


SPECIAL ISSUE INTRODUCTION

Selection Bias and Social Science History

Kris Inwood^{1*}  and Hamish Maxwell-Stewart²

¹Economics and History, University of Guelph and ²History, University of Tasmania

*Email: kinwood@uoguelph.ca

Abstract

Historians and social scientists routinely, and inevitably, rely on sources that are unrepresentative of the past. The articles in this special issue of the journal illustrate the widespread prevalence of selection bias in historical sources, and the ways in which historians negotiate this challenge to reach useful conclusions from valuable, if imperfect sources.

Bias occurs when researchers select information for study that is unrepresentative of the distribution of characteristics in the wider population they seek to analyze. In Barbara Geddes's pithy summary of the problem: "[T]he cases you chose affect the answer you get" (Geddes 1990: 131). While those who research contemporary phenomena minimize this risk using randomized trials and other controls, this is rarely a viable option for historians. Because the records and other collections held in archives, libraries, and museums have already been created, the researcher must make assumptions about the extent to which the resulting data is representative of the particular aspect of the past they wish to study. Thus, for historians, there are two aspects to selection bias: first, the degree to which the information they utilize is representative of the wider record collections from which that material has been drawn; and second, the degree to which those collections reflect the historical realities upon which the researcher wishes to shed fresh light. Both forms of bias can be systematically introduced because of inadequate research design, or they may be instrumental—in that they might result from the particular preconceptions of the researcher.

While historians may go to lengths to mitigate the extent to which selection bias impacts upon the results of their work, it is difficult (if not impossible) to eliminate the problem. When it comes to the study of the past, selection is an ever-present and particularly slippery issue. As Peter Novik once quipped, getting history right can feel like "nailing jelly to the wall" (Novik 1988: 3). It is thus not surprising that the issue of selection bias has preoccupied the minds of historians for generations. Arguments about selection and the most appropriate ways to mitigate its affects have powerfully shaped approaches to the study of the past. At times they have also contributed to antagonisms within the discipline. While the adoption of social science methods might help to guard against some forms of selection bias, the critics of quantification are keen to point out that such methods can usher in new forms of

selection. The aim of this special edition of *Social Science History* is to explore some of the wider dimensions of selection bias and to canvas ways in which social science historians might seek to identify, test for, and address these. It is also hoped that the issue promotes a wider constructive conversation about the ever-present challenge of selection and the new directions our work might take in a digital age.

At its root, the problem of selectivity is compounded by the way in which the theoretical perspective of the researcher inevitably shapes the narrative they construct. In a memorable phrase the British philosopher Michael Oakeshott argued that “The past is a field in which we exercise our moral and political opinions, like whippets in a meadow on a Sunday afternoon” (Oakeshott 1962: 165). It is difficult for the researcher not to be instrumentally selective. Thus, as David Cannadine has shown, successive generations of economic historians have sought to make sense of the Industrial Revolution in ways that mirror the concerns and preoccupations of an evolving present (Cannadine 1984). Some have argued that it is important to distinguish between examples of cultural relativism, such as that described by Cannadine, and more pernicious forms of cultural bias (McCullagh 2000: 65). Many explanations for the comparative economic advantage achieved by Europeans, for example, have been implicitly or explicitly located in characteristics perceived to be peculiar to Western cultures and presumed lacking elsewhere (Goody 2006: 305). This might be described as selection bias on a continental scale—a phenomena manifest in a tendency to see the West as dynamic and other regions as passive (Said 1978). It is not difficult to think of similar examples in relation to the treatment of women, First Nations, and other disadvantaged or marginalized groups in some past histories.

As Barbara Geddes points out, the problem of selecting cases for study on the dominant variable is particularly widespread amongst non-quantitative-minded researchers (Geddes 1990: 131–32). As Lara Putman recently described, most historians do not “build in systematic checks against omitted variable bias.” Instead, as they shift through archival accounts, they construct narratives of causation. Thus, an encounter with a single record might be sufficient to transform an historian’s thinking (Putman 2016: 384). In this type of research environment participant bias might be expected to flourish.

