"Unveiling the Truth: How Cancer Survivors Walk, Move, and Overcome Falls"

Secondary Analysis Research

Abstract

In secondary data analysis (SDA) studies, investigators use data col- lected by other researchers to address different questions. Like prima- ry data researchers, SDA investigators must be knowledgeable about their research area to identify datasets that are a good fit for an SDA. Several sources of datasets may be useful for SDA, and examples of some of these will be discussed. Advanced practice providers must be aware of possible advantages, such as economic savings, the ability to examine clinically significant research questions in large datasets that may have been collected over time (longitudinal data), generating new hypotheses or clarifying research questions, and avoiding over- burdening sensitive populations or investigating sensitive areas. When reading an SDA report, the reader should be able to determine that the authors identified the limitation or disadvantages of their research. For example, a primary dataset cannot "fit" an SDA researcher's study exactly, SDAs are inherently limited by the inability to definitively ex- amine causality given their retrospective nature, and data may be too old to address current issues.

Secondary analysis of data collected by another researcher for a different purpose, or SDA, is increasing in the medical and social sciences. This is not surprising, given the immense body of health care–related research performed worldwide and the potential beneficial clinical implications of the timely expansion of primary research (Johnston, 2014; Tripathy, 2013). Oncology advanced practitioners should understand why and how SDA studies are done, their potential advantages and disadvantages, as well as the importance of reading primary and secondary analysis research reports with the same discriminatory, evaluative eye for possible applicability to their practice setting.

To perform a primary research study, an investigator identifies a problem or question in a particular population that is amenable to the study, designs a research project to address that question, decides on a quantitative or qualitative methodology, determines an adequate sample size and recruits representative subjects, and systematically collects and analyzes data to address specific research questions. On the other hand, an SDA addresses new questions from that dataset previously gathered for a different primary study (Castle, 2003). This might sound "easier," but investigators who

carry out SDA research must have a broad knowledge base and be up to date regarding the state of the science in their area of interest to identify important research questions, find appropriate datasets, and apply the same research principles as primary researchers.

Most SDAs use quantitative data, but some qualitative studies lend themselves to SDA. The researcher must have access to source data, as opposed to secondary source data (e.g., a medical record review). Original qualitative data sources could be videotaped or audiotaped interviews or transcripts, or other notes from a qualita- tive study (Rew, Koniak-Griffin, Lewis, Miles, & O'Sullivan, 2000). Another possible source for qualitative analysis is open-ended survey questions that reflect greater meaning than forced-response items.

SECONDARY ANALYSIS PROCESS

An SDA researcher starts with a research question or hypothesis, then identifies an appropriate dataset or sets to address it; alternatively, they are familiar with a dataset and peruse it to identify other questions that might be answered by the available data (Cheng & Phillips, 2014). In reality, SDA researchers probably move back and forth between these approaches. For example, an investigator who starts with a research question but does not find a dataset with all needed variables usually must modify the research question(s) based on the best available data.

Secondary data analysis researchers access primary data via formal (public or institutional archived primary research datasets) or informal data sharing sources (pooled datasets separately collected by two or more researchers, or other independent researchers in carrying out secondary analysis; Heaton, 2008). Therearenumeroussourc- es of datasets for secondary analysis. For example, a graduate student might opt to perform a secondary analysis of an advisor's research. University and government online sites may also be useful, such as the NYU Libraries Data Sources (https:// guides.nyu.edu/c.php?g=276966&p=1848686) or the National Cancer Institute, which has many subcategories of datasets (https://www.cancer. gov/research/resources/search?from=0&toolTyp es=datasets_databases). The Google search engine

is useful, and researchers can enter the search term "Archive sources of datasets (add key words related to oncology)."

In one secondary analysis method, researchers reuse their own data—either a single dataset or combined respective datasets to investigate new or additional questions for a new SDA.

