SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

1: What is primary data and secondary data in Statistics and research methods? | eNotes

Start studying Chapter 14 - Secondary analysis and official statistics. Learn vocabulary, terms, and more with flashcards, games, and other study tools.

New focus, Additional in-depth analysis Applied and Methodological Table 1: Methodological features of the secondary analyses undertaken [8] Though different, these secondary analyses share some similarities. In every case, the materials corpus have been taken from the VERBATIM database with the collaboration of the sociologist who had collected many of the original research interviews. This shows how important it is to rely on the experience and knowledge of the expert of a sociological field. Secondary analysis such as we practice always results from the collaboration between both researchers and disciplines. During the inquiry which had been commissioned by the head of the Division, the sociologists were eager to understand how those researchers adapt their work to the important changes in the company loss of the monopoly of electricity production and distribution, adjustment to competition. But they never asked questions about stress-related problems or mental health at work. The initial inquiry had allowed us to collate a very important amount of qualitative data, which could not have been completely analysed. Choices always have to be made during the analysis of data, especially in an industrial context when the sponsor of the study is eager to get its results quickly. Thus there are constraints related to efficiency which compel researchers to focus on some aspects only and to neglect others. The principal interest of such a secondary analysis was obviously to be able to explore a thematical field which had not been addressed initially. In this case, the secondary analysis was conducted on one set of data by the sociologists who had originally collected the data and carried out the initial study. First that stress at work had not been a subject that had been asked about in any depth during the initial inquiry, and second that the data were already two years old and things were changing very quickly. Our conclusion was that there was a genuine problem with stress in the workplace and that we could help elucidate more clearly the different types of interpretation about it as they related to different functions of the company. However, to be able to say more about it, it would be necessary to undertake a new inquiry focusing on this subject. In this case, the secondary analysis had been useful in order to quickly form an opinion about the problem and to define more precisely the nature of a new inquiry. In France, night units 1 are cheaper than day units for customers who have made a special tariff agreement with EDF. For example, in this case, they turn on their dishwasher and laundry machine at night. The marketing division needed precise information about the way people deal with that tariff. Since the statistics could not answer their questions properly it was considered important to obtain qualitative data to understand better the needs of customers. In this case it was quite easy to carry out such research even on a significant amount of data, as it corresponds to a precise terminology, in French "heures creuses", that is the concept of night units for electricity tariffs. As we will see further on, the problem of the corpus definition is much more complex in other cases. We chose to keep all the extracted studies even when they did not precisely focus on tariff problems and to re-use all the interviews collected where people spoke about their experience of "heures creuses" and night-time domestic use of electricity. It was also extremely useful to define categories of customers in relation to their different patterns of behaviour. For example, their relationship to "heures creuses" depends on several criteria such as: For example, we had the opportunity to re-use a very important study composed of a lot of interviews about rural and poor populations, which could misrepresent the behaviour of the global French population. Thus we had to be very careful about the sample definition imposed by the database contents and to resist the temptation to give quantitative results. The secondary analysis could only give qualitative information to be confirmed by a quantitative survey if necessary. Fortunately, the number of studies and interviews involved in the secondary analysis was sufficient to ensure a reliability of qualitative results. Results of this research were to contribute to an EDF website, devoted to providing advice on the use of thermic systems. Compared to the previous study we described earlier that related to the price of electricity , regarding the corpus definition, this secondary analysis featured a distinct problem. This is the reason for the choice of a larger corpus that of electric comfort in which those topics are more likely to occur. This secondary analysis has been agreed upon

because of a large amount of archived interviews and studies, arising from diverse research situations and samples and relating to different domestic electric systems. First, the problems are linked to the initial objectives associated with these research limits; second, the analysis process; and third, the kind of software used. These had an overly technical orientation and were uneasy to examine as such, due to the secondary nature of this research and to the verbal status of the available data. Moreover, limits due to the secondary character of this analysis reduced the scope of the initial objectives in that the primary data were not designed to examine the issue of "deviated" practices. These practices depended on a diversity of aims, the samples, and the interview guides. As a result, the theme of practices does appear in a very different manner through primary studies and data. Limits concerning the verbal content of our data also reduced the scope of goals. Thus, a typology of practices was impossible to derive for the following reasons: Lastly, this re-analysis illustrates the methodological problems related to the study of any practice available in these data. People talk about their practices in different ways and sometimes practices are deduced from the general context. Thus we can imagine that "deviated" practices are not meaningful for people, but two other possibilities might be proposed: This way to carry out our analysis suggests a holistic analysis procedure, which aims at preserving the specific content of primary data: The data have been explored through two types of software devoted to text analysis. The first stage of the analysis related to the content trends of the whole corpus and has been carried out using the French Automatic Textual Data Analysis software Tropes. The latter helped to highlight the themes appearing in our data and their importance in terms of frequency. It also offered results related to the linguistic specificities of the data. Through this first analysis we have been able to make a general description of the whole corpus in a somewhat "quantitative" approach. Finally, there was a need for a more complete description of the use of thermic systems, and a decision was taken to study a unique interview using Atlas. This secondary analysis was an exploratory study focusing on many aspects linked to renovation, for example, what is the object of renovation? What are the problems related to it? How are decisions to renovate taken? Who are the different actors involved in renovation? In cases where surveys had not been purposely designed to study renovation, this topic had sometimes been raised spontaneously or had sometimes been a marginal topic undertaken as part of the primary study. The secondary analysis was chosen because of the diversity of research situations corpus, samples, objectives, guides of interviews, etc. As a result, the context in which the theme of renovation was raised varied and this might have lead to the possibility of a more multidimensional study of the situation regarding renovation. It is interesting not only because of the holistic method used, but also because of the active role of the analyst during the interpretation of results. The researcher manages to identify the biases due to the primary data, and also the results are not seen as representative of the situation surrounding renovation, but rather as informing some issues related to this topic. We also had to adopt a holistic kind of analysis centred on the thematic context in which the renovation theme appeared in the whole data set. Afterwards, using the same type of software we had to make a more in-depth analysis of the renovation theme, by viewing the data more precisely, as well as the variables that were linked to its different aspects. Indeed, the theme of renovation is referred to in a different manner depending on the research context e. It may be a spontaneous theme or one that is induced, or it may be dominant or marginal throughout interviews. Lastly, its meaning varies among actors tenants, landlords, caretakers, etc. The researcher is then obliged to interpret results on renovation as being non-representative. Also, this kind of renovation was significantly linked among other variables to people aged 40 and below, whose houses had been built after In spite of these results, obtained through the Alceste software, we did not take into account these findings, which were considered as consequences of gathering our primary corpus, that is the secondary data analysed mainly represented: Thus, we were able to interpret these results as the consequence of our corpus structure, regarding its contents and its variables. Additionally, another finding seemed to oppose the existing statistics on renovation. The complete renovation of housing does not appear through our findings, which focus mainly on renovation of thermic systems, or on the partial renovation of housing. In this investigation, we collaborated closely with the sociologist who helped us to select the primary studies and who had undertaken some of the original studies. Methodological Issues Reflexivity Through the studies described above, we can see that the researcher adopts a reflexive position during the secondary analysis, thereby

