SAGE researchmethods cases

Conducting a Correlational Study Using Publicly Available Data to Investigate a Nationwide Nocebo Effect

Contributors: Kate MacKrill & Keith J. Petrie
Pub. Date: 2020
Product: SAGE Research Methods Cases: Medicine and Health
Methods: Research questions, Experimental design, Correlation
Disciplines: Medicine
Access Date: March 24, 2020
Academic Level: Postgraduate
Publishing Company: SAGE Publications Ltd
City: London
Online ISBN: 9781529733815
DOI: https://dx.doi.org/10.4135/9781529733815

Abstract

The dominant methodologies in most areas of research are experiments and correlational observational studies. In health psychology and specifically research on the nocebo effect, this is no different. Although experiments are considered the gold standard, they still have limitations, and correlational studies can be conducted instead. In certain instances, researchers conducting a correlational study can make use of publicly available data. This case discusses a correlational study using publicly available adverse drug reaction data to investigate the effect of media coverage on subsequent side effect reports in New Zealand. We identified three main stages in the process of conducting correlational research with publicly available data. This type of study has advantages, such as being able to investigate naturally occurring situations where an experiment would be unethical. However, there are also a number of challenges to keep in mind, such as being unable to prove causal relationships between variables and the difficulties of using data you have not collected yourself.

Learning Outcomes

By the end of this case, students should be able to

- Understand some of the limitations of experimental research and in what circumstances correlational studies can be of value
- Understand the types of research where publicly available data can be utilized
- Understand the process of conducting a correlational study using publicly available data
- Discuss the advantages and disadvantages of correlational research and the challenges of using publicly available data

Project Overview and Context

Our area of research focuses on the psychological phenomenon known as the nocebo effect. The nocebo effect is defined as the adverse response from a placebo tablet or active medication that cannot be explained by the known pharmacology of the drug (Barsky et al., 2002). Instead, the nocebo effect occurs when an individual expects side effects, pays more attention to their own internal sensations, and misattributes common benign symptoms to the medication (Petrie & Rief, 2019). These negative expectations can come from a range of sources, such as verbal information from the doctor or medication leaflets, and can cause some people to believe they are more sensitive to drug side effects (Horne et al., 2013). Currently, our research focuses specifically on how social observation can transmit side effects and, in other words, how observing another person experience medication side effects can influence an individual's expectations and subsequent side effect reports through the nocebo effect.

Our recent study, which is the topic of this case, investigated the effect of media coverage on side effect

reporting following a medicine change in New Zealand. In 2018, two of New Zealand's largest print media outlets released a series of newspaper and online articles detailing patients' complaints following a brand change to a generic antidepressant. The articles, published in February and April, highlighted some patients' views that the new generic was less effective than the original medication and that they had experienced increased side effects, such as nausea, headaches, and suicidal thoughts. There were approximately 45,000 patients taking this particular antidepressant (venlafaxine) in New Zealand; however, the media reports focused on only three patient accounts.

When the media coverage of this medicine switch occurred, we hypothesized that it would be associated with an increase in adverse drug reaction reports. The gold standard of research methodologies in psychology is experimental studies, and these have been conducted to identify the underlying mechanisms of how the nocebo effect occurs (Faasse & Petrie, 2016). However, for our particular situation and hypothesis, a correlational observational design was more appropriate.

Experimental studies of the nocebo effect, often conducted in clinical laboratory settings, involve informing participants of the intended effects of a medication they are about to take as well as the side effects (when actually study participants receive a placebo). To investigate the effect of social observation, participants are instructed to wait for the medication to take effect while ostensibly watching a fellow participant who has taken the same medication. This person is actually a study confederate and when asked how they are feeling by the experimenter they either complain of side effects or report no problems, depending on the participant's group allocation.

Studies that have utilized this paradigm have shown that female participants who see another person report side effects are more likely to report side effects themselves, and that side effect modeling also reduces the beneficial placebo effect of the tablet irrespective of gender (Faasse et al., 2015).

The advantage of experiments is that they can establish cause and effect in a relationship; for example, seeing another person report side effects causes the observer to also report side effects. However, there are a number of limitations to this methodology. For example, experiments often take place in a laboratory with the careful manipulation and control of variables, which can limit the generalizability of the findings to a real-world context. Also, it can be impractical or unethical to conduct an experiment in a naturally occurring situation. In our case, it is impossible to randomly allocate members of the general population to either be exposed to negative media coverage or not. To overcome the limitations of experimental research, we conducted an observational correlational study to investigate the effect of media coverage on side effect reporting.