The need to guard against such issues by attempting to show the extent to which particular phenomena are representative of the distribution of all occurrences is a cornerstone of quantitative approaches to history. Yet, those who object to quantification are wont to point out that there are many aspects of past human experience that cannot be empirically studied. To leave out these, they argue, is a biased act—a form of statistical filtering. As David Hackett Fischer expressed the problem, there is a danger that a factor’s perceived importance might rest on the extent to which it can be quantitatively measured. As a result, a fixation on the fungible might direct the researcher to follow the numbers into some causal research backwater (Fischer 1971: 90–94).

A related charge is that quantitative history is statist history, in that it tends to draw on official records such as census enumerator’s returns, registers of birth’s deaths and marriages, or military enlistment records at the expense of other forms of evidence. If it does this in ways that are innocent of power relations the researcher may be blind to the circumstances that shaped the official record in the first place

(Evans and Thorpe 1992). It is all too easy to forget the ways in which the data-collection process can impact on data quality and integrity. Analysis of the interviews conducted with former slaves by the Work's Progress Administration in the 1930s reveals that 76 percent of those interviewed by white researchers indicated that the food supplied on the plantation had been good, but that this dropped to 46 percent where the interviewer was black. As Norman Yetman observes, the etiquette of Southern race relations is likely to have introduced forms of self-censorship that could easily be missed by subsequent researchers (Yetman 1984: 188).

A great deal of historical data was recorded as the result of a conversation that occurred when a census enumerator knocked on a door, or a soldier fronted up before the recruiting sergeant, or a prisoner was measured and questioned in a watchhouse. While not always charged with race dynamics, these encounters are likely to have been shaped by other power dynamics. This is a particular concern for longitudinal researchers. The information recorded against the same individual in different record series often varies. Such differences can reveal much about the circumstances under which those data were collected. It is all too easy to forget that each of the many records that might make up a reconstructed life course consist of recorded responses to a question, subtle variations in which might elicit different responses. In some records, for example, prisoners might be asked how they are currently employed and in others to report the skills they had acquired (Maxwell-Stewart 2016). The desire on the part of information collectors to allocate data to discreet categories might lead to further distortions. Some historians argue, for example, that the notion of caste as currently understood in India was largely a creation of the nineteenth-century census (Samarpudra 2011).

Archives in this sense are never inert spaces. Record collection was always designed to suit a particular purpose and usually that purpose was quite different from the manner in which a historian might wish to put that data to use. As Ann Laura Stoler has argued, because of this it is important to read an archive along the grain before attempting to co-opt data and use it in ways not envisaged by the organizations and individuals who originally collected that material (Stoler 2008). It is precisely because historians cannot control the circumstances under which their data were collected, that they need to reconstruct the processes that shaped record collection to gauge the degree to which past practices led to omissions, imprecision, or other imperfections likely to impact on the quality of subsequent analysis. A check on associated archival correspondence, for example, can sometimes reveal the extent to which the official agencies had misgivings about the quality of data or the patchiness of its coverage (Van de Valle 1974: 18).

There is a danger that the ease of access that has accompanied widespread digitization will increase the risk of selection bias. The perils associated with mining information remotely without taking the care to place that data into its wider archival context has raised particular concern (Putman 2016). Yet, it is also the case that the availability of ever-larger numbers of digitized data sets present researchers with opportunities to explore selection processes across record series. Triangulation is a powerful and often underestimated research technique that can be employed as an effective check against selection bias and record omissions. It can also be used to analyze variations in the way information is recorded between record groups. By

identifying who is enumerated in different administrative data sets, and comparing differences in the way they are described, it is often possible to reconstruct much about the process of record formation.

