

# Contents

American automobiles and workers' dreams H. F. Moorhouse

Economic brokerage and peripheral underdevelopment in the Isle of Lewis Peter G. Mewett

The persistence of religion: conservative Protestantism in the United Kingdom Steve Bruce

Contradictions in the educational theory of Gintis and Bowles Mike Cole

Research based on administrative records Gatherme Hakim

Entrée and exit Frank Coffield and Carol Borrill

Extended review Shirley Dex

Book reviews

Books received

Notes on contributors

University of Keele

# Research based on administrative records

# Catherine Hakim\*

This review is concerned with a type of research confined mainly (but not exclusively) to researchers working in organisations which produce or maintain large quantities of information as a result of their activities, information which is not labelled as research data but can be used as a basis for research work. Administrative records are easily overlooked as a potential source of data for research, and they present special problems that are not covered in the existing methodological literature produced by academics. But the problems are not unsurmountable, and the result is often research of a type that could not be done by any other means at all.

There are cases of independent researchers (usually academics) doing research based on administrative records. <sup>1</sup> But the great majority of such studies are carried out by researchers working inside the relevant organisation (such as a government department, local authority, hospital or area health authority), and probably few end up in published, and hence accessible, reports. <sup>2</sup>

The literature on social research methods concentrates almost exclusively on new data collection, whether by survey or other approaches, and secondary analysis of existing data sources is rarely dealt with.<sup>3</sup> Although documentary evidence is frequently used in social research, the relevant methodological literature is concerned primarily with the 'reading' and interpretation of non-quantitative (and non-quantifiable) single-unique documents and records.<sup>4</sup> The only literature of which I am aware on the methodology of converting the information in records into quantitative data is that which is now being developed around the re-analysis of nineteenth-century census records deposited in public archives.<sup>5</sup> But many of the problems that must be confronted in the coding and analysis of data from *administrative* records do not occur at all, or in the same way, in the coding and analysis of census information—if only because the census

489

## Catherine Hakim

is a specially designed (and unique) data collection and not carried out as the by-product of administrative functions. Discussions of the secondary analysis potential and limitations of published statistics provide a very limited introduction to the analysis of data in administrative records, especially as certain records are little used for statistical purposes.\*

Administrative records are a collection of documents containing mainly factual information compiled in a variety of ways (from informants and respondents) and used by organisations (both in the public and private sector) to record the development and implementation of decisions and activities that are central to their functions. They are large in number; are usually high in content rigidity and narrow in content scope; and are often available for considerable stretches of time. Government records can offer complete censuses of the populations or universe in question. Administrative records are traditionally held on paper (in forms, files, card indexes and so forth), but are increasingly held and updated on computers. The main reference here is to examples of records held in the public sector.

## 'Types of administrative record

Given the enormous range of administrative records and of the types of information recorded and its uses, no single classification can hope to do justice to them all. I have used two criteria to develop the typology presented in Table 1, but other criteria might be used to differentiate further within each category. The two primary criteria would appear to be: (a) the type of information recorded and its primary, administrative, uses; and (b) the nature of the recording process. Most of the time, these two factors are very closely related, so that some combinations would be rare or non-existent.

A three-fold classification of the primary uses of administrative records is offered in Table 1. Statutory decision-making with reference to minima only refers to records which are kept in the context of implementing legislation or regulations which set out minimum entitlements or minimum requirements only. For example Wages Councils lay down minimum wages in certain trades, but do not set any maximum rates of pay. Health and Safety legislation lays down minimum standards for safety, but inspections (and the records) are concerned only with ensuring that the minima are observed, and not with recording full details of those cases where the minima are exceeded.

Statutory decision-making with universal applicability refers to records which are kept in the context of implementing legislation un Table 1 Classification of data available from administrative records

	Primary administrative uses of the records	Nature of recording process		
		Routine	Regular	Special
1	Statutory decision-making with reference to minima only og WI assessments. HSE inspections	A		P
141-	Statutory decision-making with universal applicability eg UB and SB claimants, income tax records	в	D	G
X	Record-keeping for service-delivery functions eg Jobcentre records electoral register Industrial Tribunal records and related ACAS case files	C	E	H
8	Data collections carried out purely for statistical and research purposes (and hence with no administrative uses) but relying to a large extent on the existing administrative rystem for the data collection process, eg Census of Employment, New Earnings Survey, Retail Price Index, strikes data, population census.		-1	

regulations which set out the entitlement of all applicants. For example legislation on income tax, on unemployment benefit and supplementary benefit determines the amounts to be collected from or paid to each person, and the records are a routine and necessary part of the task of ensuring that all people are treated according to the rules, consistently and uniformly.

Many records are created in the context of service-delivery activities where legislation does not lay down the precise, or minimum, entitlement or requirement. For example the Jobcentres run by the Manpower Services Commission do not have a statutory inhligation to counsel or advise every person who registers as looking for work; thus any records of the number of times a registrant was

epunselled would not be as comprehensive and routine as if there were a statutory obligation, they are more likely to be adequate for the service-delivery function, but no more

In the early days, administrative records rarely contained more than the bare minimum of essential and routinely collected data. More recently, additional information has been collected in certain records, either on a regular basis or as a special exercise (in which case they were often termed a 'survey'). Additional information might be required for management purposes, for monitoring service-delivery or the operation of statutory activities, or for statistical purposes. <sup>7</sup> This suggests the three-fold classification of the recording process in Table

Data recording is *noutine* when it is an expected and necessary part of the work of administrators, in some way central to their work. In this case, errors should be minimal or non-existent, coverage should be comprehensive, consistent and reliable. Any biases in the data recorded should be systematic and consistent because they would be part of the organisational and work context.

In some records, additional information is recorded on a regular basis. For example MSC Jobcentres have been required to record the ethnic origin of unemployed registrants for management information and statistical purposes, even though they do not need to use this information themselves in helping people to find jobs (type E data in Table 1); with the change to voluntary registration (in 1982) the equivalent information will be collected by unemployment benefit offices instead (type D data in Table 1), 8 The recorders' assiduity and care here will normally be reduced, with a higher incidence of errors and random variation in recordings. The errors and random variation will not be detected by the recording office (since they do not use the information in the course of their work) but may become apparent when analyses of the data are carried out. In some cases additional information is created rather than collected. For example ACAS clerks are required to create a code for the industry in which Industrial Tribunal applicants worked (using the Standard Industrial Classification and information on the employer's name and type of business). Analyses of the data from these records reveal marked variations in some years in the industrial distribution of IT applicants which is due to error and inconsistency in the classification of industry by clerks.

Administrative records may also contain additional information for ad hoc or special studies. These are sometimes referred to as 'surveys' or 'research exercises'. The quality of the additional information recorded is less predictable and may or may not be affected by systematic or random bias arising out of the fact that it will normally be collected and recorded in conjunction with the other routinely required information. Although one-off exercises can attract interest, and hence greater care in the recording, they are usually carried out at the same time as (and in addition to) existing work, so that there is often less time available for special care, norwithstanding the interest. One example is the 'special drive' carried out by the Wages Inspectorate in winter 1978/79 during which inspectors were required to record additional information on homeworkers' reasons for doing homework and length of experience in the trade, items of information that were not central to their main task of inspecting wages, and not routinely recorded (type F in Table 1).