Correlational research assesses the relationship between variables, without manipulating them as you would in an experiment (Wood & Brink, 1998). Our study builds on the small number of correlational studies that have also investigated media coverage and the subsequent change in the reporting of side effects (e.g., Faasse et al., 2012; Matthews et al., 2016; Nielsen & Nordestgaard, 2016). In these correlational studies, the researchers examine the association between the rate of adverse drug reaction reports and the timing of the media coverage. Although we might hypothesize that there is a causal relationship between these variables,

this cannot be explicitly determined in correlational studies. However, a sudden change in the reporting rate directly after the media indicates the likely presence of a nocebo response in a naturalistic setting.

Furthermore, it is often not possible for researchers to collect their own data—the study may be reactive rather than proactive, data collection would be too time-consuming and costly, or they do not have access to the population of interest. In such cases, publicly available data can be used. Publicly available data refer to any data collected by governmental agencies or research institutions that are freely available for the public to use. For example, the study by Matthews and colleagues that investigated the effect of media coverage of statins used data from the Clinical Practice Research Datalink (www.cprd.com)—a database that collects anonymized patient data from a network of primary health care services across the United Kingdom—and found that negative news stories about statins were associated with more patients stopping their statin tablets. Sometimes two sets of publicly available data can be accessed to further show the downstream health effects of the nocebo response. For example, Nielsen and Nordestgaard (2016) used data from the Danish Registry of Medicinal Products and Danish Patient Registry to show that, after negative statin stories in the Danish media, more patients discontinued taking their statin tablets and that discontinuation was associated with an increased risk of a heart attack and a risk of death from cardiac causes.

The increase in use of the Internet and search engines means that there is now more opportunity to use publicly available or search aggregation data to examine research questions. For example, researchers can use Google Trends to track how specific Internet searches change over time or during a specific period of interest (Petrie et al., 2019) or even what information people want to know from Wikipedia after an infectious disease outbreak (Tausczik et al., 2012). It is likely with advances in artificial intelligence tools that potential for this type of research will increase over time.

The following sections discuss our recent correlational study that examined the impact of news stories about a nationwide medication switch to a generic antidepressant using publicly available adverse drug reaction data in New Zealand. We go through the design of this study and how it worked in practice, including challenges and the lessons learned.

Section Summary

- Although experimental studies can identify cause and effect, the highly controlled environments of laboratories can limit the generalizability of the findings. To overcome this limitation, correlational observational studies can examine “real-world” situations.
- Correlational studies examine the relationship between variables; however, it is often not possible for researchers to collect their own data, as it may be too time-consuming, expensive or they do not have access to the population of interest. In such situations, publicly available data can be used instead.
- Publicly available data can come from a range of sources such as national health databases or Google Trends.

Research Design

To investigate the effect of the media coverage on side effect complaints, we conducted a correlational study using publicly available data from the New Zealand Centre for Adverse Reactions Monitoring (CARM). The study is correlational as we investigated the relationship between the release of the media articles and changes in side effect reporting. We examined the rates of side effect reporting before and after the media coverage as well as comparing the reporting rate of the side effects mentioned in the articles to those side effects that were not mentioned.

Unlike typical correlational studies where researchers first identify participants and then determine recruitment strategies, utilizing pre-existing publicly available data means these steps are redundant. The data we used for this study were patient adverse drug reaction reports collected by CARM. CARM has an online service where health care providers and patients can report a suspected adverse reaction to a medicine or vaccine (in fact the vast majority of reports come from doctors). The data from CARM are primarily used by Medsafe (New Zealand's medicine safety agency) to monitor potential medication problems. Overall, we had planned three main stages in the process of conducting this correlational study with publicly available data.

Stage 1: Retrieving the Data

Accessing publicly available data can range in difficulty. Often, it can be quite simple with some online databases allowing the public to enter a search query, which brings up relevant information or data files that can be downloaded. We intended to use a database called "Suspected Medicine Adverse Reaction Search" (SMARS), which is maintained by Medsafe. This database compiles all the adverse reaction reports received by CARM into one online repository. The purpose of this database is to increase the availability of official information to the New Zealand public. Selecting a medicine and a time period in the database website produces a list of all the individual adverse reaction reports. We wanted all adverse reaction reports for venlafaxine from October 2017 to May 2018—a period that extended from before until after the media coverage. This timeframe was chosen to allow a sufficient number of months before the onset of the media coverage to determine the consistent pre-media rate of adverse reaction reporting.

