



SECONDARY DATA IN RESEARCH – USES AND OPPORTUNITIES

¹Fellipe Silva Martins

²Júlio Araujo Carneiro da Cunha

³Fernando Antonio Ribeiro Serra

INTRODUCTION

We wrote this editorial– in both the Iberoamerican Journal of Strategic Management and Podium – we thought about future opportunities for research, especially in a topic that still has room for further development: the use of secondary data. Sports management research in Brazil, which is the key focus of Podium, has gained wide acceptance within the management research academia in Brazil. Along with this maturation process, this now reinforced area needs to step up its game (pun intended) in terms of methodological approaches. This means that both the qualitative and quantitative sides of the equation must now become more focused in improving results. In strategic management, on the other hand, the use of secondary data is now common and widespread, but not as much in Iberoamerican countries. In this editorial, we will focus more specifically on the use and opportunities in employing secondary data.

Cite it like this:

Martins, F., Cunha, J., & Serra, F. (2018). Secondary Data in Research – Uses and Opportunities. *PODIUM Sport, Leisure And Tourism Review*, 7(3), 1-4. <https://doi.org/10.5585/podium.v7i3.316>

¹ PhD in Business Administration University Nove de Julho – UNINOVE. São Paulo, (Brazil). E-mail: silvamartinsfellipe@gmail.com Orcid id: <https://orcid.org/0000-0003-3918-7231>

² PhD in Business Administration University of São Paulo – USP, São Paulo, (Brazil). E-mail: juliocunha@yahoo.com Orcid id: <https://orcid.org/0000-0002-1435-055X>

³ PhD in Materials Engineering and Chemical and Metallurgical Processes from Pontifical Catholic University of Rio de Janeiro – PUC, (Brazil). E-mail: fernandorserra@gmail.com Orcid id: <http://orcid.org/0000-0002-8178-7313>



Why not primary data?

Whereas first-hand data is intrinsically good, as it most likely provides trustworthiness and transparency about the phenomena researchers focus on, it is also likely to be fruit of hard work, expensive to obtain or gather and, overall, limited. To use a personal example (although it was in a different subject), during a masters' degree research one of the authors decided to understand diversification strategies and financial performance in portfolios and personally obtained data from almost 70 companies in Brazil. The only problem is that most of such data is not usually disclosed as it is sensitive to most companies. In result, it took almost a year to convince enough companies to disclose their data. Years later, he discovered that some public and private organizations collected reasonably the same amount of data, and it would have been much easier to convince one or two organizations to share their data instead – saving months of work (not to mention very costly phone bills at the time).

As such, it is probably obvious by now that using secondary data may be a great advantage for researchers. However, its indiscriminate use is something that should be discussed. Its uses most commonly come with a long list of caveats (some of which we will discuss further on) that are not usually discussed, or even reported in academic papers. Taking care of appropriately using secondary data, substantiating their choice, justifying their context and report any limitations are the essential steps to be taken.

What is secondary data?

In simple terms, secondary data is every dataset not obtained by the author, or “the analysis of data gathered by someone else” (Boslaugh, 2007:IX) to be more specific. Secondary data may include data that has been previously gathered and is under consideration to be reused for new questions, for which the data gathered was not originally intended (Vartanian, 2010).

As an example, a researcher wants to understand football match attendance and how to improve communication towards attendees in a stadium and defines a theoretical question. He or

she could gather data first-hand. However, he or she wants to optimize their research chronogram and looks for ready-to-use data and finds an association that collects data about stadium attendance, including at least parts or variables of interest to the research. That dataset may come in handy but was not designed to answer the question.

As a second example, a team of researchers want to understand the influence of regional institutions in location choice of cross-border acquisitions. Whereas gathering data first-hand for dozens of countries is certainly possible (however long this may take), it much easier to find world databases that gather at least a subset of the intended data and complement them with smaller national statistics databases – which was the case in Falaster et al., (2018).

After these examples it should be obvious that whereas employing secondary data is very useful, it also comes along with some serious (but manageable) caveats

Where to find secondary data?

The answer to this question would obviously depend on one's research gap, what kind of theory is being used to explain it, what kind of phenomenon is being analyzed and so on. Since secondary data was not intended for the purposes of the study, care must be taken to select an adequate database accordingly.

First, we would suggest governmental sources – *if* they come from transparent and trustworthy governmental agencies. In that sense, one should use both one's common sense, as well as performing a literature search to verify whether this database has already been successfully used before in academia and. Most governmental agencies run surveys every now and then to collect data that interests them. In this group I would also include international agencies such as the UN and its affiliates that are usually trustworthy. When using secondary data to run tests, especially panel data and time series analysis, you should always observe whether there were changes in the data collection mechanisms, formulae, or other aspects that may change the significance or even the very meaning of your potential findings.