Example of a Secondary Data Analysis

An example highlighting this method of reusing one's own data is Winters-Stone and colleagues' SDA of data from four previous primary studies they performed at one institution, published in the *Journal of Clinical Oncology* (JCO) in 2017. Their pooled sample was 512 breast cancer survivors (age 63 ± 6 years) who had been diagnosed and treated for nonmetastatic breast cancer 5.8 years (± 4.1 years) earlier. The investigators di- vided the cohort, which had no diagnosed neu- rologic conditions, into two groups: women who reported symptoms consistent with lower-extremity chemotherapy-induced peripheral neuropathy (CIPN; numbness, tingling, or discomfort in feet) vs. CIPN-negative women who did not have symptoms. The objectives of the study were to define patient-reported prevalence of CIPN symptoms in women who had received chemotherapy, compare objective and subjective measures of CIPN in these cancer survivors, and examine the relationship between CIPN symptom severity and outcomes. Objective and subjective measures were used to compare groups for manifestations influenced by CIPN (physical function, disability, and falls). Actual chemotherapy regimens administered had not been documented (a study limitation, but regimens likely included a taxane that is neurotoxic); therefore, investigators could only confirm that symptoms began during chemotherapy and how severely patients rated symptoms.

Up to 10 years after completing chemotherapy, 47% of women who had received chemotherapy were still having significant and potentially life-threatening sensory symptoms consistent with CIPN, did worse on physical function tests, reported poorer functioning, had greater disability, and had nearly twice the rate of falls compared with CIPN-negative women (Winters-Stone et al., 2017). Furthermore, symptom severity was re-

lated to worse outcomes, while worsening cancer was not.

Stout (2017) recognized the importance of this secondary analysis in an accompanying editorial published in JCO, remarking that it was the first study that included both patient-reported subjective measures and objective measures of a clinically significant problem. Winter-Stone and others (2017) recognized that by analyzing what essentially became a large sample, they were able to achieve a more comprehensive understanding of the significance and impact of CIPN, and thus to challenge the notion that while CIPN may improve over time, it remains a major cancer survivorship issue. Thus, oncology advanced practitioners must systematically address CIPN at baseline and over time in vulnerable patients, and collaborate with others to implement potentially helpful interventions such as physical and occupational therapy (Silver & Gilchrist, 2011). Other primary or secondary research projects might focus on the usefulness of such interventions.

ADVANTAGES OF SECONDARY DATA ANALYSIS

The advantages of doing SDA research that are cited most often are the economic savings—in time, money, and labor—and the convenience of using existing data rather than collecting primary data, which is usually the most time-consuming and expensive aspect of research (Johnston, 2014; Rew et al., 2000; Tripathy, 2013). If there is a cost to access datasets, it is usually small (compared to performing the data collection oneself), and detailed information about data collection and statistician support may also be available (Cheng & Phillips, 2014). Secondary data analysis may help a new investigator increase his/her clinical research expertise and avoid data collection challenges (e.g., recruiting study participants, obtaining large-enough sample sizes to yield convincing results, avoiding study dropout, and completing data collection within a reasonable time). Secondary data analyses may also allow for examining more variables than would be feasible in smaller studies, surveys of more diverse samples, and the ability to rethink data and use more advanced statistical techniques in analysis (Rew et al., 2000).

Secondary Data Analysis to Answer Additional Research Questions

Another advantage is that an SDA of a large dataset, possibly combining data from more than one study or by using longitudinal data, can address high-impact, clinically important research questions that might be prohibitively expensive or time-consuming for primary study, and potentially generate new hypotheses (Smith et al., 2011; Tripathy, 2013). Schadendorf and others (2015) did one such SDA: a pooled analysis of 12 phase II and phase III studies of ipilimumab (Yervoy) for patients with metastatic melanoma. The study goal was to more accurately estimate the long-term survival benefit of ipilimumab every 3 weeks for greater than or equal to 4 doses in 1,861 patients with advanced melanoma, two thirds of whom had been previously treated and one third who were treatment naive. Almost 89% of patients had received ipilimumab at 3 mg/kg (n = 965), 10 mg/kg (n = 706), or other doses, and about 54% had been followed for longer than 5 years. Across all studies, overall survival curves plateaued between 2 and 3 years, suggesting a durable survival benefit for some patients.