Finally, Table 1 identifies one other category (type X) which strictly speaking does not belong there at all. This covers specially designed data-collections that are carried out solely for statistical purposes but rely to a large extent on the existing administrative system for the data collection process. In this case administrative staff are used (for example for the mail-out and mail-back parts of a survey) not administrative records. The task is typically completely unrelated to the day-to-day work of local office staff, and the resulting information is used by headquarters staff (nowadays statisticians in the Government Statistical Service). The Census of Employment and New Earnings Survey are carried out in this way. Arguably the censuof population falls into this category because the census offices rely in large part on the network of local Registrars' offices for the conduct of the census operation, even though enumerators are recruited specially for the job. Less well-known examples are the use of local office staff to collect information on prices (for the Retail Price Index) and on strikes. All of these specially designed data collections result in published reports, whereas many types of administrative records are never used to produce statistics. Type X data does not constitute an example of administrative records; it is identified in Table 1 primarily to make the distinction-one that is easily overlooked by academics who make use of official statistics.9

One implication of the classification offered in Table 1 is that as the range of information in records increases there will tend to be an associated decline in the quality of the information. Record-keeping becomes a very variable activity when it is for internal management purposes, especially if the object of the exercise is invisible to the recorders (and potentially misunderstood by them also), so that the quality of the data (in terms of completeness and consistency) will be

#### Catherine Hakim

lower. And 'special' data collection exercises carried out within the context of administrative work are almost certain to be of lower quality than routinely collected information, where quality control is implemented by virtue of the fact that the information will be used by the recorders themselves. Special data collections are carried out in this way because they are cheaper than the alternative of a specially designed research survey, but they do not compare in terms of data quality.<sup>10</sup>

The second implication is that the data in administrative records is very heterogeneous, in terms of coverage, content and quality, on both axes in the table. Since a single set of records might contain three or more types of data, an overall assessment of their quality would be ruled out and the nature of each data item would need to be separately established. In the case of type X data collections carried out solely for research and statistical purposes, published reports on the results, which comment on the quality of the information obtained, will always be available.<sup>11</sup> But for administrative records proper, such reports are non-existent or rare, even when data from the records ends up in published statistics. <sup>12</sup>

When trends over time are being studied, it will also be necessary to assess the degree of continuity in the treatment of any data item. A particular item of information may change its status, from 'special' to 'routine' for example, as the relevant legislation and regulations are changed. One example of this is the information on occupational pensioners among the registered unemployed. Until the late 1970s entitlement to flat-rate unemployment benefit was not dependent in any way on receipt of an occupational pension; so data on occupational pensioners among the registered unemployed was available in the records only when specially collected by Jobcentres as part of a 'survey' exercise on the characteristics of the unemployed. As a result of new legislation in the early 1980s, benefits were reduced for persons aged 60 and over who were in receipt of an occupational pension of over £35 a week, so in the 1980s the information had to be routinely collected by benefit offices in the process of assessing statutory benefit entitlement. The data item was thus transferred from type H to type B in Table 1, with major implications for data quality even though the information was consistently obtained from administrative records. Even if the data content remains the same in principle, a transfer of the data collecting/recording function from one organisation to another can in practice have a substantive impact on the data content.

The typology offered in Table 1 does not cover other criteria that

might be used to differentiate between types of data held in administrative records. For some types of record, such as those offering information on suicide or crime, problems of values and attitudes will often arise and may affect the recording process. Also the data items of most interest in these records are those that are necessarily the most problematic (irrespective of the source): information on the intentions and motivations of actors, on the causes of an observed event. These sorts of records are however untypical of the great majority of administrative records and are not considered here. They are unfortunately the example most commonly presented in university courses on official statistics, even though they are atypical. <sup>13</sup>

#### Some examples of administrative records

The three examples described here concern records which are compiled by or accessible to the Department of Employment, primarily because I have some knowledge of them. They cover most of the types of data classified in Table I and illustrate the range of data which might be used for research analyses.

#### Wages Inspectorate records

Although there is no statutory national minimum wage in Britain (as in the United States and most other countries), the Wages Council system sets legal minimum wages (or more precisely, rates of pay) in certain industries and trades. The Wages Inspectorate carries out a policing function to ensure that these minima are observed, to oversee the payment of arrears due to employees who have been underpaid, and to prosecute employers who persistently fail to pay the minimum rates. The Wages Inspectorate keeps records of inspections of the rates of pay and earnings in each of the establishments they visit; these can number anywhere from 30,000 to 50,000 establishments each year. The records are treated as confidential, unless they result in a prosecution for underpayment, in which case they are part of the evidence presented in court. All the records are kept indefinitely so long as the employer stays in business; records for businesses that have closed down are destroyed after a lapse of time. The Wages Inspectorate publishes an annual report on the results of each year's work. The statistical content is limited to a count of the number of establishments and employees inspected, the number where underpay ments were found, and the total amount of arrears of pay collected from employers and paid to their employees. The information routinely recorded by the Wages Inspectorate (type A in Table 1) in

#### Catherine Hakim

very limited, consisting essentially of: weekly and hourly earnings, hours worked per week, and the legal minimum rate of pay applicable to each worker. In the winter of 1978/79 a 'special drive' on homework was carried out in London and Birmingham, during which a special effort was made to inspect the pay records of all establishments in the clothing trade, where homeworking was concentrated. For this special exercise, a new form was employed which required Wages Inspectors to collect some additional information on homeworkers, for example 'reasons for homework' and 'experience in the trade' were to be recorded. This specially collected additional information (type F in Table 1) was no longer collected after the special drive period. The results of the special drive were reported in a Department of Employment Press Notice in early 1979, but this only gave figures on the number of illegal underpayments discovered (which was nil for homeworkers). Findings based on the additional specially collected information were not reported. The records of inspections carried out in the winter of 1978/79 in London were later used as the basis of a more intensive and detailed research analysis. The study made use both of the routinely collected data on carnings, hours and rates of pay and of the additional special information recorded about homeworkers, 14

#### Industrial Tribunal records

Industrial Tribunals are a more recent institution, created in the mid-1960s for the resolution of disputes about employment relationships between employers and employees. The scope and range of disputes which can be presented to the tribunals have been enlarged greatly since then, as legislation on employment relationships has been extended, for example by the Equal Pay Act and equal opportunines legislation. Tribunals are now responsible for hearing unfair dismissal applications and cases concerning disputes relating to sex or race discrimination, equal pay, redundancy, maternity rights, time off for union business, and statutory guarantee payments among others. In some years as many as 41,000 cases have been lodged.

Applications are made to the Central Offices of Industrial Tribunals in London and Glasgow where each is allotted a case number and a confidential file is opened which contains all correspondence and documents relating to the case and the eventual decision of the tribunal. Shortly afterwards, in most jurisdictions, applications are sent for information and conciliation action to regional offices of the Advisory Conciliation and Arbitration Service (ACAS) where separate confidential files are opened, inaccessible to tribunal staff, which contain papers generated by the conciliation process. These two sets of files, and particularly the last, provide the basis for regular statistical accounts of the tribunal and conciliation system and their results, based in some jurisdictions on counts of all cases and in others on a 10 per cent sample. Results are published, usually annually, in the *Employment Gazette*. The statistical content of these regular reports is limited to a broad description of the characteristics of applicants (who are almost invariably employees), the types of complaint brought, and the eventual outcome of the case. Occasional special exercises are also undertaken to provide more detail, and these are sometimes based on tribunal office files.

The information recorded in both tribunal and ACAS case files is variable, according to the circumstances of the case, and while there are certain procedures which have to be observed (producing some standardisation in the documents, particularly in the initial application), there are no set questionnaires or schedules which reduce all the information into a standardised format. However, custom and practice has evolved in relation to the administrative work of tribunals, so that certain information is customarily recorded precisely; for example careful notes are almost always made of those attending any meetings, and on the status of the parties' representatives (e.g. trade union officials, lawyers, employer associations). Case files thus contain information spanning across types C, E and H in Table 1, but not in a highly standardised form, and the ACAS files differ from tribunal records in coverage and content.