We explore some of these questions in this special issue examining selection bias in topics relating to historical demography, economic history, and health. The first article in this collection by Eric Schneider illustrates the value of triangulating with school, military, and prison records from Japan, Australia, Britain, and the United States in his exploration of selection bias encountered in the identification of the historical growth pattern of children. A scarcity of evidence permitting the researcher to follow the same children over time has forced scholars to infer growth using the apparent change in average heights between different groups of children at consecutive ages. Any influence that changes the composition of these groups, or leads to the misreporting of true ages, has the potential to bias the measure of growth. The bias may or may not be sufficiently powerful to alter the qualitative assessment of particular hypotheses. A particularly awkward problem arises if the bias is known to be age-related because there is no easy way to separate age-related bias from age-related physical growth. Schneider proposes tests for selection bias that may be useful in particular circumstances, but he also acknowledges that there is no substitution for taking the “time and effort to study the selection mechanisms into a sample very carefully.” Only in this way is it possible to make a sound judgment about the presence of selection bias and the severity of its consequences for the problem at hand.

A particularly clear example of selection bias is provided by Ewout Depauw’s treatment of recidivism in prison records. Depauw reviews recent evidence by historians and analyzes the personal characteristics and patterns of repeat incarceration for 19,300 people held in Belgian prisons from 1838 to 1902. Depauw builds on a careful reading of previous studies to suggest ways in which incarceration figured in the life course of individuals captured in his data. He argues that prisoners were a heterogeneous lot and that repeat offenders differed in systematic ways—by sex, birthdate, social class (occupation), stature, and other characteristics. Depauw establishes beyond doubt that the people who ended up in prison were a selected subset of the Belgian population and that the nature of selection varied by type of offense and by the experience of repeated incarceration.

The study of Belgian prisoners makes clear that generalizing from their experience to the broader population is risky without some understanding of the processes by which people end up in prison, once or repeatedly, and for what kind of offences. A similar insight arises from a comparison by Inwood et al. of 8,000 Australian soldiers and 4,000 Australian prisoners born from 1870 to 1890. Studies of the long-run evolution of human health rely heavily on reports about soldiers and prisoners because in most societies they were the only people to receive standardized medical exams. This article shows that the soldiers and prisoners differed from each other and from the wider population in terms of age, birthplace, occupation, and life-course experience. Prisoners on balance came from more limited social circumstances and continued their physical growth later in adolescence, although interestingly there was little difference in adult stature between the two groups after controlling for observable characteristics.

Soldiers are the focus of three more articles in this collection. It is well known that a voluntary military enlistment is particularly likely to be unrepresentative because some people decide to enlist while others do not, and the reasons underlying this decision complicate any inference from the enlistment sample to the broader population. Data derived from a compulsory universal conscription are generally understood to escape this problem. Quanjer and Kok provide a sobering corrective using the records for more than 4,000 men born from 1850 to 1890 and captured by the Historical Sample of the Netherlands. These authors demonstrate that the cumulative impact of mortality and migration mean that the men who entered compulsory universal conscription were already unrepresentative of their cohort. Quanjer and Kok further report that some of those who did enlist still managed to escape the compulsory medical exam, and that this group differed systematically from those who were examined. Thus, conscription does not eliminate the concern for selectivity, although the authors argue that their source is nevertheless useable for research with caution.

The Fourie et al. article examine the records of 31,000 English-born soldiers from the same cohort who enlisted in England, Canada, Australia, and South Africa. The authors argue that differences between the scale and technology of the Anglo-Boer War 1899–1902 and World War I 1914–18 influenced the stature of the men who decided to enlist in each conflict. Thus, representativeness of a military sample depends on the forces shaping military demand for labor as well as more commonly considered influences of labor supply. For this example, at least, it is clear that data from the two conflicts should not be pooled uncritically.

The studies of Dutch and English soldiers show that the selectivity of military enlistment is even more complex than has been appreciated. A study of American soldiers by Ariel Zimran goes some way toward suggesting a possible solution. Zimran provides a usefully intuitive understanding of the selectivity of military service for 26,000 soldiers in the US Civil War 1860–65. The author draws from a conceptual tradition that is reasonably common in economics although its extension to historical sources requires considerable ingenuity. His method provides a way to identify if a particular sample is representative and, if it is not, the consequences of selection bias on both observable and unobservable characteristics for the testing of hypotheses about the broader population.