enhancing the quality of re-analysis. This position varies across the different stages of the analysis, depending on the object of the study and its theoretical and empirical properties. In these studies, the following situations were relevant: Concerning the "holistic" procedure, in some of the secondary analyses we undertook a contextualisation of the topic that was the focus of the secondary study taking into account all of the topics featured in our data. This procedure was sometimes made necessary by the goals and specificities of the research. Moreover, we suggested this way of secondary analysis as an additional means to elucidate links with primary data and their thematic specificities. Secondly, the use of the French software helped us to carry out this method of analysis. There are programmes, that originated in the French sociolinguistic tradition, that contribute to the analysis of huge amounts of data; relying on linguistics and statistics, and that permit a quasi-automatic textual data analysis without the analyst imposing his or her own units of analysis 3. We tried to use these softwares in a complementary way and to answer different empirical questions through the use of them. These secondary analyses are unique in France, where this methodological trend is absent from everyday practice. The approaches have certain peculiarities, advantages and limits. Let us examine some peculiarities and advantages first: The re-analyses were conducted in an industrial research context, and this is the reason why they all concern specific empirical problems. This is uncommon compared to many secondary analyses which usually re-analyse two studies or a smaller number of interviews 4. These archived studies have the advantage of being similar in their quality and in the kind of information they provide. In fact, the archived studies are standardised and homogeneous with regards to: Those properties permit their combined use more easily than in other cases of secondary analysis. The secondary analyst has the chance to collaborate with researchers who have produced the primary studies. He or she also has the possibility to consult with the experts of the different topics throughout the various stages of a secondary analysis. Thus, a secondary analyst will be less exposed to the "risk of misrepresentation or misappropriation" THORNE, , p. There are few within the department who undertake secondary analysesâ€”only four among the 30 researchers working in this department have conducted secondary analyses since the VERBATIM database was created. This shows a difficulty in the take-up of use of the available database. This is, however, a feature of many technological innovations. If this fact appears discouraging at a somewhat quantitative level, it seems significant at a qualitative oneâ€”that secondary analysts differ from the other researchers in this sociology department by their sensitivity to methodological issues, by their frequent use and knowledge of software analysing qualitative data, by their openness to English-speaking research practices, as well as by their various scientific fields linguistics, psychology, political studies. Most researchers find that archiving and secondary analysis raise some key problems, including: There are four different types of attitudes that appear to be shown by researchers: Some researchers find the use of software necessary in order to carry out a secondary analysis. Thus, this represents an obstacle, as some do not use them in their research, either due to lack of knowledge, or because they reject their usefulness.

2: Secondary Analysis and Official Statistics (Chapter 14) | readinglists@leicester

the analysis of official statistics and records; opportunity for content analysis and secondary analysis Content analysis The systematic classification and study of content of mass media which could include either latent or manifest content.

Experiments are best for topics where the researcher controls a situation and manipulates an independent variable. Survey research is best for topics where the researcher asks questions and learns about reported attitudes and behavior. Content analysis is for topics that involve the content of messages in cultural communication. Existing statistics research is best for topics that involve information collected by large bureaucratic organizations. Public or private organizations systematically gather many types of information. Such information is collected for policy decisions or as a public service. It is rarely collected for purposes directly related to a specific research question. Thus existing statistics research is appropriate when a researcher wants to test hypotheses involving variables that are also in official reports of social, economic and political conditions. These include descriptions of organizations or people in them. Often, such information is collected over long periods. For example, existing statistics can be used by researcher who wants to see whether unemployment and crime rates are associated in cities across a 20 year period. As part of the trends, say in development, researchers try to develop social indicators for measuring the well being of the people. A social indicator is any measure of wellbeing used in policy. There are many specific indicators that are operationalization of well-being. It is hoped that information about social well being could be combined with widely used indicators of economic performance e. The main sources of existing statistics are government or international agencies and private sources. An enormous volume and variety of information exists. Researchers who conduct existing statistics research spend many hours in libraries or on the internet. There are so many sources of existing statistics like: Secondary Survey Data Secondary analysis is a special case of existing statistics; it is reanalysis of previously collected survey or other data that was originally gathered by others. As opposed to primary research e. Secondary analysis is increasingly used by researchers. It is relatively inexpensive; it permits comparisons across groups, nations, or time; it facilitates replication; and permits asking about issues not thought by the original researchers. There are several questions the researcher interested in secondary research should ask: Are the secondary data appropriate for the research question? What theory and hypothesis can a researcher use with the data? Is the researcher already familiar with the substantive area? Does the researcher understand how the data were originally gathered and coded? Large-scale data collection is expensive and difficult. The cost and time required for major national surveys that uses rigorous techniques are prohibitive for most researchers. Fortunately, the organization, preservation, and dissemination of major survey data sets have improved. Today, there are archives of past surveys open to researchers e. Reliability and Validity Existing statistics and secondary data are not trouble free just because a government agency or other source gathered the original data. Researchers must be concerned with validity and reliability, as well as with some problems unique to this research technique. It occurs when someone gives a false impression of accuracy by quoting statistics in greater detail than warranted by how the statistics are collected and by overloading detail. For example, in order to impress an audience, a politician might say that every year , persons, instead of saying 3 million persons, are annually being added to the population of Pakistan. Official policies and procedures specify definitions for official statistics. For example, a researcher defines a work injury as including minor cuts, bruises, and sprains that occur on the job, but the official definition in government reports only includes injuries that require a visit to a physician or hospital. Many work injuries as defined by thee researcher will not be in the official statistics. Another example occurs when a researcher defines people unemployed if they would work if a good job was available, if they have to work part-time when they want full-time work, and if they have given up looking for work. The official definition, however, includes only those who are now actively seeking work full or part-time as unemployed. The official statistics exclude those who stopped looking, who work part-time out of necessity, or who do not look because they believe no work is available. Another validity problem arises when official statistics are a proxy for a construct in which the researcher is really interested. This is necessary because the researcher cannot collect