Although both sets of case files are confidential, bona fide researchers have sometimes been given permission to have access to tribunal files to extract data for research purposes. Some studies have been carried out by academic researchers who were granted access to completed case files in specific tribunal offices, others by professional social scientists in the Department of Employment. Some studies have been based on the datatapes containing anonymous and limited information extracted from the ACAS files for the production of the routine statistics on the tribunal system.<sup>15</sup>

#### Unemployment register and benefit records

Administrative records relating to the unemployed have as long a instory as the Wages Inspectorate records. Their most important, and most widely known, use is in the compilation of official statistics on the registered unemployed. But the records are also used as the sampling frame for interview surveys of the unemployed and for research analyses based on samples drawn from the records.

#### Catherine Hakim

The information available in the records has varied in line with changes in legislation and procedural arrangements. Traditionally the unemployed were required to attend at only one government office to register themselves as seeking employment and to claim unemployment and/or social security benefits. For a decade (from 1973 to 1982) the procedures were changed so that (compulsory) registration as seeking work took place in local offices of the Manpower Services Commission (Jobcentres or, for the young, Careers Advisory Offices) while claiming benefits took place at the local offices of the Department of Employment and/or the Department of Health and Social Security. The regulations were changed in the early 1980s su that from October 1982 the unemployed would again attend only one office to claim benefits, with registration at Jobcentres made voluntary. Whatever the procedures in force, it is clear that registration as seeking work and entitlement to benefits have always been closely associated, so that even in the decade where records were separately maintained by MSC, DE and DHSS offices, some items of information would be available from the records of all three (either because it was collected afresh by each one, or because information was exchanged between offices), while other information was uniquely recorded in one office only. The information recorded on unemployed benefit-claimants varies with the rules determining eligibility and entitlement (including, for example, the number of dependents for whom benefit will be paid), and includes a record of benefits actually paid, which can vary over the course of a spell of unemployment. The information recorded on those registered as seeking work includes that which is necessary for helping people to find jobs (for example occupation and whether seeking full-time or part-time work) and z variety of additional information required for statistical purposes (such as sex, country of birth, the industry in which the last job was held). Thus the information held in the records spans across types B, D, G, C, E and H in Table 1, with a particular data item potentially transferring from one category to another, as noted above. 16

From time to time the Department of Employment has carried our 'surveys' of the characteristics of the unemployed for management information and statistical purposes. These special exercises are in fact based almost entirely on extracting existing information from the records in local offices, but some additional information might be recorded—some of it based on the personal judgment of the local office staff (for example their views on a person's employability and artitude to work) and some of it being obtainable only by questioning the unemployed registrants (for example on whether they had an occupational pension or not in 1976). Local office staff were also responsible for implementing any sampling procedures applied in the study.<sup>17</sup> Research analyses carried out by academic researchers have sometimes relied on the datasets produced by these departmental special exercises, <sup>10</sup> but some have been done quite independently, with the researcher himself extracting and coding information from the records, usually those at a single local office.<sup>10</sup>

In the early 1980s, both the DHSS records relating to unemployed benefit-elaimants, and the MSC records relating to persons registered as seeking work, were computerised. Research analyses of data from the computerised records have had to rely on the coding procedures and classifications already applied to the information, but rather more complex sampling designs could be carried out reliably. Also special exercises to extract data from the records for a specified sample of claimants or registrants became feasible. The DHSS cohort study of unemployed benefit-claimants provides an early example of this use of the records for research purposes. For all men in the study sample who agreed to an interview, information on the benefits actually paid out to them were extracted from the records and added to the interview data. The procedure was adopted in preference to asking people about the benefits they received, to avoid problems of recall. Also, for all those who gave their permission, information on income and tax was extracted from the Inland Revenue records and added to the interview data. In this example, data from administrative records contributed only one part of the information used in the research study and was analysed in combination with information obtained through personal interviews with unemployed claimants. Use of the (computerised) records as a sampling frame for an interview survey and as the basis for research analyses was conflated into a single study. 20

#### Research procedures

There are essentially three procedures in research based on administrative records: the selection of a sample from the records; the application of coding or classification routines so as to convert the information into a standard format (amenable to computer analysis if necessary); and the analysis and interpretation of the results. To a large extent these are standard procedures typical of any research exercise. Those features which distinguish records-based research are described.

The selection of a sumple from the records can be problematic if the

#### Catherine Hakim

and is to obtain a random or representative sample. Almost all records form a continuous seriest they are constantly being added to or updated, often on a daily, weekly or monthly basis. There is thus no static universe from which to sample, and any sample will be a time-specific sample. This is also true of research surveys of course, but it is more obvious in the case of records-based research, since the researcher usually has a great deal of freedom to specify precisely the particular time period to be selected for a study. For example it is often possible to select more than one sample, from two different years for instance, in order to study changes over time or changes resulting from, say, the modification of policy or procedures. More generally administrative records can be used for single-time studies, before-and-after studies, time series, studies comparing regions (or other areas) within the country, and even for longitudinal studies of a particular cohort or of nationally-representative samples of cases.<sup>21</sup>

Most government records are maintained on a national basis, and constitute a census (in that information is available for all cases in the specified category), so that in principle nationally representative samples can be obtained by the researcher. In practice this is only feasible when the records, or some basic information derived from them, are routinely collated in a single location. The computerisation of records does just this, as there will almost invariably be a single computer holding the relevant records for the whole country. But some agencies have a regional network and records are maintained in regional, or other local, offices. Unless it is feasible to arrange for local offices to select samples, and also extract the required information, it is necessary for researchers to travel around the network of offices to do this themselves. In practice, researchers will more commonly limit their sample to just a few (or just one) specified locations, and accept that the sample so obtained cannot be nationally representative. Official statistics produced routinely from the records, or other information, may be available to indicate what is lost in the process of restricting the sample, and any limitations that therefore apply to the research findings. For example the Wages Inspectorate records are maintained in a network of some eighteen regional offices, and as yet no study has ever attempted to draw a nationally representative sample from all of them.

Some records concern one-off events, such as the birth and death registers, the probate register or the industrial tribunal records. Other records cover all the population within specified categories, and update the information on this group at regular intervals, such as the Inland Revenue tax records. Yet a third type of record contains information on people who flow into and out of a specified status or category, but remain there for varying periods of time. Examples here are unemployment and social security records. In this case one will have to choose between a flow sample or a stock sample. A stock sample is taken from all the records for people within the specified category, such as all persons registered as unemployed at a given time. A flow sample is taken from the records for people moving into (or out of) a specified category-for example all persons 'newly' registering as unemployed in a given period of weeks (or months), or all persons moving into the 'one-year or more of unemployment' category, or all persons 'coming off' the register (i.e. ceasing to maintain continuous registration). The choice of a flow or stock sample can have major implications for a research design. For example flow samples of the unemployed have a very different composition from stock samples of the unemployed as there are significant differences between groups with different durations of continuous unemployment. Similarly a sample of all prison inmates will have different characteristics from a sample of people newly committed to prison. Decisions as to which type of sample is most appropriate for a particular study can sometimes be informed by official statistics produced from the records in question. These will show whether there is any major duration effect on the composition of each type of sample. There usually will he.