Irrespective of prior therapy, ipilimumab dose, or treatment regimen, median overall survival was 13.5 months in treatment naive patients and 10.7 months in previously treated patients (Schadendorf et al., 2015). In addition, survival curves consistently plateaued at approximately year 3 and continued for up to 10 years (longest follow-up). This suggested that most of the 20% to 26% of patients who reached the plateau had a low risk of death from melanoma thereafter. The authors viewed these results as "encouraging," given the historic median overall survival in patients with advanced melanoma of 8 to 10 months and 5-year survival of approximately 10%. They identified limitations of their SDA (discussed later in this article). Three-year survival was numerically (but not statistically significantly) greater for the patients who received ipilimumab at 10 mg/kg than at 3 mg/kg doses, which had been noted in one of the included studies.

The importance of this secondary analysis was clearly relevant to prescribers of anticancer therapies, and led to a subsequent phase III trial in the same population to answer the ipilimumab dose

question. Ascierto and colleagues' (2017) study confirmed ipilimumab at 10 mg/kg led to a significantly longer overall survival than at 3 mg/kg (15.7 months vs. 11.5 months) in a subgroup of patients not previously treated with a BRAF inhibitor or immune checkpoint inhibitor. However, this was attained at the cost of greater treatment-related adverse events and more frequent discontinuation secondary to severe ipilimumab-related adverse events. Both would be critical points for advanced practitioners to discuss with patients and to consider in relationship to the particular patient's ability to tolerate a given regimen.

Secondary Data Analysis to Avoid Study Repetition and Over-Research

Secondary data analysis research also avoids study repetition and over-research of sensitive topics or populations (Tripathy, 2013). For example, people treated for cancer in the United Kingdom are surveyed annually through the National Cancer Patient Experience Survey (NCPES), and questions regarding sexual orientation were first included in the 2013 NCPES. Hulbert-Williams and colleagues (2017) did a more rigorous SDA of this survey to gain an understanding of how lesbian, gay, or bisexual (LGB) patients' experiences with cancer differed from heterosexual patients.

Sixty-four percent of those surveyed respond- ed (n = 68,737) to the question regarding their "best description of sexual orientation." 89.3% indicated "heterosexual/straight," 425 (0.6%) indicated "lesbian or gay," and 143 (0.2%) indicated "bisexual." One insight gained from the study was that although the true population proportion of LGB was not known, the small number of self-identified LGB patients most likely did not reflect actual numbers and may have occurred because of ongoing unwillingness to disclose sexual orientation, along with the older mean age of the sample. Other cancer patients who selected "prefer not to answer" (3%), "other" (0.9%), or left the question blank (6%), were not included in the SDA to correctly avoid bias in assuming these responses were related to sexual orientation.

Bisexual respondents were significantly more likely to report that nurses or other health-care professionals informed them about their diagnosis, but that it was subsequently difficult to contact nurse specialists and get understandable answers from them; they were dissatisfied with their interaction with hospital nurses and the care and help provided by both health and social care services after leaving the hospital. Bisexual and lesbian/gay respondents wanted to be involved in treatment decision-making, but therapy choices were not discussed with them, and they were all less satisfied than heterosexuals with the information given to them at diagnosis and during treatment and aftercare—an important clinical implication for oncology advanced practitioners.

Hulbert-Williams and colleagues (2017) proposed that while health-care communication and information resources are not explicitly homophobic, we may perpetuate heterosexuality as "normal" by conversational cues and reliance on heterosexual imagery that implies a context exclusionary of LGB individuals. Sexual orientation equality is about matching care to individual needs for all patients regardless of sexual orientation rather than treating everyone the same way, which does not seem to have happened according to the surveyed respondents' perceptions. In addition, although LGB respondents replied they did not have or chose to exclude significant others from their cancer experience, there was no survey question that clarified their primary relationship status. This is not a unique strategy for persons with cancer, as LGB individuals may do this to protect family and friends from the negative consequences of homophobia.

