

Using Secondary Data in Doctoral Research

By Jim Goes and Marilyn Simon

Find this and many other dissertation guides and resources at
www.dissertationrecipes.com

When considering research topics and research designs, most doctoral students decide to collect primary data. Primary data are data that are proactively gathered for a specific research purpose, and both the choice of instrumentation for data collection and use of the resulting data are focused on addressing the specific purpose and the alignment among the research question, planned analyses, and types of data collection proposed (Simon & Goes, 2013) The most attractive feature of primary data is that they are customized to the particular study being completed, and available for use and analysis once gathered.

There are, however, challenges involved with gathering primary data.. Not only must a population and sample be identified clearly, but as researchers we must convince our target population to respond. People are constantly surveyed, interviewed, asked for feedback, etc., and often resist the time and effort necessary to participate in market or academic research.

Regardless of whether interviews, surveys, focus groups, or some other approach is being used, one of the most difficult problems facing researchers is gaining an acceptable level of response to enable the planned analyses. This may be less difficult with some qualitative designs, like phenomenology, because the necessary sample size is usually smaller. In many cases, however, students end up with fewer responses than needed, and must adapt their analyses and statistical approach to account for a less than ideal sample size.

There is another approach, and it can be quite effective even though among doctoral students it is rarely used. Few students take advantage of **secondary sources of data**, even when they are readily available. Secondary data consists of data that were collected for a different purpose, but can be repurposed for use in a different study. These data may be publicly available for students to use, such as census data, statistical agencies, federal agencies, academic publications, and trade associations. There are also data considered internal to an organization that might be obtained with permission, such as standardized tests in an educational institution, or annual reports from businesses, which may require a royalty or payment to use the data. Data from national, state, and local governments is often available free of charge, although it may take some work to access. In some cases, like that of census data, large portions of the data may even be downloaded off the Internet.

For example, I (Jim) completed my dissertation years ago based on secondary data obtained from the State of California. The data consisted of profile, operational, and business data on California hospitals. Hospitals were required to gather these data and submit it to the state on an annual basis as part of the state regulatory framework. Even better, the data were required to be audited, providing strong assurance of quality. After some digging, I found that the data were available from the state just for the asking. What a deal! Even though it took some work, a lot of phone calls, and even a trip to Sacramento (state capitol) to obtain the data, it was well worth it.

The core challenge in using secondary data usually revolves around finding ways to make the data fit with the purpose of the new study and the research questions being asked. In my case, I was studying processes of strategic and structural change in hospitals over a period of a dozen years. The unit of analysis was the organization (hospital), and very few researchers in my field had access to any sort of longitudinal, organization-level data like this. The real work in using this data was in figuring out how to adapt and apply it to fit my research goals, and how to craft variables and measures that were credible indicators of the kinds of phenomena I was interested in studying. In some cases, I had to transform the data, create constructs and measures by combining multiple indicators, and overall be very inventive coming up with measures that served as proxies for the actual variables I wanted to study.

In the end, it was all worth it. I was not required to gather any primary data, which meant no turn downs by participants, no missing data, and no pleading e-mails to members of the study population asking them to help me. However I did spend far more time on preparing, cleaning, and recoding the secondary data to fit my purposes than I would have with a primary study. The result, however, was a database and a study that was unusual in the field and had scope and breadth far beyond most primary studies, enabling me to answer questions that had not yet been investigated fully because of the absence of primary data on the topic.

So as you are considering the methodology and design for your dissertation or other research study, consider secondary data. Rather than gravitating immediately to a strategy that involves primary data collection, think about how you can gain access to secondary data, already collected and available, that can address your research questions. You may even find that a mixture of primary and secondary data will make for a stronger study and faster time to completion of your study or dissertation.

References

- Cowton, C. J. (1998). The use of secondary data in business ethics research. *Journal of Business Ethics, 17*(4), 423-434.
- Huston, P., & Naylor, C. D. (1996). Health services research: reporting on studies using secondary data sources. *CMAJ: Canadian Medical Association Journal, 155*(12), 1697.
- Nicoll, L. H., & Beyea, S. C. (1999). Using secondary data analysis for nursing research. *AORN journal, 69* (2), 428-428.
- Simon, M. & Goes, J (2013).
- Szabo, V., & Strang, V. R. (1997). Secondary analysis of qualitative data. *Advances in Nursing Science, 20*(2), 66-74.
- Smith, E. (2008). *Using secondary data in educational and social research*. Open University Press.