Once the sample design has been selected, the second task will be to identify the type (or types) of information available in the records according to the classification in Table 3, and decide whether routine information only will be used or whether additional information (regular or ad hoc) will also be used. The amount (and difficulty) of work involved in the *conversion to a standard format* is determined in part by the nature of the information used, and in part by the extent to which the information has already been processed and coded. Although each set of records is unique and presents different problems and potential, some of the more common issues can be noted.

In a research survey, one knows from the questionnaire whether a particular question has been asked or nor. And when the answer to the question is missing, one knows also whether this was because the respondent refused to answer it, because the interviewer forgot to ask it, or because the interviewer decided the question was not applicable to the respondent, as these reasons are usually separately identified in the coding frame. With administrative records the process of finding out whether particular information was obtained or not, and why information is missing, is less clear-cut, because there is no standard

#### Catherine Hakim

documentation (questionnaire and coding record) available to the researcher. Documentation usually has to be created for the study.

In a survey, there is usually a written briefing document issued to all interviewers, supplemented usually by personal briefings which amplify the instructions given in the briefing document. An important part of these training materials concerned with the definitions and meanings applied both to the questions and to the (closed) answers in the survey questionnaire. Coding guides provide further information on how unusual or unforeseen (written-in) replies were coded or classified. There may be little or no equivalent background material for the information held in administrative records. Indeed in the case of 'special' exercises, the only way of knowing exactly how and why information was recorded in a particular way may be to talk to the recorders. In other cases the information may be contained in departmental circulars (issued from headquarters to local offices with written instructions on sampling and/or data to be extracted) which may be incomprehensible to someone not familiar with the system, so that consultation with the recorders is still required. There may also be some degree of regional or local variation in the type of information recorded on something as broad as 'reasons for X' which needs to he clarified. The (operational) definitions and classifications applied in practice may depart from the original instructions either because these were too open-ended (hence ambiguous) or because they were inoperable in their original form (usually because they were not rested). 22

Lack of familiarity with the content of records makes it difficult to set up in advance a standard coding frame to be applied to the information extracted from records. One method of overcoming the problem is to split the data extraction process into two stages: first copying information as given in the records, then later coding the information into a standard format.<sup>23</sup> If the copying process is done manually, the classifications adopted will be based on one complete 'reading' of the data and knowledge of the full range of responses. It will also clarify whether a single item of information can be coded on the basis of two or more classificatory schemes in order to fully tap the range of information recorded under a single heading. An obvious example here is the heading 'occupation' which may consist uf information on main occupation, on whether currently working or unemployed, on whether a second occupation or second job is also held and so on. However, the scope for recoding and creating new derived variables is usually more limited in the case of computerised records, where the information has already been coded to some extent. The analysis and interpretation of data from administrative records differs from similar work based on research surveys in two ways. First, because of the variable quality of data even from the same records, the variable validity and reliability of the findings may need to be stressed (and explained), particularly when both 'routine" and 'special' data have been used. Second, there may be a need to clarify whether, and if so how and why, the findings differ from those of any routine analyses of the records that result in published statistics.

It seems reasonable to expect that, in most cases, administrative records will prove less amenable than research surveys to secondary analyses with a focus quite different from that of the primary analysis. The data content is more limited and focused towards the primary, administrative, uses. So most research uses of the records will consist of some combination of four procedures: (a) applying a somewhat different perspective to the topic, as reflected in different conceptual framework and coding operations, (b) making full use of any contextual information collected but little used in previous analyses, (c) applying more extensive and more complex analysis routines to the dataset, and (d) extending the interpretations that can be placed on the data by setting them in the context of data or information from other sources and other studies. Any departures from, or modifications to, the conceptual frameworks and classifications applied in previous analyses of the data, and any new or different analytical concepts splied, must be spelt out clearly, to explain why the results present a different picture from that offered by previous analyses, and to avoid the danger that findings will be dismissed as incorrect because they are out of line with any reports previously published. 24

This overview of the procedures for quantitative research based on administrative records suggests that this type of research requires a combination of the skills of the *data analyst* (and hence some statistical skills) and the skills of the researcher using *documentaryvidence*. The latter are more common among historians than among social scientists engaged in quantitative research based on recently collected datasets.<sup>20</sup> The problems confronting the user of historical documents are rather greater than those presented by contemporary documentary evidence. The user of contemporary documents can, for example, consult some of the people responsible for the records, to obtain documentation on their compilation. This type of discussion with role-holders is not an interview, nor yet a discussion between research colleagues with similar interests and concerns. It is a crucial part of the research process, but one which has no label, no guidelines, and which can easily go awry. The third skill required for research

#### Catherine Hakun

based on administrative records is the ability to discuss the nature of the records on a without-prejudice basis with the recorders to *abtain valuable documentation* on the data which is not readily accessible. Interestingly, this third aspect of the method tends to be played down in reports by academics on records-based research, with little more than a cryptic acknowledgement of 'assistance'. This is somewhat ungracious, but it also devalues the important contribution of the data recorders to any study of this nature.<sup>26</sup>

#### Methodological problems

At the general level, the research procedures applied to records-based studies do not differ fundamentally from those applied to specially collected data. At the detailed level the problems arising are often different and demand an innovative approach to their solution. Although the problems, and solutions, will always be specific to a particular set of records and to a particular research use of them, this action gives the flavour of what is involved. The discussion is illustrated in the main by a study of the pay and earnings of homeworkers in Wages Council industries based on Wages Inspectorate records,<sup>[27]</sup> but illustrative examples from other records are also offered. The main feature of records-based studies is that the records exist already and cannot be created to the researcher's specification. The approach to research design has to be back-to-front. Instead of designing the study, and then collecting data in accordance with it. one gets full details of how the data was collected and/or recorded, and then identifies (or recognises) the research model to which it corresponds.

For example the records from the 1978/79 'special drive' of the Wages Inspectorate were identified as corresponding perfectly to the model of a sample of 500 homeworkers with a 'control' sample of 500 inworkers employed in the same area, the same industries and the same firms. Recognition that the data already available constituted a particular research design (even if this was accidental) pointed the way to standard analysis procedures, and also indicated that the standard strengths and weaknesses of marched samples would be present.

Identification of standard research designs and research procedures can be inhibited by the fact that the language of social scientists differs from that of the data producers. In the Wages Inspectorate study it was found that a high proportion of the records for homeworkers did not contain full information on pay and earnings. At first it was assumed that the missing data was attributable to the problems of contacting the homeworkers in their own homes, and hence to non-contacts (due to incomplete addresses) and refusals (from homeworkers who did not wish to be 'interviewed' by an inspector)-This raised the question of non-response bias, and some time was spent assessing its significance. Surprisingly, there was no evidence of non-response bias. Discussions with the Wages Inspectorate subscquently clarified that the procedures they used equated in practice with a type of sampling procedure, although the term as such was not used. Inspectors were not required to visit all homeworkers on a firm's lists, only to assess the wages of a sufficient number to assure themselves that the firm was not paying rates below the statutory minima to any homeworkers. Recognition of the sampling principle implied in the data collection procedure eliminated the question of non-response bias and led to the conclusion that the interviewed homeworkers were representative of all homeworkers in the sample (as had been shown already by the tests for 'non-response' bias).

The main point here is that the data collection and recording procedures may not be consciously and intentionally designed along specified methodological lines, but the results may nevertheless correspond in practice to standard methods or research designs. Recognition of them, *post hoc*, enables conclusions to be drawn about the representative nature of the dataset or the appropriate forms of analysis. But the process of model-recognition is left entirely to the researcher, with few hints being offered by those who produce the records.