Hulbert-Williams and others (2017) identified that this dataset might be useful to identify care needs for patients who identify as LGBT or LGBTQ (queer or questioning; no universally used acronym) and be used to obtain more targeted information from subsequent surveys. There is a relatively small body of data for advanced practitioners and other providers that aid in the assessment and care (including supportive, palliative, and survivorship care) of LGBT individuals—a minority group with many subpopulations that may have unique needs. One such effort is the white paper action plan that came out of the first summit on cancer in the LGBT communities. In 2014, participants from the United States, the United Kingdom, and Canada met to identify LGBT communities' concerns and needs for cancer research,

clinical cancer care, health-care policy, and advocacy for cancer survivorship and LGBT health equity (Burkhalter et al., 2016).

More specifically, Healthy People 2020 now includes two objectives regarding LGBT issues: (1) to increase the number of population-based data systems used to monitor Healthy People 2020 objectives, including a standardized set of questions that identify lesbian, gay, bisexual, and transgender populations; and (2) to increase the number of states and territories that include questions that identify sexual orientation and gender identity on state-level surveys or data systems (Office of Disease Prevention and Health Promotion, 2019). We should help each patient to designate significant others' (family or friends) degree of involvement in care, while recognizing that LGB patients may exclude their significant others if this process involves disclosing sexual orientation, as this may lead to continued social isolation of cancer patients. This SDA by Hulbert-Williams and colleagues (2017) produced findings in a relatively unexplored area of the overall care experiences of LGB patients.

DISADVANTAGES OF SECONDARY DATA ANALYSIS

Many drawbacks of SDA research center around the fact that a primary investigator collected data reflecting his/her unique perspectives and questions, which may not fit an SDA researcher's questions (Rew et al., 2000). Secondary data analysis researchers have no control over a desired study population, variables of interest, and study design, and probably did not have a role in collecting the primary data (Castle, 2003; Johnston, 2014; Smith et al., 2011).

Furthermore, the primary data may not include particular demographic information (e.g., respondent zip codes, race, ethnicity, and specific ages) that were deleted to protect respondent confidentiality, or some other different variables that might be important in the SDA may not have been examined at all (Cheng & Phillips, 2014; Johnston, 2014). Although primary data collection takes longer than SDA data collection, identifying and procuring suitable SDA data, analyzing the overall quality of the data, determining any limitations inherent in the original study,

and determining whether there is an appropriate fit between the purpose of the original study and the purpose of the SDA can be very time consuming (Castle, 2003; Cheng & Phillips, 2014; Rew et al., 2000).

Secondary data analysis research may be limited to descriptive, exploratory, and correlational designs and nonparametric statistical tests. By their nature, SDA studies are observational and retrospective, and the investigator cannot examine causal relationships (by a randomized, controlled design). An SDA investigator is challenged to decide whether archival data can be shaped to match new research questions; this means the researcher must have an in-depth understanding of the dataset and know how to alter research questions to match available data and recoded variables.

For example, in their pooled analysis of ipilimumab for advanced melanoma, Schadendorf and colleagues (2015) recognized study limitations that might also be disadvantages of other SDAs. These included the fact that they could not make definitive conclusions about the relationship of survival to ipilimumab dose because the study was not randomized, had no control group, and could not account for key baseline prognostic factors. Other limitations were differences in patient populations in several studies included in the SDA, studies that had been done over 10 years ago (although no other new therapies had improved overall survival during that time), and the fact that treatments received after ipilimumab could have affected overall survival.

READING SECONDARY ANALYSIS RESEARCH

Primary and secondary data investigators apply the same research principles, which should be evident in research reports (Cheng & Phillips, 2014; Hulbert-Williams et al., 2017; Johnston, 2014; Rew et al., 2000; Smith et al., 2011; Tripathy, 2013).

- Did the investigator(s) make a logical and convincing case for the importance of their study?
- Is there a clear research question and/or study goals or objectives?
- Are there operational definitions for the variables of interest?