Many of the articles in this collection analyze large collections of records that might be characterized as “big historical data.” The largest samples are seen in the article by Antonie et al., who analyze six sets of linked Canadian census data ranging in size from 550,000 to 1,250,000 observations. Each observation describes a single person observed at the beginning and the end of a decade. The authors argue that the kinds of people whose records in different years can be matched successfully are unrepresentative of the population—even if the matching criteria are designed to minimize bias. The solution, unavoidably, is to reweight the linked records during any substantive analysis. The authors explore the limitations that may arise from a relatively small number of records in some of the cells being reweighted.

People are the individual units of observation in each of the mentioned articles. The final article in the collection, a sophisticated study of inequality in the Manhattan housing market from 1880 to 1910, takes the family rental unit as the basic unit of observation. Rowena Gray points out that inequality measured with

income typically implies a different profile than a consumption-based measure, and that evidence for the latter is available from an earlier date. Gray identifies reasons why, within the housing evidence, the least expensive and the most expensive housing tends to be understated. She further argues that this selection bias is a less pressing problem for some parts of the city than others, and that it changes little over time. These observations illustrate the importance of understanding the mechanisms for selection into the sample, as is recommended by Eric Schneider.

The eight articles collected in this special issue illustrate the widespread prevalence of selection bias in historical sources, and the ways in which analysts nonetheless manage to reach useful conclusions in spite of these challenges. If there is any single lesson from this collection, it is that the evidence of selection bias should not dissuade us from using valuable, if imperfect historical sources. Rather, we are challenged to take the time to understand the limitations of our sources and to exercise cautious ingenuity as we reach for robust conclusions. The challenge of working with incomplete and selected sources does not present an obstacle to historical research; rather it is the very stuff of historical research and should be welcomed as such.

References

- Cannadine, David** (1984) "The present and the past in the English Industrial Revolution 1880–1980." *Past and Present* 103: 131–72.
- Evans, Raymond, and Thorpe William** (1992) "Power, punishment and penal labour: Convict workers and Moreton Bay." *Australian Historical Studies* 25 (98): 90–111.
- Fischer, David Hackett** (1971) *Historians' Fallacies: Toward a Logic of Historical Thought*. New York: Routledge.
- Geddes, Barbara** (1990) "How the cases you choose affect the answers you get: Selection bias in comparative politics." *Political Analysis* 2: 131–50.
- Goody, Jack** (2006) *The Theft of History*. Cambridge: Cambridge University Press.
- Maxwell-Stewart, Hamish** (2016) "The state, convicts and longitudinal analysis." *Australian Historical Studies* 47 (3): 414–29.
- McCullagh, C. Behan** (2000) "Bias in historical description, interpretation, and explanation." *History and Theory* 39 (1): 36–66.
- Novik, Peter** (1988) *That Noble Dream: The "Objectivity Question" and the American Historical Profession*. Cambridge: Cambridge University Press.
- Oakeshott, Michael** (1962) *Rationalism in Politics and Other Essays*. New York: Methuen.
- Putman, Lara** (2016) "The transnational and the text-searchable: digitized sources and the shadows they cast." *American Historical Review* 121 (2): 377–402.
- Said, Edward** (1978) *Orientalism*. London: Pantheon.
- Samarpdra, Padmanabh** (2011) "Census in colonial India and the birth of class." *Economic and Political Weekly* 46 (33): 51–8.
- Stoler, Ann Laura** (2008) *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton, NJ: Princeton University Press.
- Van de Valle, Etienne** (1974) *The Female Population of France in the 19th Century: A Reconstruction of 82 Departments*. Princeton, NJ: Princeton University Press.
- Yetman, Norman R.** (1984) "Ex-slave interviews and the historiography of slavery." *American Quarterly* 36 (2): 181–210.