A set of records will sometimes have missing information, in that no entry will be found in a clearly labelled box or space in the forms used, or information routinely recorded on a file is absent. In many cases it will be possible to deduce or impute the missing data item. In most others it will be possible to investigate the range of reasons for the missing information and to carry out tests for non-response bias on the basis of other data items in the records.

Where a large number of cases are being routinely recorded, simplified recording procedures are often adopted, in order to save nime. The most common is the use of ditto marks to indicate that the information for consecutive cases is the same as for the first fully recorded case. Sometimes even ditto marks are omitted, so that a blank entry leads to a suspicion of missing information. For example in the Wages Inspectorate records many items of information were left blank and were apparently missing. Discussions with the recorders -larified those cases where ditto marks were applicable, and where the missing data could thus be inserted.<sup>28</sup>

#### Catherine Hakim

Where an item of information rarely has a value different from the norm, it may also be left blank except for the exceptions to the rule. For example the Wages Inspector's decision as in whether a worker was statutorily underpaid, and whether they were below 'ordinary' status in their work efficiency, was recorded only for the rare cases where workers were determined to be statutorily underpaid or sub-ordinary. In all other cases (the vast majority) this item was left blank and could have been mistaken for missing information. (For example a feasibility study based on a small sample of the records would suggest that the information was never recorded at all.) Again, knowledge of the simplified recording procedures adopted in practice allowed this missing data to be imputed for all cases.

It would appear that unwritten rules on simplified recording procedures are developed most commonly on routine or regularly recorded data items rather than on special items. In the latter case, the only reliable procedure would be to treat all blank entries as missing information, and to carry out systematic checks for non-response bias on *each* item of information with a high level of missing data. If there is a high level of randomly distributed missing data, this can severely limit the range of analyses applied to the dataset.

Given the absence of documentation on a set of records, the process of identifying the data available in the records is not always straightforward. In particular records may contain *data that is not explicitly recorded*. An obvious example is information on sex, which can almost always be deduced from names alone, or from the attached titles (Mr, Mrs or Miss) even if not explicitly recorded. In the Wages Inspectorate records there is no single item of information which specifies whether a worker is an in-factory worker or a homeworker. But the procedures adopted in practice by inspectors allow each worker's status to be reliably deduced. Inspectors would invariably group together all inworkers and all homeworkers in separate lists for each establishment. Also, while names were listed for all workers, home addresses would only be listed for homeworkers. Thus workers could readily be classified to the inworker or homeworker category even though the inspection forms did not explicitly state this information.

In some cases information not recorded can be derived from other information by straightforward calculation. For example the Wages Inspectorate records on assessments of inworkers' pay did not include information on their average hourly earnings. As the records did provide data on weekly earnings and hours worked per week, a straightforward calculation provided the data on hourly earnings for inworkers. When records consist of case files and all paperwork connected with a case, the process of identifying and extracting items of information from the papers is even more extensive than with records compiled on forms. For example the case files of Industrial Tribunals usually contain papers from which it is possible to identify whether an applicant was represented, or assisted by, a Trades Union official, even though representation of the applicant (and the employer) is not separately and formally recorded as an item of information on any form.

It would appear that routine, regular and special data collections may all provide instances of information being available in the records even though it is not explicitly recorded.

By far the most common difficulty encountered in records-based research is *inconsistently recorded information*. Whether records are used for statutory decision-making or for management purposes, their common feature is a focus on the details of each individual case. There is hence less standardisation than in data from research surveys, for example, and some similarity to case-study material. A major task in the coding process will be to devise a standard set of response categories to which the varied information recorded can be reduced. This will generally be a great deal easier for routine and regular data items than for special data items, and easier for data extracted from forms than from case-files.

For example, in the Wages Inspectorate records, it was not uncommon for the inspector to record details of earnings and hours worked over a four-week period instead of a one-week period. Also the assessment procedures differed between workers on time-rates and those on piece-rates (paid by the hour or by the piece). For workers paid on piece-rates, the assessment procedure might differ between those working solely on one type of work (and hence a single piece-rate) and those doing different types of work, to which a variety of piece-rates applied. Thus the information recorded for an assessment might vary from a few items of information at one extreme to a long and complicated list of piece-rates, size of batches, and earnings over a four-week period at the other extreme. For the purposes of a research analysis covering over 1,000 cases, it was necessary to condense and summarise the information into a set of standard data items, so that comparisons between homeworkers and inworkers could be made. Since all the information consisted of numerical values (hours worked, piece-rates, earnings and so on) a complex system of averaging and estimating was applied to reduce the detailed information into a set of standard data items (average weekly earnings,

#### Catherine Hakim

average hours worked per week and average hourly earnings) for all workers in the sample: This meant that the data on which the analysis was based differed substantively in many cases from the detailed information on which the inspector's assessment was based, a point that had to be emphasised in the interpretation of the research results.

Another example of estimating procedures being used to convert inconsistently recorded information into a standard format is given by Baxter's study based on unemployment records. In this case the objective was to complete a quarterly time series on unemployment spells among chronic job changers!<sup>29</sup>

In the case of the additional information specially collected by the Inspectorate during the winter of 1978/79, the causes of inconsistently recorded information were attributable to the data quality problems observed earlier with reference to special data items. For about 40 per cent of homeworkers the information on 'length of experience in the trade' was not recorded. When the information was recorded, it was with a marked degree of random variation. For example the homeworker's total experience in the trade might be recorded, and/or their experience as a factory inworker, and/or their experience as a homeworker only. On 'reasons for homework' references to children were by far the most common entry, but the entries were usually ambiguous and unspecific (for example 'children', 'young children', or 'family'). From this information it was not possible to differentiate between homeworkers with children below school age (who are always at home), those with children of school age only (who are not at home during the day in term time at least), and those with older unmarried children (who would not present the same child-care problems as the first two groups).

Similarly it was not possible to differentiate between those who had originally begun working at home because of their children (and were now continuing out of habit and convenience) and those homeworkers who were currently working at home because of young children. The fact that very variable meanings could be attached to these responses was emphasised by cross-tabulations with data on homeworkers' ages and the length of their experience as homeworkers (which might run to 20 years). In this case there was no possibility of clarifying or amplifying on the information available in the records and the weakness of the data had to be taken into account in the interpretations placed on the results.

In any set of records where data is recorded other than for statutory decision-making purposes, there is the possibility of a systematic slant or bias in the recordings, arising from the organisational or work. context or from uses of the data. (The term bias is used here in its non-pejorative sense of a swaying influence.) This type of problem in the most elusive and least visible. It will sometimes come to light from analyses of the data extracted from records; sometimes it will only come to light from discussions with those directly involved in making up the records. One well-known example is the tendency for trade unions to offer inflated rather than deflated figures on union membership derived from their administrative records, as unions exert voting influence in the TUC according to their membership size.