- Did the authors acknowledge the source of the original data and acquire ethical approv- al (as necessary)?
- Did the authors discuss the strengths and weaknesses of the dataset? For example, how old are the data? Is the dataset sufficiently large to have confidence in the results (adequately powered)?
- How well do the data seem to "fit" the SDA research question and design?
- Does the methods section allow you, the reader, to "see" how the study was done (e.g., how the sample was selected, the tools/instruments that were used, as well their validity and reliability to measure what was intended, the data collection process, and how the data was analyzed)?
- Do the findings, discussion, and conclusions—positive or negative—allow you to answer the "So what?" question, and does your evaluation match the investigator's conclusion?

Answering these questions allows the advanced practice provider reader to assess the possible value of a secondary analysis (similarly to a primary research) report and its applicability to practice, and to identify further issues or areas for scientific inquiry. ●

Disclosure
The author has no conflicts of interest to disclose.

References

Ascierto P. A., Del Vecchio, M., Robert, C., Mackiewicz A., Chiarion-Sileni V., Arance, A.,…Maio M. (2017). Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: A randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncology, 18*(5), 611–622. https://doi.org/10.1016/S1470-2045(17)30231-0

Burkhalter, J. E., Margolies, L., Sigurdsson, H. O., Walland, J., Radix, A., Rice, D.,…Maingi S. (2016). The National LGBT Cancer Action Plan: A white paper of the 2014 National Summit on Cancer in the LGBT Communities. *LGBT Health, 3*(1), 19–31. https://doi.org/10.1089/lgbt.2015.0118

Castle, J. E. (2003). Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing,* 35(5), 287–290. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14593941

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: Opportunities and implementation. *Shang- hai Archives of Psychiatry, 26*(6), 371–375. https://dx.doi.org/10.11919%2Fj.issn.1002-0829.214171

Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research, 33*(3), 33–45.

Hulbert-Williams, N. J., Plumpton, C. O., Flowers, P., McHugh, R., Neal, R. D., Semlyen, J., & Storey L. (2017). The cancer care experiences of gay, lesbian and bisexual patients: A secondary analysis of data from the UK Cancer Patient Experience Survey. *European Journal of Cancer Care, 26*(4). https://doi.org/10.1111/ecc.12670

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

Office of Disease Prevention and Health Promotion. (2019). Lesbian, gay, bisexual, and transgender health. Retrieved from https://www.healthypeople.gov/2020/topics-objectives/topic/lesbian-gay-bisexual-and-transgender- health

Rew, L., Koniak-Griffin, D., Lewis, M. A., Miles, M., & O'Sullivan, A. (2000). Secondary data analysis: New perspective for adolescent research. *Nursing Outlook, 48*(5), 223–239. https://doi.org/10.1067/mno.2000.104901

Schadendorf, D., Hodi, F. S., Robert, C., Weber, J. S., Margolin, K., Hamid, O.,…Wolchok, J. D. (2015). Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *Journal of Clinical Oncology, 33*(17), 1889–1894. https://doi.org/10.1200/JCO.2014.56.2736

Silver, J. K., & Gilchrist, L. S. (2011). Cancer rehabilitation with a focus on evidence-based outpatient physical and occupational therapy interventions. *American Journal of Physical Medicine & Rehabilitation, 90*(5 Suppl 1), S5–S15. https://doi.org/10.1097/PHM.0b013e31820be4ae

Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., & Steinman M. A. (2011). Conducting high-value secondary dataset analysis: An introductory guide and resources. *Journal of General Internal Medicine, 26*(8), 920–929. https://doi.org/10.1007/s11606-010-1621-5

Stout, N. L. (2017). Expanding the perspective on chemotherapy-induced peripheral neuropathy management. *Journal of Clinical Oncology, 35*(23), 2593–2594. https://doi.org/10.1200/JCO.2017.73.6207

Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges (letter). *Iranian Journal of Public Health, 42*(12), 1478–1479.

Winters-Stone, K. M., Horak, F., Jacobs, P. G., Trubowitz, P., Dieckmann, N. F., Stoyles, S., & Faithfull, S. (2017). Falls, functioning, and disability among women with persis- tent symptoms of chemotherapy-induced peripheral neuropathy. *Journal of Clinical Oncology, 35(23)*, 2604–2612. https://doi.org/10.1200/JCO.2016