original data. For example, the researcher wants to know how many people have been robbed, so he or she uses police statistics on robbery arrests as a proxy. But the measure is not entirely valid because many robberies are not reported to the police, and reported robberies do not always result in an arrest. Another validity problem arises because the researcher lacks control over how information is collected. All information, even that in official government reports, is originally gathered by people in bureaucracies as part of their job. A researcher depends on them for collecting organizing, reporting, and publishing data accurately. Systematic errors in collecting the initial information e. Stability reliability problems develop when official definition or the method of collecting information changes over time. Official definitions of work injury, disability, unemployment, literacy, poverty, and the like change periodically. Even if the researcher learns of such changes, consistent measurement over time is impossible. Equivalence reliability can also be a problem. For example, studies of police department suggest that political pressures to increase arrests are closely related to the number of arrests. It could be seen when political pressures in one city may increase arrests e. Researchers often use official statistics for international comparisons but national governments collect data differently and the quality of data collection varies. Inferences from Non-Reactive Data: It is difficult to use unobtrusive measures to establish temporal order and eliminate alternative explanations. In content analysis, a researcher cannot generalize from the content to its effects on those who read the text, but can only use the correlation logic of survey research to show an association among variables. Unlike the case of survey research, a researcher does not ask respondents direct questions to measure variables, but relies on the information available in thee text.

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

3: Secondary Data - Meaning, its advantages and disadvantages

Secondary data analysis can be literally defined as "second-hand" analysis. It is the analysis of data or information that was either gathered by someone else (e.g., researchers, institutions, other NGOs, etc.) or for some other purpose than the one currently being considered, or often a combination of the two (Cnossen).

Quick revise Official statistics This mainly refers to data already collected by governments, e. Official statistics are seen as scientific because they are collected in a highly standardised way. For example, some data has to be registered by law. Government surveys such as the Census, the General Household Survey and the British Crime Survey are viewed as highly reliable and objective in their design and execution. Positivist sociologists suggest that official statistics are useful to sociologists because they are already available. Their use saves time, effort and money. They give a wide-ranging picture of social phenomenon. They have excellent comparative value in that they allow examination of trends over time. Interpretivists however, are very critical of statistics because they have been collected for non-sociological purposes. They may be socially constructed and therefore tell us more about the people who collect them than about the social phenomena in question. Mass media reports Sociologists have used newspaper and magazine articles, television programmes, advertisements and films as sources of secondary data. In particular, interpretivist sociologists have used newspaper reports to give insight into past events and social concerns. The sociology of moral panics is one such area. Some sociologists use content analysis to systematically analyse the content and meaning of media messages. Analysis of mass media is relatively cheap because such reports are already available. Samples of people are not necessary. Content analysis is repeatable if used in its simple category form. It is allegedly objective because it does not interfere with what is being researched. It results in quantitative data. However, its scientific status has been questioned by positivists. It can be very subjective and the same content may be interpreted in different ways by different sociologists. The method also implicitly assumes that the media have an impact on the audience of lasting significance. The evidence for this is mixed and certainly not proven. Content analysis rarely asks the audience what it thinks and assumes that the audience reacts in the same way. Media content may tell us more about the values and attitudes of journalists than those of society. Historical documents Historical documents such as government reports and white papers, historical treatises, diaries and even novels from a particular period may add qualitative insight into social problems. However, the reliability and validity of any historical document should be assessed by asking four key questions: How authentic is it? Does it have credibility? Might it involve exaggeration, deception and justification leading to bias? Is the document representative or typical? Documents such as diaries may only represent the views of an articulate minority and provide a selective picture of society. Personal documents These are documents such as diaries, letters, autobiographies and biographies. Sociologists may even examine photographs and the inscription on gravestones. Some sociologists ask people taking part in their research to keep a diary documenting activities and feelings. This is known as time-budgeting. However, all these methods may be too subjective. People may be more concerned with justifying their activities than objectively recounting their experiences.

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

4: Oxford University Press | Online Resource Centre | Multiple choice questions

Secondary analysis is the practice of using secondary data in research. As a research method, it saves both time and money and avoids unnecessary duplication of research effort. Secondary analysis is usually contrasted with primary analysis, which is the analysis of primary data independently collected by a researcher.

