

# Key Concepts in Social Research

## Secondary Analysis

Contributors: Geoff Payne & Judy Payne  
Editors: Geoff Payne & Judy Payne  
Book Title: Key Concepts in Social Research  
Chapter Title: "Secondary Analysis"  
Pub. Date: 2004  
Access Date: October 17, 2013  
Publishing Company: SAGE Publications, Ltd  
City: London  
Print ISBN: 9780761965428  
Online ISBN: 9781849209397  
DOI: <http://dx.doi.org/10.4135/9781849209397>  
Print pages: 214-219

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

*Secondary analysis is the re-analysis of either qualitative or quantitative data already collected in a previous study, by a different researcher normally wishing to address a new research question.*

*Section Outline: Data-sets are never fully used. Re-use for new purposes. Access to published or raw data? Archive sources of data-sets. Secondary analysis for student dissertations. Data-sets: suitability and re-defining measurements for new concepts. Are data-sets available? Secondary analysis in qualitative methods: ethics; the person in the setting.*

Because of time and other career pressures, no study analyses every aspect of the data collected. This creates opportunities for other researchers to re-examine and re-use data from previous studies. Sometimes a data-set is re-analysed to develop the original topic, such as Payne's re-calculations from the main Nuffield Mobility Study table (2003). More typically, the second use of the data-set is for a different purpose than its original intention. Oral histories of Welsh miners, for example, have been re-used to explore theories of social capital, while employment data have been re-analysed to investigate the socio-economics of racism (Bloor 2002; Iganski et al. 2001: **Documentary Analysis; Official Statistics**).

The term for this is 'secondary analysis', meaning the use of data, collected in one project, in a second study. This goes beyond just quoting from a publication, reproducing tables essentially as they appear in the original document, or cross-referencing something as a source. To qualify as secondary analysis, the data must be used as if the second researcher had collected them, i.e. *evaluated in detail, re-processed*, and placed as evidence in *an argument that is different* from the first study (Dale et al. 1988). The great attractions of secondary analysis are that already collected data are so quickly to hand, with virtually no fieldwork cost. On the other hand, we are restricted to the quality of the original research: for instance, if there have been technical errors in data collection (**Bias**), we cannot normally correct for them retrospectively.

[p. 215 ↓ ] Data availability has been crucial in the changing status of secondary analysis. There is a great difference in accessibility between hand-recorded interviews still stored in dusty piles of old questionnaires, and a coded, cleaned and categorised

data-set downloaded over the internet. In the past, whole sets were not released, while new tabulations were not provided, or took so long and cost so much to extract that it was uneconomic. Researchers were largely dependent on adapting published tables, which could to some extent be manipulated, but fell far short of the full set of raw data (i.e. for each respondent, individual responses to each question) ready to be processed into whatever format needed. In the mid-1970s, a secondary analysis study of industrial convergence theory involved re-coding the occupation tables from five Censuses by hand and calculator, collapsing the more detailed classifications from 1921 into the smaller units used in 1971 (Payne 1977). This took the researcher two months to complete, before any real analysis could start.

Today, official sites e-mail data-sets at little or no cost. For instance, modern Census data are now available to use 'on every researcher's desktop'. This influences what is processed in what order, the structures in which data are held, and the technical media of their 'publication'. While not all data are released ('small area statistics' are adapted to prevent identification of anyone contained therein), geographical boundaries, age groupings, detailed social class categorisations, etc., can largely be chosen by the researcher.

Major studies now copy their data to a central archive on completion of their first analysis. Indeed, Britain's main funder, the Economic and Social Research Council (ESRC), makes this a condition of its research grants. Data-sets and coding documents describing their format are held at the UK Data Archive (<http://www.data-archive.ac.uk> and [darchive-userservices@essex.ac.uk](mailto:darchive-userservices@essex.ac.uk)), part of the national Economic and Social Data Service (ESDS) set up in 2003 to co-ordinate holdings of archived records. The Archive does not run a data-processing service, but does make datasets available, at a very low cost, to bona fide researchers. Other parts of ESDS hold data on **Longitudinal Studies**, government surveys (**Official Statistics**) qualitative and historical data ([qual@essex.ac.uk](mailto:qual@essex.ac.uk) and [hds@essex.ac.uk](mailto:hds@essex.ac.uk)) and international comparative surveys. Electronic catalogues list the variables and sample sizes, facilitating choices from among the original studies.