Discussions with the Wages Inspectorate revealed a systematic slant in the information recorded under the heading 'reasons for homework' that was not shown up in any of the analyses and would otherwise have remained undetected. Because the information on reasons was recorded in the context of a pay assessment, there was a strong propensity, in practice, to record those reasons that were relevant to a pay assessment rather than to obtain full information on reasons. Thus the reasons actually recorded were those that might hinder efficient working methods or contribute to low skill or output. Knowledge of this systematic slant did not of course invalidate the data, but it justified a somewhat different interpretation being placed on the research findings than would otherwise have been the case. It supported an interpretation of references to children as the reasons for homework as subjectively valid: the homeworker who referred to children or family commuments was saying something about her priorities which could be expected in practice to have an influence on her work output and productivity even if children of school age or above did not present an objective impediment to working outside the home, 30

Other examples of systematic bias or slant in the information in records come from the unemployment register maintained by the MSC local offices (Jobcentres and the like). The recording and coding of occupation in the unemployment register has a consistent downward bias in occupational grade, which produces a disproportionate number of people in the lowest category of unskilled 'general labourer'. This is because occupation is recorded, for operational purposes, as the occupation considered most appropriate for finding someone a job, taking account of the local unemployment situation, local job opportunities and any other relevant factors. It is not necessarily the last occupation they held, nor their usual or main lifetime occupation as 'a local office may decide that a wider range of unskilled jobs might be suitable for an unemployed man, and they may therefore classify him as seeking a 'general labouring' job."<sup>31</sup>

#### Catherine Hakim

The nature and direction of this particular bias in the unemployment register records is known from how the system works, but is also reflected in differences between the occupational distribution of the unemployed as shown by register statistics and by interview surveys (such as the biennial Labour Force Survey). <sup>31</sup>

A second example concerns additional information collected for the June 1976 'survey' of unemployed registrants. As noted earlier this 'survey' was based almost entirely on extracting existing information from the records for a sample of registrants, but local office staff were also required to record their assessment of the person's employability (or prospects of getting work) and attitude towards work. Analysis of the results showed such a strong overlap between poor prospects of obtaining work and an 'unenthusiastic' attitude towards looking for work as assessed by local office staff that it is questionable whether these were really two independent items of information. Almost all (96 per cent) of the men who were judged by the local office to have "good' or 'reasonable' prospects of getting work were also judged to be 'keen' to find a new job, while only half (53 per cent) of the men who were judged to have 'poor' prospects of finding work were judged by local office staff to have an 'unenthusiastic' attitude to getting work. That assessments of attitude were determined by, or a reflection of, appraisals of job prospects was underlined by the experience of the men in the following six months: about two-thirds of the men with 'good' or 'reasonable' prospects left the register for employment, compared with one-third of those with 'poor' prospects. Among those with 'poor' prospects, attitude towards work made no difference to the likelihood of getting work: both among those with 'keen' attitudes and those with 'unenthusiastic' attitudes, one-third left the register for employment within six months.35 Here again, knowledge of the slant in the information recorded on registrants did not invalidate the data, but supported a very different interpretation being placed on the results of the analysis than would otherwise have been the case. In this case, the results were used to illustrate the processes surrounding discouraged workers.

In some studies, researchers are not allowed access to the original records to extract data themselves, but are provided with data specially extracted for them by the records producers. In these cases, the researchers' main contribution will be to extend interpretation of the data by analysing them in conjunction with data from other sources. A problem that often arises is the *incompatability of operational definitions* of a variable in the different sources. One example here is the Home Office study of arrest rates for whiten and ethnic minority groups based on data specially extracted from police records for London and on population census data on whites and ethnic minorities in London. Since the data extracted from police records was already coded and could not be further manipulated, the solution here was to obtain special tabulations of the relevant census data, based on definitions of ethnic groups which matched the data from the records as closely as possible, <sup>34</sup>

So far only six types of methodological problem have been illustrated. There may be other types, occurring perhaps only in particular types of record. However, the examples given have also illustrated how each problem is open to solutions. Sophisticated research designs can be recognised as present in a dataset, even if the data cannot be created to order. Incomplete information can be made good, gaps supplied by deduction or from information supplied by the recorders, or the reasons for missing information can be clarified. Although slants of one sort or another often arise in data derived from records, knowledge of them can clarify the interpretations placed on findings. And although records often contain very varied information on each particular case, it is usually possible to reduce this to a standard form of data amenable to computer processing and systematic analyses. The difficulties are real, but not unsurmountable.

#### Conclusions

This review has been concerned with a type of data that is widely used in central and local government yet is almost entirely unacknowledged in the literature on social research methodology which is produced almost exclusively by academics.<sup>35</sup> This may explain (if it does not excuse) why academic reports on records-based research typically give almost no information on their methodology, the problems encountured and solutions adopted.

The common element in all the records discussed here is that they are very large in number, national in coverage, sometimes constitute complete censuses of the events recorded (sometimes representative samples), are usually available for considerable stretches of time and are usually high in their content rigidity and narrow in content scope. They can be converted into quantified data on a reasonably systematic basis for quantitative analysis. The business of coding and classifying the information oneself is what distinguishes this type of research from that based on statistics derived from administrative records. The fact that one has no control over the original recording process, but can usually gain access to the recorders to discuss their procedures

with them, is what distinguishes this type of research from the truly reactive type based on interviews and questionnaires, and from the truly non-reactive type based on historical archival records.<sup>36</sup>

Administrative records can provide information that is unique, either because no other existing source provides information on the topic (as in the example of Industrial Tribunal records) or, an even stronger argument, no other data collection method could provide the information required. One of the examples discussed here-Wages Inspectorate records on the earnings and skill level of homeworkers -is of the second type. A third advantage is that records can be used to study the administrative process itself, as illustrated by studies based on local authority housing or social service records. These arc sometimes used to shed light on the decision-making process in the delivery of services, for example the criteria applied in rating housing need, or need for a place in a day nursery, or to assess whether racial discrimination affects the allocation of applicants to Council housing estates. Similiarly company personnel records can be used to assess whether racial or sex discrimination appears to be operating at establishment level. A fourth advantage is that studies can be mounted very quickly and relatively cheaply compared to the timescales and costs involved in specially designed research surveys. And where the topic involved is sensitive for some reason, there are advantages to using unobtrusive and non-reactive methods instead of interviewing people directly about the issue.

Given the limited information typically available in records, and its variable quality, it seems unlikely that a records-based study could ever carry the whole burden of proof for any research conclusions. Throughout this article, reference has been made to the need to use other sources of data to corroborate, complement and clarify the results of analyses of administrative records. The case for records-based research is much stronger if this type of study contributes one part—but not the whole—of a research project.<sup>37</sup> This conclusion would be even stronger in the case of policy research.<sup>18</sup>

The trend towards computerising administrative records, both in central and local government, may lead to expanding interest in records-based research. Computerisation usually entails centralisation of records previously dispersed in the files of local offices, so that samples of the records can more easily be extracted for analysis. Also an increasing number of datasets of this type are finding their way into the SSRC Data Archive at the University of Essex. The Archive already holds a few datasets based on administrative records, most of which have never previously been accessible, such as the Department of Education and Science datasets on educational institutions, manpower and related aspects of the educational system, and the Department of Employment data on strikes. The computerisation of unemployment records has made it possible for some datasets derived from these files to be deposited at the Archive. The DHSS cohort study data is being deposited, and this contains data from unemployment benefit records. As complex multi-source datasets such as these become more common, other surveys of this type may become available. Developments in data archiving may thus lead academic social scientists to take greater interest in the research potential of this type of data.

But ultimately the case for records-based research will rest on the value of the studies they produce. The main example discussed here was the study based on Wages Inspectorate records. Although only one part of a research programme on homeworking, it contributed a key element and information that could not have been obtained from an interview-based survey, however large and well-funded. Homeworkers are notoriously vague about their earnings, and the range of earnings varies a good deal. The study based on Wages Inspectorate records was the only element in the research programme that was based on *factual* data on homeworkers' earnings, and had the added attraction of providing earnings data for a matched sample of inworkers as well. Finally, the records provided data on the statutory rates of pay applicable to each worker, and on the inspector's assessment of the worker's pay, information that would have been almost impossible to gain by any other method, however well-funded. The difficulties, both practical and substantive, of records-based research will only be overcome if the value of the results is seen to justify the research expertise required by this type of work, so that the reports produced will ultimately be the most eloquent testimony in favour of the method.