Data revision[edit] Even after they have been published, some official statistics may be revised. In order to understand the accuracy of economic data and the possible impact of data errors on macroeconomic decision-making, the Federal Reserve Bank of Philadelphia has published a dataset [17] that records both initial real-time data estimates, and subsequent data revisions, for a large number of macroeconomic series. Data Sources[edit] There are two sources of data for statistics. Primary, or "statistical" sources are data that are collected primarily for creating official statistics, and include statistical surveys and censuses. Secondary, or "non-statistical" sources, are data that have been primarily collected for some other purpose administrative data, private sector data etc. Statistical survey or sample survey[edit] A statistical survey or a sample survey is an investigation about the characteristics of a phenomenon by means of collecting data from a sample of the population and estimating their characteristics through the systematic use of statistical methodology. The main advantages are the direct control over data collection and the possibility to ask for data according to statistical definitions. Disadvantages include the high cost of data collection and the quality issues relating to non-response and survey errors. There are various survey methods that can be used such as direct interviewing, telephone, mail, online surveys. Census[edit] A census is a complete enumeration of a population or groups at a point in time with respect to well-defined characteristics population, production. Data are collected for a specific reference period. A census should be taken at regular intervals in order to have comparable information available, therefore, most statistical censuses are conducted every 5 or 10 years. Data are usually collected through questionnaires mailed to respondents, via the Internet, or completed by an enumerator visiting respondents, or contacting them by telephone. An advantage is that censuses provide better data than surveys for small geographic areas or sub-groups of the population. Census data can also provide a basis for sampling frames used in subsequent surveys. The major disadvantage of censuses is usually the high cost associated with planning and conducting them, and processing the resulting data. It contains information on a complete group of units. An advantage is the total coverage even if collecting and processing represent low cost. It allows producing more detailed statistics than using surveys. Different registers can be combined and linked together on the basis of defined keys personal identification codes, business identification codes, address codes etc. Moreover, individual administrative registers are usually of high quality and very detailed. A disadvantage is the possible under-coverage that can be the case if the incentive or the cultural tradition of registering events and changes are weak, if the classification principles of the register are not clearly defined or if the classifications do not correspond to the needs of statistical production to be derived from them. There are different types of registers: Using the existing administrative data for statistical production may be approved by the public because it can be seen as a cost efficient method; individuals and enterprises are less harassed by a response burden; data security is better as fewer people handle it and data have an electronic format. Even though different types of data collection exist, the best estimates are based on a combination of different sources providing the strengths and reducing the weakness of each individual source. Official Statistics presentation[edit] Official statistics can be presented in different ways. Analytical texts and tables are the most traditional ways. Graphs and charts summarize data highlighting information content visually. They can be extremely effective in expressing key results, or illustrating a presentation. Sometimes a picture is worth a thousand words. Graphs and charts usually have a heading describing the topic. There are different types of graphic but usually the data determine the type that is going to be used. To illustrate changes over time, a line graph would be recommended. This is usually used to display variables whose values represent a regular progression. Stacked bar chart showing the sectoral contribution to total business services growth, for members of UNECE. For categorical data, it is better to use a bar graph either vertical or horizontal. They are often used to represent percentages and rates and also to

compare countries, groups or illustrate changes over time. The same variable can be plotted against itself for two groups. An example of this is the age pyramid. Pie chart can be used to represent share of per cent. Pie charts highlight the topic well only when there are few segments. Stacked bar charts , whether vertical or horizontal, are used to compare compositions across categories. They can be used to compare percentage composition and are most effective for categories that add up to per cent, which make a full stacked bar chart. Their use is usually restricted to a small number of categories. Tables are a complement to related texts and support the analysis. They help to minimize numbers in the description and also eliminate the need to discuss small variables that are not essential. Tables rank data by order or other hierarchies to make the numbers easily understandable. They usually show the figures from the highest to the lowest. Another type of visual presentation of statistical information is thematic map. They can be used to illustrate differences or similarities between geographical areas, regions or countries. The most common statistical map that is used is called the choropleth map where different shades of a colour are used to highlight contrasts between regions; darker colour means a greater statistical value. This type of map is best used for ratio [21] data but for other data, proportional or graduated symbol maps, such as circles, are preferred. The size of the symbol increases in proportion to the value of the observed object. Release[edit] Official statistics are part of our everyday life. For most citizens, the media provide their only exposure to official statistics. Television is the primary news source for citizens in industrialized countries, even if radio and newspapers still play an important role in the dissemination of statistical information. On the other hand, newspapers and specialized economic and social magazines can provide more detailed coverage of statistical releases as the information on a specific theme can be quite extensive. Official statistics provides us with important information on the situation and the development trends in our society. Users can gather information making use of the services of the National Statistical Offices. The development of computing technologies and the Internet has enabled users - businesses, educational institutions and households among others- to have access to statistical information. The Internet has become an important tool for statistical producers to disseminate their data and information. People are able to access information online. The supply of information from statistical agencies has increased. Today the advanced agencies provide the information on their websites in an understandable way, often categorized for different groups of users. Several glossaries have been set up by different organizations or statistical offices to provide more information and definitions in the field of statistics and consequently official statistics. Quality criteria to be respected[edit] The quality criteria of a national statistical office are the following: There principles apply not only to the NSO but to all producers of official statistics. Therefore, not every figure reported by a public body should be considered as official statistics, but those produced and disseminated according to the principles. Adherence to these principles will enhance the credibility of the NSO and other official statistical producers and build public trust in the reliability of the information and results that are produced. Relevance[edit] Relevance is the first and most important principles to be respected for national statistical offices. When releasing information, data and official statistics should be relevant in order to fulfil the needs of users as well as both public and private sector decision makers. Impartiality[edit] Once the survey has been made, the NSO checks the quality of the results and then they have to be disseminated no matter what impact they can have on some users, whether good or bad. All should accept the results released by the NSO as authoritative. Users need to perceive the results as unbiased representation of relevant aspects of the society. Dissemination[edit] In order to maximize dissemination, statistics should be presented in a way that facilitates proper interpretation and meaningful comparisons. To reach the general public and non-expert users when disseminating, NSOs have to add explanatory comments to explain the significance of the results released and make analytical comments when necessary. There is a need to identify clearly what the preliminary, final and revised results are, in order to avoid confusion for users. All results of official statistics have to be publicly accessible. There are no results that should be characterized as official and for the exclusive use of the government. Moreover, they should be disseminated simultaneously. Independence[edit] Users can be consulted by NSOs but the decisions should be made by statistical bodies. Information and activities of producers of official statistics should be independent of political control. Moreover, NSOs have to be free of any political interference that could influence their work and thus, the

