RESEARCH · MARKET AND BUSINESS RESEARCH

Primary Versus Secondary Market Research

The difference between primary and secondary data, the strengths and weaknesses of each, how to assess them, and how to combine them in business research.

ModQu research article · 1,635 words · 15 references (APA 7th edition)

Every research project has to answer an early practical question: can the information needed be found in data that already exist, or must new data be collected? Data collected by the researcher for the problem at hand are **primary data**. Data originally collected by someone else, for another purpose, are **secondary data** (Hox & Boeije, 2005). The distinction concerns the relationship between the data and the current study, not the type of data: a customer survey is primary data for the firm that commissioned it and secondary data for anyone who later reuses it. This article explains each type, compares their strengths and weaknesses, and sets out how to combine them in business research.

Secondary research

Secondary research, sometimes called desk research, is the analysis of existing information. Stewart and Kamins (1993) divide secondary sources into **internal** sources, which are held inside the organisation, and **external** sources, which are produced by others.

Internal sources include sales records, customer databases, loyalty programme data, website analytics, customer service logs, warranty claims and earlier

research reports. They are often the cheapest and most relevant data a firm has, and they are frequently underused.

External sources include:

- **Official statistics:** census data, labour force surveys, household spending surveys and trade statistics published by national statistical agencies such as Statistics Canada.
- **Industry and trade publications:** reports from trade associations, regulators and industry bodies.
- **Commercial and syndicated data:** retail scanner data, consumer panels and market reports sold to many clients by research firms.
- **Academic research:** published studies, many of which also release their datasets for reuse.
- **Company publications:** annual reports, investor presentations and regulatory filings by competitors and customers.
- **Digital traces:** online reviews, social media posts and search trends.

Retail scanner data illustrate how valuable secondary sources can be. Bucklin and Gupta (1999) interviewed 41 executives at data suppliers, packaged-goods manufacturers and consulting firms. They reported that scanner data had been used most successfully for decisions about consumer promotions, trade promotions and pricing, and less successfully for decisions about product strategy, advertising and distribution.

Primary research

Primary research collects new data designed for the specific question. The main methods are:

- **Surveys**, administered online, by telephone, by post or face to face, which measure attitudes, behaviour and characteristics across a sample (Groves et al., 2009).
- **Interviews and focus groups**, which explore views and reasons in depth with smaller numbers of people (Krueger & Casey, 2015).
- **Observation and ethnography**, in which researchers watch how people use products in natural settings (Arnould & Wallendorf, 1994).
- **Experiments**, including online A/B tests, which assign people at random to different versions of an offer to measure cause and effect (Kohavi et al., 2020).

Because the researcher controls the design, primary data can be matched exactly to the decision: the right population, the right questions, the right time period and the right level of detail.

Comparing the two

Primary and secondary data compared

How the two types of data differ on six practical criteria

Criterion	Primary data	Secondary data
Fit to the question	Designed for this decision	Collected for another purpose
Cost	Higher: design and fieldwork	Low; often free or already held
Time	Weeks to months	Hours to days
Control over quality	Full: method is known	Judged from documentation
History and scale	Limited to this study	Long time series, large samples
Confidentiality	Owned by the firm	Usually open to competitors

Based on Hox & Boeije (2005); Stewart & Kamins (1993); Johnston (2014) ModQu.com

Figure 1. How primary and secondary data compare on six practical criteria. Based on Hox and Boeije (2005), Stewart and Kamins (1993) and Johnston (2014); table by ModQu.

Cost and speed. Secondary data are usually much cheaper and faster to obtain. Much official data is free, and internal data have already been paid for. Johnston (2014) notes that secondary analysis is an efficient use of data that took considerable resources to collect. Primary research takes time to design, field and analyse.

Fit. The main weakness of secondary data is that they were collected for someone else′s purpose. Definitions, categories, time periods and populations may not match the question (Hox & Boeije, 2005). A government survey may group households in age bands that do not match a firm′s target market, or report at a national level when the firm needs regional figures.

Quality and control. The researcher who collects primary data knows exactly how it was done. With secondary data, the user must judge quality from documentation that may be incomplete. Stewart and Kamins (1993) recommend asking who collected the data, for what purpose, how they were collected, what was collected and when, and whether the results are consistent with other sources.

