresponse, weighting, and variance estimation are handled appropriately. Non-probability methods can provide deep explanatory or theoretical insight when their purpose and limits are stated honestly. Questionnaires, interviews, observations, records, and digital measures each produce characteristic forms of error, and sample size cannot compensate for biased recruitment or invalid measurement. Reliable research therefore combines a well-defined population, defensible sampling, suitable collection methods, ethical participation, accessible administration, valid measurement, careful handling of missing data, and transparent reporting. Mixed-methods designs are strongest when their samples and analyses are connected to a clear integration purpose rather than combined superficially. The credibility of both quantitative and qualitative research ultimately depends on whether researchers can explain whose evidence was collected, how it was collected, what it can support, and where uncertainty remains.

References

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105–124.

Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12*(2), 281–316.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77–100.

American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*.