A wide range of statistical information is available for comparative purposes at the international level. Of particular relevance are those produced by EUROSTAT for the European Union; the World Health Organization; and the Organization for Economic

Co-operation and [p. 216 ↓ ] Development (OECD). All produce statistics relating to health and social conditions on an annual basis (for example, the OECD's *Annual Health Data*). In addition, ad hoc reports are produced frequently. All of these sources are listed at the organisations' websites on the internet (**Internet and Other Searches**).

These resources offer an excellent prospect for undergraduate dissertations. No student preparing a thesis could collect and process thousands of interviews. Indeed, the data have often been collected by highly competent interviewers, described, and made professionally acceptable through academic publication. Their quality can be demonstrated by the fact that well-established findings have already come from them. Speed, cheapness, quality and legitimacy are readily on offer, obtainable from ESDS through one's supervisor.

However, secondary analysis dissertations come with three 'health warnings'. A few departments retain dissertation assessment regulations that award marks for data collection per se, so not giving full credit for using archived sources: this should be checked in advance with one's supervisor. Second, the focus of a secondary analysis thesis shifts to the *ideas* being explored, and the data *analysis* (because there are no marks to gain from *data collection*). The re-analysis has to be well handled if good marks are to be obtained. Third, and here we return to a problem faced in all secondary analysis, the data really must be capable of supporting the new uses planned.

As with official statistics, each study has involved a series of operational definitions and practical procedures. When re-using data, we need to ensure that the samples are large enough and representative of the people or organisations that interest us. When studies talk about, say, children, households, inner cities, migrants or religiosity (topics selected from a recent journal), do these things mean what we mean by the same words? Formal records like coding instructions are often incomplete: the original team shared implicit knowledge. We also need studies that included all the factors we want to explore. This limits what we can re-analyse, and how confident we can be about our new findings.

To take one example in more detail, Iganski et al. (2001) wanted to know how the position of British minority ethnic groups had changed since the 1960s. There had been several individual sociological studies, but these used inconsistent definitions and did not cover the whole country. Nor were there repeated government surveys providing

all the answers. The biennial General Household Survey had good questions but for samples too small to give estimates without potentially large sample error (**Sampling: Estimates and Size**). The Census had a very large sample but [p. 217 ↓ ] asked about 'country of birth', so mis-allocating expatriate whites and British-born members of the ethnic minorities. The annual New Earnings Survey had questions about jobs, industries and incomes, but no ethnicity data. The quarterly Labour Force Survey (LFS) became the main source, but because of its sample size, surveys had to be merged, and only five ethnic groupings could be considered. The LFS income data were unreliable, so it was necessary to substitute employment categories, following the LFS's use of the Registrar General's 'socio-economic groups'.

This illustrates how even in quantitative research, secondary analysis encounters difficulties. In qualitative research, where secondary analysis is still relatively new, there is less agreement about what can be justified. The participants did not agree to the secondary analysis: is re-use ethical? In one sense, secondary analysis is an **Unobtrusive Method**, but of course the original research may have been highly intrusive. If the data depended on the unique interaction of the original researcher with people and events (**Qualitative Methods**), how can a subsequent analyst claim access to that original understanding (Hammersley 1997)? Again, the archived format matters: simple re-analysis of interview transcripts or recordings may be possible, but these probably require re-coding. Even more work is involved in re-processing field notes or original observations, but these seldom survive in easily intelligible form. These practical problems help to explain the still limited use of secondary analysis.