results. They should not make any political advice or policy-perspective comments on the results released at any time, even at press conferences or in interviews with the media. Transparency[edit] The need for transparency is essential for NSOs to gain the trust of the public. They have to expose to the public the methods they use to produce official statistics, and be accountable for all the decisions they take and the results they publish. Also, statistical producers should warn users of certain interpretations and false conclusions even if they try to be as precise as possible. Furthermore, the quality of the accurate and timely results must be assessed prior to release. But if errors in the results occur before or after the data revision , [22] they should be directly corrected and information should be disseminated to the users at the earliest possible time. Producers of official statistics have to set analytical systems in order to change or improve their activities and methods. Confidentiality[edit] All data collected by the national statistical office must protect the privacy of individual respondents, whether persons or businesses. But on the contrary, government units such as institutions cannot invoke statistical confidentiality. All respondents have to be informed about the purpose and legal basis of the survey and especially about the confidentiality measures. The statistical office should not release any information that could identify an individual or group without prior consent. After data collection, replies should go back directly to the statistical producer, without involving any intermediary. Data processing implies that filled-in paper and electronic form with full names should be destroyed. International standards[edit] The use of international standards at the national level aims to improve international comparability for national users and facilitate decision-making, especially when controversial. Moreover, the overall structure, including concepts and definitions, should follow internationally accepted standards, guidelines or good practices. Their aim is to guide countries in the dissemination of their economic and financial data to the public.

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

5: Secondary analysis of existing data: opportunities and implementation

Chapter 13 Secondary analysis and official statistics Authored by Tom Owens Bryman and Bell: Business Research Methods: 3e This preview has intentionally blurred sections. Sign up to view the full version.

Gemma Hendy , Posted on: We provide analysis and insight into the data Companies House collects. The team behind the statistics There are 15 of us within the wider team, and we all do a range of things. Throughout the team, there are data compilers and analysts who have a range of responsibilities â€" including coding, providing visualisations and statistical analysis, to name a few. The people on the team have a lot of different skills for this. Some of the team have been in the organisation quite a long time. Others are still new and learning their way around just like me! The average length of time on the team is 5. About our official statistics One of my main responsibilities is the official statistics produced by Companies House. We publish 2 different official statistics releases: Incorporated companies in the UK and Companies register activities. Our statistics include the analysis of companies on the register by the period of incorporation, companies removed from the register and other corporate bodies we deal with, such as limited partnerships and limited liability partnerships. The release includes collective statistics for the United Kingdom, plus separate statistics for England and Wales, Northern Ireland, and Scotland. One of the main figures commented on is the size of the register â€" the total number of companies and other corporate bodies in the UK. Our statistics are used by the government, the public, public bodies and businesses, for a variety of reasons. This includes monitoring policy changes, understanding the market and making location decisions for companies. Our official statistics are also used for secondary analysis, such as spotting trends and patterns. The results of the analysis can prompt the users to dig deeper into the data. User needs are important. We want to spend our time producing relevant statistics for our users. Any feedback we receive will be taken into consideration, and we may use this to make changes to our next publication, due in the summer of We want to understand what matters to you and the impact of our data on you, your work and the decisions you make based on our data. Access the consultation on Companies House statistical products. The survey will only take 10 to 15 minutes to complete. The consultation will close on 30 September

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

6: Statistics - www.enganchecubano.com

Setting a reading intention helps you organise your reading. You can filter on reading intentions from the list, as well as view them within your profile.. Read the guide Ã—.

To view a copy of this license, visit http: Summary The secondary analysis of existing data has become an increasingly popular method of enhancing the overall efficiency of the health research enterprise. But this effort depends on governments, funding agencies, and researchers making the data collected in primary research studies and in health-related registry systems available to qualified researchers who were not involved in the original research or in the creation and maintenance of the registry systems. The benefits of doing this are clear but the barriers are many, so the effort of increasing access to such material has been slow, particularly in low- and middleincome countries. This article introduces the rationale and concept of the secondary analysis of existing data, describes several sources of publicly available datasets, provides general guidelines for conducting secondary analyses of existing data, and discusses the advantages and disadvantages of analyzing existing data. Background A typical mental health research project begins with the development of a comprehensive research proposal and is hopefully followed by the successful acquisition of funding; the researcher then collects data, analyzes the results, and writes-up one or more research reports. Another less common, but no less important, research method is the analysis of existing data. The analysis of existing data is a cost-efficient way to make full use of data that are already collected to address potentially important new research questions or to provide a more nuanced assessment of the primary results from the original study. In this article we discuss the distinction between primary and secondary data, provide information about existing mental health-related data that are publically available for further analysis, list the steps of conducting analyzes of existing data, and discuss the pros and cons of analyzing existing data. Of course, there are cases where the distinction is less clear. One example would be the analysis of data by a researcher who has no connection with the data collection team to address a research question that overlaps with the hypotheses considered in the original study. Another example would be when a member of the original research team subsequently revisits the original hypothesis in an analysis that uses different statistical methods. These situations commonly occur in the analyses of large-scale population surveys where the research questions are generally broad e. Sources of existing data Existing data can be private or public. To maximize the output of data collection efforts, researchers often assess many more variables than those strictly needed to answer their original hypotheses. Often times, these data are not fully used or explored by the original research team due to restrictions in time, resources, or interest. However, if the research team is willing to share their data with other researchers who have the interest, skills, and resources to conduct additional analyses, this can greatly increase the productivity of the research team that conducted the original study. This type of exchange usually involves an agreement between the data collection team and the data analysis team to clarify details about data sharing protocols and how the data should be used. There are several publically available health-related electronic databases that can be used to address a variety of research topics. A few examples follow. However, these statistics are generally at the country-level so regional or population subgroup-specific data are not usually available. Another similar source is data available on the website of the Institute of Health Metrics and Evaluation at the University of Washington in the United States http: Located at the University of Michigan in the United States, ICPSR is a membership-based network that includes 65, datasets from over 8, discrete studies or surveys, including a number of largescale population surveys conducted in the United States and other countries. The website provides online analysis tools to generate simple descriptive statistics including frequencies and cross-tabulations. The website also provides technical support in data analysis and in the identification of potential data sources. In order to download data, users need to register with the system. For example, the United States Census Bureau http: Users interested in more information about publicly available health-related data can refer to Secondary data sources for public health: A practical guide by Boslaugh. Conducting a secondary analysis of existing data There are two general approaches for analyzing existing data: In the research question approach, researchers have an a priori hypothesis or a question in mind and then