> Received 2 August 1982 Finally accepted 30 March 1983

"The author is a Principal Research Officer with the Department of Employment but this article was written in a private capacity, any views expressed are those of the author and not necessarily those of the Department. Grateful acknowledgments are due to colleagues whose comments on an earlier draft helped improve this paper, in particular W.R. Hawes.

#### Notes

See for example P. M. Blau. The Dynamics of Bureaucrusy, Chicago: University of Chicago Press, chapters 2-6; T. Leggatt, Management style and economic success in industry', in Sociological Theory and Survey Research, (ed.) T. Leggatt, London, Sage, 1974, pp. 185-205; R. Price and G. S. Bion, 'Union growth revisited: 1948-1974 in perspective's British Journal of Industrial Relations, vol. 14, no. 3, 1976, pp. 339-55; D. W. Marsden, 'Pilot project on the working of internal labour markets' research report to the Social Science Research Council, 1981; P. L. Payne, The Early Scottish Limited Companies, 1856-1895: An Historical and Analytical Survey, Edinburgh, Scottish Academic Press, 1980; G. F. Dreher, 'The role of performance in the turnover process', Academy of Managment Journal, vol. 25, no. 1, 1982, pp. 137-47; W. D. Rubenstein, Men of Property, London, Croom Heim, 1981; E. Gittus, 'A study of the unemployed of Merseyside', in Merseyside: Social and Economic Studies, (eds) R. Lawton and C. M. Cunningham, London, Longman, 1970, pp. 324-73; J. L. Baxter, "The chronic job changer: a study of youth unemployment', Social and Economic Administration, vol. 9, no. 3, 1975, pp. 184-206; R. McNabb and N. Woodward, "The effect of recurrent spells upon unemployment duration", British Journal of Industrial Relations, vol. 20, no. 1, 1982, pp. 105-8; J. Wrench and G. Lee. 'Piecework and industrial accidents: two contemporary case studies'. Succology, vol. 16, no. 4, 1982, pp. 512-25; M. McConville and J. Baldwin, Courts, Protecution, and Conviction. Oxford, Clarendon Press, 1981; B. Chiplin, 'An alternative approach to the measurement of ses discrimination: an illustration from university entrance. The Economic Journal, vol. 91, 1981, pp. 988-97; W. S. Siebert and P. J. Sloane, "The measurement of sex and marital status discrimination at the workplace. Economica, vol. 48, no. 190, 1981, pp. 125-41.

- 2 See for example J. Parker and K. Dugmore, Colour and the Allocation of GLC Housing: the Report of the GLC Lettings Survey 1974-75, Research Report no. 21, London, Greater London Council, 1976; Department of Employment, 'Characteristics of the unemployed: sample survey, June 1976, 'Some further characteristics of the unemployed' and 'Characteristics of the unemployed: regional analysis', Department of Employment Gaactie, vol. 85, 1977, pp. 559-74, 965-75 and 1122-6, R. McNabb, N. Woodward and J. Barry, Unemployment in West Cornwall, Research Paper no. 8, London: Department of Employment (see pp. 17 ff.); P. Stevens and C. F. Willis, Race, Crime and Arrests, Home Office Research Study no. 58, HMSO, 1979; R. Leete and S. Antheny, 'Divorce and remarriage: a record-linkage study', Population Trends, no. 16, 1979, pp. 5-11.
- 3 Both of the two main texts on secondary analysis focus on data from interview surveys, and the latter makes only brief references to research based on administrative records. H. H. Hyman, Secondary Analysis of Sample Surveys: Proteiples, Procedures and Potentialities, New York John Wiley, 1972; C. Hakim, Secondary Analysis in Social Research, London, Allen & Unwin, 1982, pp. 6-7, 143-4.

4 C. Davies, 'Making sense of the census in Britain and the USA: the

changing occupational classification of the position of nurses', Sociological Review, vol. 28, no. 3, 1980, pp. 581-609; C. Hakim, 'Cansus reports an documentary evidence: the census commentaries 1801-1951', Sociological Review, vol. 28, no. 5, 1980, pp. 551-80, J. Platt, 'Evidence and proof in documentary research', Sociological Review, vol. 29, no. 1, 1981, pp. 31-66.

- 5 See for example E.A. Wrigley (ed.), An Introduction to English Historical Demography, London, Weidenfeld and Nicolson, 1966; E.A. Wrigley (ed.), Nineteenth Century Society: Enays in the Use of Quantitative Methods for the Study of Social Data, Cambridge University Press, 1972; and Chapter 6 in C. Hakim, Secondary Analysis in Social Research, London, Allen & Unwin, 1982, pp. 82-94.
- 6 For example Hindess's concluding summary notes (page 47) that scress to the original schedules may allow the reconstruction of categories not identified in published statistics, and re-analyses of the information along different lines. But he does not go on to discuss records-based research. Also he deals solely with the conceptual difficulties which may arise in a secondary analysis of any data, but with no other problems at all It seems self-evident that the information in records will not necessarily be amenable to reclassification into 'sociological' variables. but they are very likely to be amenable to reclassification into the policy-relevant variables required for policy research, so that the conceptual problem may be completely overshadowed by others. See B. Hindess, The Use of Official Statistics in Sociology, London, Macmillan, 1973. Another review simply notes in passing (page 133) that government statistics are based sometimes on specially designed data collection and sometimes on data from administrative records, but does not consider the effects of this major difference. See J. Irvine, J. Miles and J. Evans, Demystifying Social Statistics, London, Pluto Press, 1979.
- 7 In many cases additional information is collected with mixed purposes: both for management information and to produce published statistics (a for example in the case of unemployment records which provide the basis for extensive and regular published statistics). The advent of the Government Statistical Service (GSS) probably played a large part in extending the amount of additional information collected (and tabularcil); however the Rayner Review of the GSS argued that only that information which was essential for management information purposes should be collected and as a result there were major cut-backs in the GSS-(a one-third cut in the Department of Employment for example).
- 8 The issue of professional ethics raised by this change (from E to D in Table 1) is of some interest and appears to have been at the root of local offices' and trade union objections to it. Benefit claimants could hardly be expected to appreciate the difference between information collected for the purpose of assessing their statutory benefit entitlement and other (non-essential) information collected only for information (and hence of a voluntary basis).
- 9 Given a statutory requirement in co-operate with some of these data collections (such as the censuses of population and of employment) it is possible that the distinction between type 2 and type X data is not always clear to respondents either. See note 8 above.

- 10 Assessments of value for money are always difficult in the case of statistical or research information, particularly as the initial (known or intended) uses would often allow for fairly rough and ready data but subsequent (unpredicted) uses often go well beyond them and really require fairly high quality data. Special exercises that are based on administrative records often appear cheaper than specially mounted surveys only because their true cost is hidden in general administrative overheads.
- Hindess's op. cit. 'critique' relies primarily on the Indian census officials' own published account of what had, and had not, worked well in the conduct of the 195] agrarian census.
- 12 Asseminents of official statistics tend to concentrate on their coverage and the level of undercount or overcount rather than on the quality of the information obtained for those who are covered, as illustrated by Garside with reference to unemployment statistics. However Lord, writing from the perspective of a producer of official statistics, is more informative about the variable quality of particular data items, and of information collected for statistical purposes but not required for operational purposes. S. Lord, 'Unemployment statistics in Britain', in The Economics of Unemployment in Britain, (ed.) J. Creedy, London, Butterworths, 1981, pp. 235-54, especially p. 246. W. F. Garside, The Measurement of Unemployment in Great Britain 1850-1979, Oxford, Basil Blackwell, 1980. It is worth noting that the quality of all categories of data in records can be affected by a major breakdown of the administrative machine-for example during periods of work overload or sprikes. See S. Keeble, 'Social security on strike', New Society, 28 October 1982, pp. 168-9.