Scale and history. Secondary sources can offer sample sizes, time series and coverage no single firm could afford. A national census or a long-running retail panel provides trends over many years and comparisons across regions.

Ownership. Primary data belong to the organisation that collected them, so competitors do not see the results. Most external secondary data are available to anyone who can find or buy them, so they rarely provide a competitive advantage on their own.

Assessing secondary sources

Before using a secondary source, check the following:

- **Purpose:** Why were the data collected? Data produced to promote an industry or product may present it favourably.
- **Method:** How was the sample drawn and how large was it? What was the response rate?
- **Definitions:** How were key terms, such as "household", "user" or "small business", defined?
- **Timeliness:** When were the data collected? Markets can change quickly.
- **Consistency:** Do other sources give similar figures? Large disagreements need explanation before either source is used.

Digital traces need particular care. Groves (2011) describes such "organic" data as by-products of digital systems that were not designed for research, so they capture only part of behaviour and their meaning is often unclear without context. Online reviews, for example, come from customers who chose to write one, who may differ from customers in general.

Assessing primary research

Primary research has its own risks. A survey can suffer from coverage error (the sampling frame misses part of the population), sampling error, non-response error and measurement error from poorly worded questions (Groves et al., 2009). Online panels have made surveys cheaper and faster, but a task force of the American Association for Public Opinion Research cautioned that results from non-probability online panels should not be treated as estimates of population values in the same way as probability samples (Baker et al., 2010). Yeager et al. (2011) compared probability and non-probability samples against benchmark data and found the probability samples were consistently more accurate. For decisions that depend on an accurate population estimate, how the sample was recruited matters as much as its size.

Using both: the usual sequence

Most textbooks recommend beginning with secondary research and moving to primary research only for the gaps that remain (Malhotra, 2019). The sequence has several advantages:

- Secondary data may answer the question outright, saving the cost of a new study.

- They help define the problem and the population, and suggest hypotheses to test.
- They provide benchmarks, such as population profiles, against which a primary sample can be checked and weighted.
- They show what is already known, so that primary research can focus on what is new.

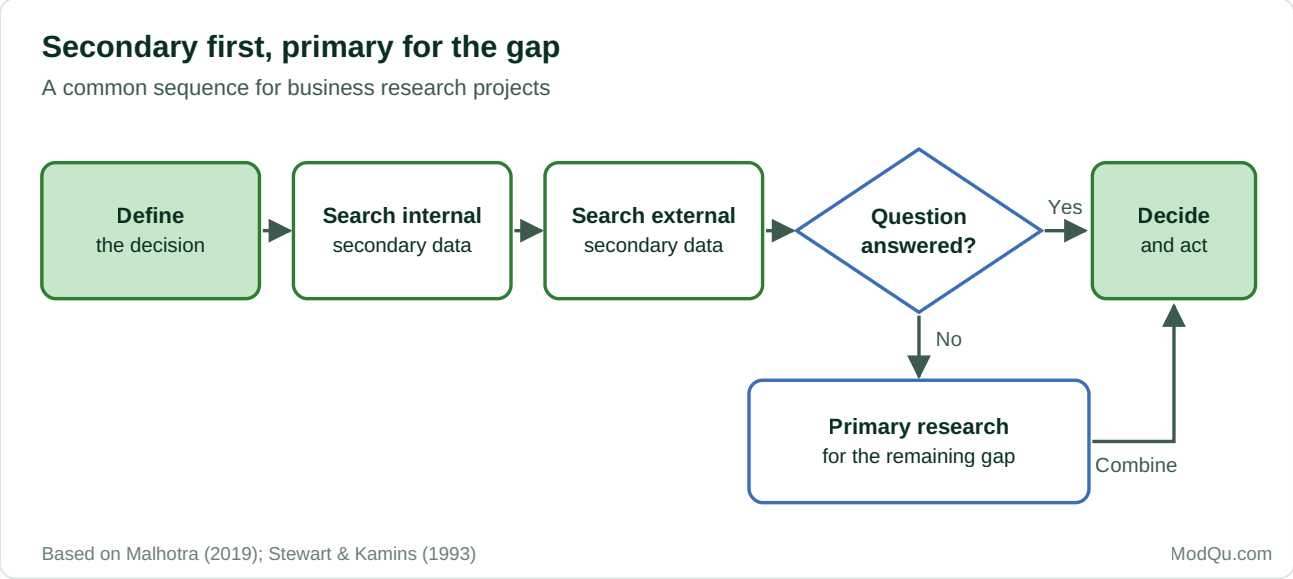


Figure 2. The usual sequence: exhaust relevant secondary sources first, then collect primary data only for the questions that remain. Based on Malhotra (2019) and Stewart and Kamins (1993); diagram by ModQu.