In other instances, gaining access to data can be a challenge and researchers may have to rely on making official information data requests. For example, in New Zealand, the Official Information Act allows people to request access to any information held by the government or government agencies such as the department of health. Both the United States and United Kingdom have a similar process called the Freedom of Information Act. It should be noted that this pertains only to information held by the government and that in certain circumstances data cannot be released or do not exist.

Stage 2: Recoding the Data

We anticipated that some degree of data recoding would be required before analysis could begin. Potentially, there would be a small mismatch between the side effects used in the newspaper articles and the actual terms

used in the CARM reports. In such circumstances, we would match the newspaper symptom with similar side effect terms that best represented them in the CARM data.

Following any side effect recoding, we would then go through the CARM data and total the number of times each of these side effects was mentioned per month. The five side effects mentioned in the media articles were suicidal thoughts, nausea, fatigue, headaches, and anxiety. To more clearly examine the changes in the reporting rate of the media-mentioned side effects, we identified five other symptoms that had been reported at similar frequencies but were not mentioned in the media, and we used these as control side effects.

Stage 3: Analyzing the Data

The strength of a correlation between two variables is usually measured statistically using a correlation coefficient, the most common being Pearson *r* correlation. This study, however, had the added detail of investigating the change in adverse reaction reporting over a specific time period and so the data were most suited to a time series analysis. A *time series analysis* is a statistical method that analyzes a sequence of data points or observations measured at regular intervals over a period of time (Madsen, 2007). This analysis is typically used to identify trends or predict future data.

Our analysis also needed to take into account the onset of the media reports, which acted as an interruption to the normal trend of the data. Therefore, we conducted an *interrupted time series analysis* in the statistical program SAS. This analysis is essentially a regression that measures the strength of a dependent variable (e.g., number of side effect reports) relative to other changing variables (presence/absence of media). In the analysis, the number of side effects was regressed against a binary independent variable that indicated the presence of the media articles by month. March and May were given the value 1 as these months directly followed the media reports, whereas all other months were coded 0. The analysis produces an estimated interruption effect, which is the change in the rate of side effect reporting from the months coded 0 and 1, and indicates whether this is a statistically significant difference. It should be noted that this is a conservative analysis as the data from months coded 1 are averaged together to create an overall interruption effect rather than a separate effect for each month.

Section Summary

- This case conducted a correlational study using publicly available data from CARM to investigate the association between media coverage and changes in the rate of side effect complaints.
- We anticipated three main stages in the process of using publicly available data: retrieving the data from CARM, recoding the raw data to match the specific side effects mentioned in the media, and analyzing the data using an interrupted time series analysis.

Methods in Action

Overall, this correlational study using publicly available data was effective in investigating our research

hypothesis—that the media coverage would be associated with greater side effect reporting following a medicine brand change. We found that the rate of adverse reaction reporting significantly increased in the months directly after the release of the media articles compared with the average rate before the coverage. Furthermore, it was largely the side effects that were mentioned in the media that experienced an increase in reporting, whereas reports of the control side effects did not change from the pre-media rate.

However, there were some deviations to our original research plan. We had not realized that the SMARS database takes a considerable amount of time to be updated, so the data file we downloaded did not contain the most recent 3 months of adverse reaction reports. This was an issue, as our study required the most recent data. To address this, we requested directly from Medsafe a summary of all adverse reaction reports for venlafaxine from October 2017 to July 2018. We were sent an Excel data file, which included the month the report was received, the patient's age, gender, the severity of the reaction, whether the patient had recovered, and up to five symptoms attributed to the medicine. The data were still relatively straightforward to retrieve; however, there was still a small delay in receiving the previous month's data, because it took a few weeks for Medsafe to receive the latest CARM reports. The data were also somewhat limited: Instead of a list of all potential side effects reported by patients, each report only contained a maximum of five symptoms.

As anticipated, there was some discrepancy between the five side effects mentioned in the media articles and the side effect categories provided by CARM. Three of the side effects, headaches, nausea, and anxiety were matched exactly to the corresponding symptoms in the CARM data. However, the other two side effects were considered broad enough to cover a range of CARM adverse reactions. Therefore, CARM reports of fatigue, lethargy, and tiredness were recoded as the broader symptom of fatigue in our analysis. Suicidal thoughts comprised suicidal tendencies, suicidal ideation, and impulses to self-harm.