13 M. Builmer, "Why don't sociologists make more use of otherwistatistics?".

- Sociology, vol. 14, no. 4, 1980, pp. 505-23; see page 508.
- 14 For further detail see C. Hakim and R. Dennis, Homeworking in Wages Council Industries, Research Paper no. 37, London, Department of Employment, 1982
- 15 Further detail on the Industrial Tribunal records is given in W.R. Hawes and G. Smith, Patterns of Representation of the Parties in Unfair Dimnissal Cases, Research Paper no. 22, London, Department of Employment, 1981. This report also reviews (on pages 11-18) the independent studies based on the records. One reason for the discrepuncica between research analysis results and routine 'official' statistics an the tribunal system is that the former are commonly based on Tribunal records while the latter are based on ACAS records.
- 16 Further detail on the unemployment records is given in Lord, op. cit., and Garside, op. cit. Although Garside covers a longer period, his focus is primarily on the statistics produced from the records.
- 17 A complete list of the 'surveys' carried out between 1923 and 1976 is given in Garside, op. cit., pp. 139-51. Results of the 1980 study are given
- in the Employment Gazette in 1983. 18 Gittus, op. cit. The Manpower Services Commission has also done
- further work with the 1976 survey.
- 19 Baxter, op. cit.
- 20 The first three reports (in the August 1980, January 1981 and September

1981 issues of the Employment Gazette) from the DHSS cohort study of the unemployed were based on the interview data alone. But later reports made use of benefit data from records in conjunction with the interview data: R. Davies, L. Hamill, S. Moylan and C. H. Smee, 'Incomes in and out of work', Employment Gazette, vol. 90, no. 6, 1982, pp. 237-43; B. Davies, The DHSS Cohort Study of Unemployed Men. Working Paper No. 2: Men whose income out of work was high compared with their incomin work, London, DHSS, 1982, For an earlier example of interview data being linked to data extracted from tax returns and social security records see W.H. Scwell and R. M. Hauser, Education, Occupation and Earnings: Achievement in the Early Career, New York, Academic Press, 1975.

- 21 The DHSS cohort study of the unemployed (see note 20 above) provide longitudinal data on benefits paid to the cohort over a period of three year. However it is not normally possible to construct samples retrospectively, and prospective studies are most common, as in the case of the DHSS study, which 'followed up' the sample for three years. This problem may now be overcome, as the computerisation of the records on unemployed benefit-claimants will lead to the creation of a standing 5 per cent national sample (based on National Insurance numbers) in order to monitor the incidence and duration of spells of unemployment (more procisely, those spells producing benefit claims) in the whole labour torce.
- 22 Hindess, op. cit., pp. 29-35, notes an example of the second type. although-because it relates to a census rather than administrative records-it was signposted in published reports.
- 23 This approach has been adopted for research analyses of nineteenthcentury census records. Since the records are now held on microfilm, in order to preserve them, researchers obtain complete copies of the microfilms for specified areas or for national samples. Even when the census records are computerised (as has been done with the complete 1801 Norwegian census and a 2 per cent sample from the 1851 British census), they are entered verbatim into computer files so as to recain complete flexibility in any coding operations applied to the information See S. Rokkan, 'Data services in Western Europe: reflections on variations in the conditions of academic institution-building', Amarican Behavioural Scientist, vol. 19, 1976, p. 452; M. Anderson et al., The national sample from the 1851 Census of Great Britain', Historical Methods Newsletter, vol. 10, no. 3, 1977, pp. 117-21.
- 24 The old adage that novel results are usually wrong, used by statisticiany as a rule-of-thumb for checking statistics prior to publication, is equally valid for research analyses based on data from administrative records. A good example of this is to be found in recent discussion of the remedies sought by applicants at Industrial Tribunals, P. Lewis relied unquestioningly on (previously unpublished) data drawn from administrative records which showed that most successful applicants in unfair dismissal cases sought reinstatement but received compensation, whereas those who were re-engaged had wanted a cash remedy. This was a novel and striking finding. But (as Dickens notes) all other evidence based on interview surveys and other methods produces opposite conclusions.

suggesting that the data drawn from the records was urueliable, and should have been questioned by the researcher. See P. Lewis, 'An analysis of why legislation has failed to provide employment protection for unfairly dismissed employees'. British Journal of Industrial Relations. vol. 19, no. 3, 1981, pp. 316-26; L. Dickens et al., "Why legislation has failed to provide employment protectiont a note', British Journal of Industrial Relations, vol. 20, no. 2, 1982, pp. 257-8; K. Williams and D. Lewis, The Afternath of Tribunal Reinstatement and Re-engagement, Research Paper no. 23, London, Department of Employment, 1981.

- 25 Platt, op. cit.
- 26 It is clear that Baxter benefited from a great deal of active collaboration with his study from local office staff, which is acknowledged as 'generous assistance'. Baxter, op. cit., pp. 184, 197, 200-4. But other academics fail to admit to such assistance and to the role played by the data producers in a study.
- 27 See note 14 above.
- 28 Similarly ditto marks, and blanks which can be taken as their equivalent, were often used in the nineteenth-century census enumerators' books, for example where all members of a household have the same surname, address or occupation.
- 29 Baxter, op. cit., pp. 188-9.
- 30 The validity of the interpretation was supported with reference to two other studies of homeworkers, one based on depth interviews: A. Cragg and J. Dawson, Qualitative Research Among Homeworkers, Research Paper no. 21, London, Department of Employment, 1981, and another based on unstructured interviews: E. Hope et al., 'Homeworkers in North London', in Dependence and Exploitation in Work and Marriage, (eds) D. L. Barker and S. Allen, London, Longman, pp. 88-108.
- 31 Department of Employment, 'Characteristics of the unemployed; analyais by occupation', Department of Employment Gazette, vol. 82, no. 5, 1974, p. 385.
- 32. Discrepancies are due also, but to a lesser extent, to the tendency among some people to slightly upgrade their occupation as reported in interview surveys. See Table 19 in C. Hakim and W. R. Hawes, The Labour Force, D291: Statistical Sources, Milton Keynes, Open University Press, 1982, p. 15.
- 33 Department of Employment, op. cit., 1977, pp. 561, 565, 966 and 970.
- 34 Stevens and Willis, op. cit., pp 46-7.
- 35 One conclusion to be drawn from this review is that the metholodogical guidelines devised by academics are partial and inadequate for professional researchers working in central and local government. The range of data sources and of research methodologies used in government go much wider than those utilised by academics. This (in addition to the fact that government research is usually multi-disciplinary while university courses are usually focused on a single discipline) may contribute to the complaint that social science graduates are often ill-equipped for government research (as voiced for example by R. V. G. Clarke, 'The effectiveness of graduate education in sociology; employment in central government research', Saciology, vol. 16, no. 4, 1981, pp. 525-30). It tiso implies that extra training in research methods has to be provided

within employing departments and local authorities.

- The classification offered by Webb, et al. is useless in relation to administrative records. See E. J. Webb et al., Unohtrustve Measures, Chicago, Rand McNally, 1966.
- 37 For example Blau's analysis of an employment agency's records was complemented by data obtained from personal interviews and three months' observation of the agency's activities. See Blau, op. cit., pp. 19-117
- 38 See for example the suitably cautions conclusions in Siebert and Sloane, UD. Cit.