Second, we would suggest data from private entities and agencies. Again, the same caveats from above must also apply to these. Compared to public agencies, not all private research is entirely committed to scientific methods of data collection – you should carefully read every single detail about it before committing to using it. There are some ways of adapting, justifying or transforming such data, but these should always be reported as potential limitations to the study. For international studies, for instance, both the Thomson-Reuters Merger and Acquisitions and the World Bank Enterprise Survey (WBES) databases are excellent sources of information.

Third, one could also consider using data obtained in private international projects. An example is data from the European Commission, an organ that collects all sorts of data about the European Union, and routinely gather some interesting statistics. Another example is the World Accident Database from the Fédération Internationale de l'Automobile (FIA - International Automobile Federation). Either of these databases may answer partially to one's research needs and may also need to be paired with other database or primary data, yet at least some effort and energy is saved in the process.

Fourth, a new avenue that has been increasingly cited and used in secondary data is web scraping. In short, web scraping is an automated process of data extraction from websites (Virgilito & Polidoro, 2017). Its use in management has already come to surface (Yerger, 2014), but much is still to be seen. Some of such databases are available online and one example is the database on downloaded papers from the pirate scientific paper website Sci-hub, which could potentially be used to understand research flows and collaboration, as well as interest in research topics (Bohanon, 2016). While most web scraping will be done through professional data science tools such as R or Python, there are a few “beginner” online tools so that most people (and researchers included) may try their hands at it – one may find a plethora of useful basic tools online.

Fifth, in the overall search for reproducibility in science, most journals started to accept the upload of database researchers gathered for their studies along with other supplementary materials. Most of them nowadays even stimulate, in the name of transparency and trustworthiness, authors to share their databases.

Whereas this trend has still not picked up entirely (maybe due to researchers' jealousy of other people using their hard-earned data, or, perhaps afraid of other researchers finding flaws in their studies), it is already possible to find ready-to-use data on journals' websites.

Interpreting secondary data-based results

As usual, care must be taken whenever using secondary data. Any problems related to the data collection – either faults or mismatches between the original theoretical target and the new study objectives – may possibly influence the interpretation of data. First, it would be the most responsible behavior as a scientist to clearly explain the limitations of using secondary data in the paper. Second, it should be also clear in the text which steps were taken in the use of the original data for the new research gap.

Second, the research should be careful in understanding whether the data is essential for the theory development, whether it reflects solely on the object of the research and whether the intended conclusions may be safely assumed from the data as they are. This is perhaps the greatest risk in using secondary data – making sure the data is statistically useful providing evidence for the theory as in the hypotheses or propositions – for more on hypotheses building, see Ferreira (2013).

Opportunities for research

Whereas commenting this would prove fruitless since the potential uses of secondary data would depend on a deeper knowledge of a given research, we can safely suggest a few points.

First, there is a plethora of theory-driven research, in which secondary data would be welcome as a shortcut for data gathering. Finding a suitable match between one's research needs and databases has become exponentially easier with database sharing online. Second, there is a whole branch of potential phenomenon-driven research that could arise from databases analyses. For instance, imagine finding a database with roughly the data related to a broad research subject. With some curiosity and some data analysis tools and skills it is perfectly possible to find some behavior in the data that has gone unnoticed or may reflect some new ideas for future theory development (even if it is a mere incremental idea).

Third, with some care, proper documentation (to ensure reproducibility and transparency) and



theoretical grounding, it is also possible to manipulate (in the sense of transforming, mind you – i.e., joining, aggregating, subtracting, etc.) variables to reach new conclusions that were not originally possible from the database as obtained.

Final remarks

Using secondary data in research has proved itself a valuable approach to finding suitable data for

REFERENCES

Bohannon J (2016). Who's downloading pirated papers? Everyone. *Science* 352(6285): 508-512. <https://doi.org/10.1126/science.352.6285.508>

Boslaugh, S. (2007). An introduction to secondary data analysis. *Secondary data sources for public health: A practical guide*, 2-10.

Falaster, C. D. ; Martins, F. S. ; Storopoli, J. E. . The Influence of Regional Institutions in Location Choice of Cross-Border Acquisitions. *European Journal Of Scientific Research*, v. 148, p. 188-201, 2018.

Ferreira, M. (2015). *Pesquisa em Administração e Ciências Sociais*. Rio de Janeiro: RJ, LTC.

one's needs. This should be used more often in research in both Iberoamerican countries as well as in sports management studies. As usual, researchers should be careful in selecting secondary data, verifying its suitability, documenting any changes or manipulations in the data, checking whether the data may be safely used to accept or refuse a given set of hypotheses. By doing so, research may be done in a quicker pace, without loss of quality and reliability.

Vartanian, T. P. (2010). *Secondary data analysis*. Oxford University Press.

Virgillito, A., & Polidoro, F. (2017). Big Data Techniques for Supporting Official Statistics: The Use of Web Scraping for Collecting Price Data. In *Data Visualization and Statistical Literacy for Open and Big Data* (pp. 253-273). IGI Global.

Yerger, C. (2014). Nontraditional undergraduate research problems from sports analytics and related fields. *Involve, a Journal of Mathematics*, 7(3), 423-430.