The analysis of the data was also more complex than expected. CARM only recorded the month the adverse reaction report was submitted, whereas more fine-grained data such as weekly or daily reports would have been more useful for our analysis. As such, we could not investigate the exact time point that the adverse reaction rate was expected to change with the onset of the media coverage.

Section Summary

- Occasionally, the predetermined research plan does not run as smoothly as expected. The study did answer our research hypothesis; however, there were some challenges that were not expected.
- Issues can include the publicly open database being infrequently updated so alternative strategies are needed to retrieve the relevant data.
- Some degree of data recoding is usually required when you are using another institution's database.

Practical Lessons Learned

Our study has highlighted the advantages as well as some more difficult aspects of conducting correlational studies with publicly available data. An important consideration when interpreting the results from such a

study is the correlational nature of the research. As it is not an experiment, where participants are randomly allocated to various groups, claims of cause and effect are more difficult to make (Grimes & Schulz, 2002). However, this does not necessarily take away from the validity of correlational research. It can be helpful to also consider alternative explanations of the findings and whether these are actually plausible (Thompson et al., 2005). For example, in our study, after considering alternatives, it seems very likely that the media coverage accounts for the sudden change in adverse reaction reports. It seems highly unlikely that the increase in side effects was caused by the change in medication as patients had switched to the new medication earlier, without an obvious increase in side effect reporting. Furthermore, the generic drugs had been tested to be equivalent and the side effects that had not been mentioned in the media did not increase to the same degree as those mentioned in the media stories.

Using publicly available data for research purposes has a number of advantages. In some situations, data collection from patients involved in a nationwide drug switch is not possible, as it can be a long, difficult, and expensive process. Utilizing publicly available data allows researchers to react quickly to sudden, unexpected situations. In this study, we could not preempt the media coverage with our own usual baseline data collection methods, such as questionnaires or patient interviews. However, a database of medicine adverse reaction reports allowed us to immediately investigate the real-time monthly impact of the media coverage and compare this to the pre-media baseline reporting rate.

Publicly available data are useful when researchers do not have the means to sample a population themselves. That being said, it is important to be aware of the sampling strategy used by the organization collecting the data. For example, in a governmental health survey, the sample will be randomly selected from a wider population (e.g., a city or a country) to ensure that it is representative and the results can be generalizable to the overall population. However, in some cases, publicly available data can be generated from a more biased sample such as one where individuals volunteer themselves to participate. It has been estimated that voluntary reports to national adverse reaction databases represent less than 10% of all potential adverse reactions (Smith et al., 1996). In the case of the CARM data, these are composed of voluntary reports, meaning that some patients who experienced an adverse reaction are likely not to have reported it. This factor needs to be kept in mind when forming conclusions from publicly available data.

A difficulty with publicly available data is having to work with the information you receive. In our situation, CARM only provided the month the adverse reaction report had been submitted, whereas a finer grain such as by week or day would have been more useful for our analysis. A lack of precision can be difficult to work with, especially when you want to examine the exact onset of an event and the subsequent effect on an outcome. In addition, when the data are not your own, you cannot influence which variables are measured or how they have been coded. For example, when patients complete an adverse reaction report on the CARM website, they are simply asked “Tell us what happened” and given a text box to describe their adverse experience. The side effects mentioned in the patient’s description are coded into specific symptom categories, and the output researchers receive only lists a maximum of five symptoms for each report. This can result in the potential loss of more detailed information about a patient’s reports; for example, we have no

way of knowing how many side effects people actually reported beyond the cutoff of five, or what symptom terms patients specifically used.

Section Summary

- Correlational studies are an effective methodology to investigate research questions. When interpreting the results from a correlational study, it is important to remember that cause and effect cannot be explicitly determined. However, by considering the plausibility of alternative explanations for the results, you may be more confident that a causal relationship between variables is present.
- A strength of using publicly available data is that researchers can react quickly to unexpected situations, but it is important to consider the sampling method used to collect the data.
- A disadvantage of publicly available data is that researchers have no influence over what variables are measured or how they have been coded.

Conclusion

Correlational studies are often viewed as inferior to experimental research; however, the conclusions drawn from these studies are still valid and in some cases this is the only type of research that can be conducted. Occasionally, the usual methodologies and timeline of correlational research can be inflexible when unexpected situations occur. In such cases, researchers should consider the benefits of publicly available datasets. Publicly available data are often underutilized in research but is nevertheless a valuable source of information. There are a number of stages in conducting correlational studies with publicly available data, from retrieving the data through to analysis strategies and interpretation of the results. Although this specific case and the research examples used focus on the use of publicly available data in a health context, this should not limit discussion of its use in other areas of research.