look for suitable datasets to address the question. In the data-driven approach researchers glance through variables in a particular dataset and decide what kind of questions can be answered by the available data. In practice, the two approaches are often used jointly and iteratively. Researchers typically start with a general idea about the question or hypothesis and then look for available datasets which contain the variables needed to address the research questions of interest. If they do not find datasets that contain all variables needed, they usually modify the research question s or the analysis plan based on the best available data. When conducting either research question-driven or data-driven approaches to the analysis of existing data, researchers need to follow the same basic steps. In the research question-driven approach this is determined before the researchers look at the actual data available in the dataset; in the data-driven approach this is determined after the researchers look through the dataset. This involves obtaining detailed descriptions of the population under study, sampling scheme and strategy, time frame of data collection, assessment tools, response levels, and quality control measures. To the extent possible, researchers need to obtain and study in detail all survey instruments, codebooks, guidebooks and any other documentation provided for users of the databases. These documents should provide sufficient information to assess the internal and external validity of the data and allow researchers to determine whether or not there are enough cases in the dataset to generate meaningful estimates about the topic s of interest. This provides information about the use of the coding pattern for each variable and about the profile of missing data for each variable. Due attention should be paid to skip patterns, which can result in large numbers of missing values for certain variables. In comprehensive surveys that take a long time to complete, skipping a group of questions that are not relevant for a particular respondent i. For example, in a survey about alcohol-related problems, the survey module typically starts with questions about whether the interviewee has ever drunk alcohol. If the answer is negative, all questions about drinking behaviors and related problems are skipped because it is safe to assume that this interviewee does not have any such problems. Prior to conducting the full analysis, these types of missing values which indicate that a particular condition is not relevant for the respondent need to be distinguished from missing values for which the data is, in fact, missing which indicate that the status of the individual related to the variable is unknown. Researchers should be aware of these skips in order to make a strategic judgment about the coding of these variables. The recoded variables should be stored in a new dataset and all syntax for the recoding of variables and for the analysis itself should be documented. Thus, close examination of the survey questionnaires and codebooks are essential to ensure that each variable in the combined dataset has a uniform interpretation throughout the study. In this case, the data set usually includes design variables for each case including sampling weight, strata, and primary sampling unit that are needed to adjust the analysis of interest such as the prevalence of a condition, odds ratios, mean differences, etc. Researchers who conduct secondary analysis of existing data should consider the design variables used in the original study and apply these variables appropriately in their own analyses in order to generate less biased estimates. Pros and cons of the secondary analysis of existing data 4. Advantages The most obvious advantage of the secondary analysis of existing data is the low cost. There is sometimes a fee required to obtain access to such datasets, but this is almost always a tiny proportion of what it would cost to conduct an original study. Also, the data posted online are usually cleaned by professional staff members who often provide detailed documentation about the data collection and data cleaning process. Moreover, teams conducting large-scale population-based surveys that are made available to others usually employ statisticians to generate ready-to-use survey weights and design variables - something that most users of the data are unable to do - so this helps users make necessary adjustments to their estimates. This is a great boon to graduate students and others who have lots of good ideas but no money to conduct the studies that could test their ideas. Researchers who would rather spend their time testing hypotheses and thinking about different research approaches rather than collecting primary data can find a large amount of data online. The increasing availability of such data online encourages the creative use and cross-linking of information from different data sources. For example, experts in hierarchical models can combine data from individual surveys with aggregate data from different administrative levels of a community e. The availability of such databases also provides statisticians with real-life data to test new statistical models. Such analyses could identify potential new interventions to existing problems that can subsequently be tested

in prospective studies. Disadvantages Inherent to the nature of the secondary analysis of existing data, the available data are not collected to address the particular research question or to test the particular hypothesis. It is not uncommon that some important third variables were not available for the analysis. Similarly, the data may not be collected for all population subgroups of interest or for all geographic regions of interest. Another problem is that to protect the confidentiality of respondents, publicly available datasets usually delete identifying variables about respondents, variables that may be important in the intended analysis such as zip codes, the names of the primary sampling units, and the race, ethnicity, and specific age of respondents. This can create residual confounding when the omitted variables are crucial covariates to control for in the secondary analysis. Another major limitation of the analysis of existing data is that the researchers who are analyzing the data are not usually the same individuals as those involved in the data collection process. Therefore, they are probably unaware of study-specific nuances or glitches in the data collection process that may be important to the interpretation of specific variables in the dataset. Sometimes, the amount of documentation is daunting particularly for complex, large-scale surveys conducted by government agencies , so users may miss important details unless they are prominently presented in the documents. Succinct documentation of important information about the validity of the data by the provider and careful examination of all relevant documents by the user can mitigate this problem. Government support for secondary analysis of existing data This paper discusses several issues related to the secondary analysis of existing data. There are definitely limitations to such analyses, but the great advantage is that secondary analyses can dramatically increase the overall efficiency of the research effort and - a secondary advantage - give young researchers with good ideas but little access to research funds the opportunity to test their ideas. Recognizing the importance of making the most of high-quality research data and of rapidly translating research findings into actionable knowledge, starting in the United States National Institute of Health, the largest funding agency for biomedical research in the world, required all projects with annual direct costs of , US dollars or more to include data-sharing plans in their proposals. Moreover, NIH has released several program announcements specifically designed to promote secondary analysis of existing datasets. Other countries and some large health care providers also make registry data available to qualified researchers. These practices ensure that other researchers not involved in the studies or in the creation and maintenance of the registries will be able to use the data generated by these big projects or by the registries to test a wide range of hypotheses. Other governments including the Chinese government , health-related non-government organizations, and other funders of biomedical research need to follow these examples. Failure to provide qualified researchers access to government-generated registry data or to government-supported research data results in a huge but unnecessary wastage of economic and intellectual resources that could be better employed to improve the health of the nation. Biography Open in a separate window Dr. Hui Cheng is an epidemiologist by training. She has published findings from studies on mental health related topics using public data. Her main interest is substance use and related problems, and public mental health. Footnotes Conflict of interest: The authors declare no conflict of interest related to this article. Secondary data sources for public health: Design and analysis 2nd Ed. Modelling the sampling design in the analysis of health surveys. Stat Methods Med Res.