When primary research is essential

Some questions cannot be answered from existing data at any price. Primary research is usually needed when:

- **The product or idea is new.** No historical data exist on how customers respond to something they have never seen.

- **The question concerns the firm′s own brand or customers.** General market data rarely show how a specific brand is perceived or why its customers leave.
- **Cause and effect must be established.** Existing data show associations; an experiment is needed to show that a change in price or message causes a change in behaviour (Kohavi et al., 2020).
- **Reasons and motives matter.** Sales data show what people bought, not why. Interviews and well-designed surveys can explore the reasons.

Collecting primary data well requires attention to how people answer questions. Tourangeau et al. (2000) describe answering a survey question as four tasks: understanding the question, retrieving relevant information from memory, forming a judgement and reporting an answer. Errors can enter at each stage, which is why question wording, order and response options are tested before fieldwork. The choice of mode also matters. Dillman et al. (2014) show that the same question can produce different answers online, by telephone and on paper, and recommend designing questionnaires so that they work consistently across the modes used.

A worked example

A regional coffee roaster is considering a line of ready-to-drink cold brew sold through grocery stores.

Secondary research first. The firm′s own sales records show which existing products sell best and in which stores. Official household spending statistics give a picture of spending on beverages by region. Trade association and market reports describe the growth of the ready-to-drink category and the main competing brands. Competitors′ websites and shelf visits give current prices and

pack sizes. Together, these sources indicate whether the category is growing, what price range is normal and which retailers carry similar products.

The remaining gap. None of these sources says whether the roaster′s own customers would buy its cold brew, at what price, or which flavours they would prefer. That is specific to the firm and its brand.

Primary research to fill it. The firm runs a short survey of its existing customers, with a conjoint exercise to test price, size and flavour combinations, and a small in-store trial in a few locations to measure actual sales. The trial matters because stated intentions are an imperfect guide to purchase, particularly for new products (Morwitz et al., 2007).

The secondary work made the primary research smaller, cheaper and better targeted. The primary work answered the questions that no existing source could.

Ethics and data protection

Both types of data raise obligations. Primary research requires informed consent and honest description of the purpose. Reuse of secondary data, especially internal customer data, must respect the terms under which they were collected and applicable privacy law. In Canada, the *Personal Information Protection and Electronic Documents Act* governs how private-sector organisations collect, use and disclose personal information in commercial activities. Aggregated or anonymised data reduce, but do not always remove, these concerns.

Summary

Secondary data are cheap, fast and broad but were designed for other purposes. Primary data fit the question exactly but cost more and take longer. The usual approach is to search what already exists, judge its quality, and then collect new data only for the questions that existing sources cannot answer.

References

Arnould, E. J., & Wallendorf, M. (1994). Market-oriented ethnography: Interpretation building and marketing strategy formulation. *Journal of Marketing Research, 31*(4), 484–504. https://doi.org/10.1177/002224379403100404

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., & Zahs, D. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly, 74*(4), 711–781.

Bucklin, R. E., & Gupta, S. (1999). Commercial use of UPC scanner data: Industry and academic perspectives. *Marketing Science, 18*(3), 247–273. https://doi.org/10.1287/mksc.18.3.247

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly, 75*(5), 861–871. https://doi.org/10.1093/poq/nfr057

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 593–599). Academic Press.

Johnston, M. P. (2014). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries, 3*(3), 619–626.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press. https://doi.org/10.1017/9781108653985

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Sage.

Malhotra, N. K. (2019). *Marketing research: An applied orientation* (7th ed., global ed.). Pearson.

Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales? *International Journal of Forecasting, 23*(3), 347–364.

Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods* (2nd ed.). Sage.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75*(4), 709–747.