Classroom Discussion Questions

Classroom Discussion Questions

1. What other sources of publicly available data could researchers use?
2. What other research questions could be answered by conducting a correlational study with publicly available data?
3. What might be potential ethical issues to consider when conducting a correlational study and also when using publicly available data?

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

Barsky, A. J., **Saintfort, R.**, **Rogers, M. P.**, & **Borus, J. F.** (2002). Nonspecific medication side effects and the nocebo phenomenon. *Journal of the American Medical Association*, 287, 622–627. https://doi.org/10.1001/jama.287.5.622

Faasse, K., **Gamble, G.**, **Cundy, T.**, & **Petrie, K. J.** (2012). Impact of television coverage on the number and type of symptoms reported during a health scare: A retrospective pre–post observational study. *BMJ Open*, 2(4), e001607. https://doi.org/10.1136/bmjopen-2012-001607

Faasse, K., **Grey, A.**, **Jordan, R.**, **Garland, S.**, & **Petrie, K. J.** (2015). Seeing is believing: Impact of social modeling on placebo and nocebo responding. *Health Psychology*, 34, 880. https://doi.org/10.1037/hea0000199

Faasse, K., & **Petrie, K. J.** (2016). From me to you: The effect of social modelling on treatment outcomes. *Current Directions in Psychological Science*, 25, 438–443. https://doi.org/10.1177/0963721416657316

Grimes, D. A., & **Schulz, K. F.** (2002). Descriptive studies: What they can and cannot do. *The Lancet*, 359, 145–149. https://doi.org/10.1016/S0140-6736(02)07373-7

Horne, R., **Faasse, K.**, **Cooper, V.**, **Diefenbach, M. A**, **Leventhal, H.**, **Leventhal E.**, & **Petrie, K. J.** (2013). Personal sensitivity to medicines scale. *British Journal of Health Psychology*, 18, 18–30. https://doi.org/10.1111/j.2044-8287.2012.02071.x

Madsen, H. (2007). *Time series analysis*. Chapman and Hall/CRC.

Matthews, A., **Herrett, E.**, **Gasparrini, A.**, **Van Staa, T.**, **Goldacre, B.**, **Smeeth, L.**, & **Bhaskaran, K.** (2016). Impact of statin related media coverage on use of statins: Interrupted time series analysis with UK primary care data. *British Medical Journal*, 353, i3283. https://doi.org/10.1136/bmj.i3283

Nielsen, S.F., & **Nordestgaard, B. G.** (2016). Negative statin-related news stories decrease statin persistence and increase myocardial infarction and cardiovascular mortality: A nationwide prospective cohort study. *European Heart Journal*, 37, 908–916. https://doi.org/10.1093/eurheartj/ehv641

Petrie, K. J., **MacKrill, K.**, **Silvester, C.**, **Dalbeth, N.**, & **Pennebaker, J. W.** (2019). OK Computer? A time analysis of Google searches about symptoms. *Clinical Psychology in Europe*, 1(2), 32774. https://doi.org/10.32872/cpe.v1i2.32774

Petrie, K. J., & **Rief, W.** (2019). Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. *Annual Review of Psychology*, 70, 599–625. https://doi.org/10.1146/annurev-psych-010418-102907

Smith, C. C., **Bennett, P. M.**, **Pearce, H. M.**, **Harrison, P. I.**, **Reynolds, D. J. M.**, **Aronson, J. K.**, &

Grahame-Smith, D. G. (1996). Adverse drug reactions in a hospital general medical unit meriting notification to the Committee on Safety of Medicines. *British Journal of Clinical Pharmacology*, 42(4), 423–429. https://doi.org/10.1111/j.1365-2125.1996.tb00004.x

Tausczik, Y., **Faasse, K.**, **Pennebaker, J. W.**, & **Petrie, K. J.** (2012). Changes in anxiety and information seeking following the H1N1 outbreak: An analysis of web blogs and Wikipedia searches. *Journal of Health Communication*, 27, 179–185. https://doi.org/10.1080/10410236.2011.571759

Thompson, B., **Diamond, K. E.**, **McWilliam, R.**, **Snyder, P.**, & **Snyder, S. W.** (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children*, 71, 181–194. https://doi.org/10.1177/001440290507100204

Wood, M. J., & **Brink, P. J.** (1998). Correlational designs. In **M. J. Wood** and **P. J. Brink** (Eds.), *Advanced design in nursing research* (2nd ed., pp. 160–177). SAGE.