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

7: Our official statistics: have your say - Companies House

Official statistics. This mainly refers to data already collected by governments, e.g. statistics relating to births, marriages, deaths, health, crime, the economy, etc. Official statistics are seen as scientific because they are collected in a highly standardised way.

Subscriptions for the hardcopy version are free to researchers with addresses in the UK. Apply by email to sru soc. Her interest in secondary analysis of qualitative data developed through the intersection of these two roles. Secondary analysis involves the utilisation of existing data, collected for the purposes of a prior study, in order to pursue a research interest which is distinct from that of the original work; The approach has not been widely used in relation to qualitative data; Various methodological and ethical issues need to be considered and are more problematic if the secondary analyst was not part of the original research team; Further work to develop the approach is required in order to see if the potential benefits can actually be realised in practice. Although the secondary analysis of quantitative data is a common and generally accepted mode of inquiry, the same cannot be said of qualitative data Hinds, Vogel and Clarke-Steffen This Update outlines some of the forms that secondary analysis of qualitative data can take, the key methodological and ethical issues that arise, and how the approach might be further developed. What is secondary analysis? Secondary analysis involves the use of existing data, collected for the purposes of a prior study, in order to pursue a research interest which is distinct from that of the original work; this may be a new research question or an alternative perspective on the original question Hinds, Vogel and Clarke-Steffen , Szabo and Strang In this respect, secondary analysis differs from systematic reviews and meta-analyses of qualitative studies which aim instead to compile and assess the evidence relating to a common concern or area of practice Popay, Rogers and Williams As will be shown below, secondary analysis can involve the use of single or multiple qualitative data sets, as well as mixed qualitative and quantitative data sets. In addition, the approach may either be employed by researchers to re-use their own data or by independent analysts using previously established qualitative data sets. Despite the fact that thus far secondary analysis of qualitative data has not been widely undertaken, there have been a few reviews of the approach e. Hinds, Vogel and Clarke-Steffen , Thorne Classification of different types of secondary analysis of qualitative data is not straightforward as there are almost as many types as there are examples. It is made more difficult by the fact that some researchers may not define their work as secondary analysis Hinds, Vogel and Clarke-Steffen These difficulties notwithstanding, forms of secondary analysis are cross-classified in Table 1 according to the focus of the analysis and the nature of the original data used. Examples of work classified in this way will be described; some cells remain empty cells 1c, 2a and b, 3a and c because appropriate examples have not yet been identified and it is not known if these forms of secondary analysis have ever been conducted there are no a priori grounds for excluding them. Forms of secondary analysis Additional in-depth analysis: Again using data from two qualitative studies on which they had worked separately, Angst and Deatrick described how children with chronic illness and their parents are involved in health care decisions cell 1b. Similarly, Kirschbaum and Knafl combined data from two studies with which they had been involved, to explore the nature and quality of parent-professional relationships across two different illness situations cell 1b. For instance, in their secondary analysis of related quantitative and qualitative data sets about claimants of Invalid Care Allowance, McLaughlin and Ritchie concentrate on the ex-carers in the original sample in order to describe the socio-economic and psychological legacies of care giving among this group cell 2c. Why do secondary analysis? There is growing interest in re-using qualitative data, reflected in the establishment of Qualidata by the ESRC. This facilitates the archiving of data from qualitative studies Corti and Thompson , Hammersley , Corti et al More generally, limited opportunities for conducting primary research and the costs of qualitative work have prompted researchers to consider maximising use of the data available to them. The advent of software to aid the coding, retrieval and analysis of qualitative data is another development which is likely to facilitate both the archiving and availability of qualitative data for secondary analytic purposes. In these respects, the impetus behind the approach is similar to the one which informed the secondary analysis of quantitative data Procter Various arguments in favour of

developing secondary analysis of qualitative studies have been put forward Hinds, Vogel and Clarke-Steffen , Sandelowski , Szabo and Strang , Thorne For example, it has been contended that the approach can be used to generate new knowledge, new hypotheses, or support for existing theories; that it reduces the burden placed on respondents by negating the need to recruit further subjects; and that it allows wider use of data from rare or inaccessible respondents. In addition, it has been suggested that secondary analysis is a more convenient approach for particular researchers, notably students Szabo and Strang However, Thorne argues that where the researcher was not part of the original research team the approach is best only employed by experienced researchers because of the particular difficulties of doing secondary analysis in an independent capacity. It should also be noted that use of the approach does not necessarily preclude the possibility of collecting primary data. This may, for example, be required to obtain additional data or to pursue in a more controlled way the findings emerging from the initial analysis. There may also be a need to consult the primary researcher s assuming that they are available in order to investigate the circumstances of the original data generation and processing. Despite the interest in and arguments for developing secondary analysis of qualitative data, the approach has not been widely adopted to date. Furthermore, as was shown above, existing studies have mainly been conducted by researchers re-using their own data rather than by independent analysts using data collected by others. This raises questions about the desirability and feasibility of particular strategies for secondary analysis of qualitative data discussed below. Methodological and ethical considerations Before highlighting some of the key practical and ethical issues which have been discussed in the literature, there are two fundamental methodological issues to be considered. The first is whether secondary analysis of qualitative studies is tenable, given that it is often thought to involve an inter-subjective relationship between the researcher and the researched. In response, it may be argued that even where primary data is gathered via interviews or observation in qualitative studies, there may be more than one researcher involved. Hence within the research team the data still has to be contextualised and interpreted by those who were not present. The second issue concerns the problem of where primary analysis stops and secondary analysis starts. Qualitative research is an iterative process and grounded theory in particular requires that questions undergo a process of formulation and refinement over time Glaser For primary researchers re-using their own data it may be difficult to determine whether the research is part of the original enquiry or sufficiently new and distinct from it to qualify as secondary analysis. There is no easy solution to these problems except to say that greater awareness of secondary analysis might enable researchers to more appropriately recognise and define their work as such. Just as the above issues have received little attention in the literature to date, so the principles of, and guidelines for, the conduct of secondary analysis remain rather ill-defined Thorne However, a number of practical and ethical considerations have been highlighted Hinds, Vogel and Clarke-Steffen , Szabo and Strang , Thorne and four key issues are summarised below. Compatibility of the data with secondary analysis: Scope for additional in-depth analysis will vary depending on the nature of the data; for example, while tightly structured interviews tend to limit the range of responses, designs using semi-structured schedules may produce more rich and varied data. A check for the extent of missing data relevant to the secondary analysis but irrelevant to the original study may also be required; for example, where semi-structured interviews involved the discretionary use of probes. More generally, the quality of original data will also need to be assessed and Hinds, Vogel and Clarke-Steffen , appendix provide a set of criteria for this task. Position of the secondary analyst: This will influence the decision over whether to undertake secondary analysis and, if so, the procedures to be followed. Secondary analysts require access to the original data, including tapes and field notes, in order to re-examine the data with the new focus in mind. This is likely to be easier if they were part of the original research team. If not, then ideally they should also be able to consult with the primary researcher s in order to assess the quality of the original work and to contextualise the material rather than rely on field notes alone. Further consultation may also be helpful in terms of cross-checking the results of the secondary analysis. Finally, whether conducting secondary analysis in an independent capacity or not, some form of contractual agreement between the secondary analyst and the primary researcher s , data archive managers, and colleagues involved in the primary research but not in the secondary analysis may have to be negotiated. Reporting of original and secondary data analysis: Ideally this

should include an outline of the original study and data collection procedures, together with a description of the processes involved in categorising and summarising the data for the secondary analysis, as well as an account of how methodological and ethical considerations were addressed Thorne Where sensitive data is involved, informed consent cannot be presumed. Given that it is usually not feasible to seek additional consent, a professional judgement may have to be made about whether re-use of the data violates the contract made between subjects and the primary researchers Hinds, Vogel and Clarke-Steffen Growing interest in re-using data make it imperative that researchers in general now consider obtaining consent which covers the possibility of secondary analysis as well as the research in hand; this is consistent with professional guidelines on ethical practice British Sociological Association Developing the approach To see if the potential of secondary analysis can be realised in practice, developmental work still needs to be undertaken. First, there should be a more comprehensive review of the literature on secondary analysis and studies which have explicitly and perhaps implicitly used this approach. This could include examination of the methods used, as well as the quality, value and impact of this work. Thirdly, there should be greater consideration of the issues involved in the secondary analysis of single, multiple and mixed data sets. Finally, some more specific guidelines are needed for researchers about the ethical issues to be considered when undertaking qualitative work that may be re-used in the future. On this note, it is encouraging that Qualidata are currently working with the ESRC to produce guidelines on collecting and preparing data for archiving and on issues of confidentiality and copyright Corti and Thompson Conclusion Despite growing interest in the re-use of qualitative data, secondary analysis remains an under-developed and ill-defined approach. Various methodological and ethical considerations pose a challenge for the would-be secondary analyst, particularly those who were not part of the primary research team. Further work to develop this approach is required to see if the potential benefits can actually be realised in practice. Readings in Medical Sociology. Critical Issues in Qualitative Research Methods. Social Research Update is published by:

8: Secondary data - Wikipedia

Why has the secondary analysis of official statistics been seen as an "unobtrusive" method? a) It increases the risk of "reactive effects" from participants b) The researcher is removed from the social settings that they are investigating.

9: Social Research Update Secondary analysis of qualitative data

Secondary data analysis, on the other hand, is the use of data that was collected by someone else for some other www.enganchecubano.com this case, the researcher poses questions that are addressed through the analysis of a data set that they were not involved in collecting.

SECONDARY ANALYSIS AND OFFICIAL STATISTICS pdf

Sardaru BALHARI: The Worship of Narsing in Kangra. [Notes and Queries . p. 176 The Canadian challenge Basic Statistics and Pharmaceutical Statistical Applications, Second Edition (Biostatistics) The development of modern sociology, its nature and growth in the United States Southwestern Elko County. Republic of Palau Guilty Sister (The Stepsisters, No 6) Appendix 2 : Legalism in the book of Galatians Program operations manual system Wonderland Avenue Exploring chemical analysis 3rd edition KJV Pitt Minion Text Bible Burgundy pigskin, O53Y Alameda and Contra Costa Counties, Ca Atlas Uefa annual plan filetype Photoreceptors and Calcium (Advances in Experimental Medicine and Biology) Grade 12 momentum and impulse notes Green in the city THE FOUR STATIONS OF LIFE Managing the modern economy The lives of john lennon The technical analysis magazine Finding Eutaw and North G Reality ColdFusion Composition of Petroleum Mixtures (Total Petroleum Hydrocarbon Criteria Working Group Series) Vancouver on the Columbia Sun and its planets: Our star Establishing Research Corporation The assault on government ; The challenge to deliberative democracy Alan Brinkley Insulin resistance-the real culprit in diabetes Wild cards by simone elkeles Masterpieces of the Vatican. Challenging the political elite Creating a Business Plan Pocket Mentor) Some things never change Discrete-time linear systems theory and design with applications Descendants of Darkness, Volume 2 The American Psychiatric Publishing Textbook of Geriatric Psychiatry (American Psychiatric Press Textbook The dramatic works of John Crowne Uncertainty of everyday life, 1915-1945 Link in a ument