COMPUTERISING THE 1861 CENSUS ABSTRACTS AND VITAL REGISTRATION STATISTICS

David Alan Gatley

David Alan Gatley is a Research Officer in the School of Social Sciences at Staffordshire University. He is currently computerising the mid-nineteenth century census abstracts and vital registration statistics. He has written several papers on the Staffordshire Potteries and Victorian Warrington.

The Victorian Census Abstracts, that is, the printed census volumes, the Annual Reports of the Registrar General and the returns of the Poor Law Commissioners are neglected sources in the study of nineteenth-century society. Yet these sources provide the researcher with a wealth of information on matters as diverse as population change, occupational structure, migration, literacy, causes of death and much more for each one of the more than 600 Poor Law Unions or Registration Districts into which England and Wales were divided in the mid-nineteenth century. Limited subsets of these data sources have been made machine-readable by other workers in this field, but until the work described in this paper, no systematic attempt had been made to create a database for most of the available sources for a single year. This has now been completed for the 1861 census.

This paper will examine the value of the database to both researchers and local historians. But, first it will briefly consider the nature of the sources, their accuracy and how they were made machine-readable.

Structure of the returns

Details of how the nineteenth-century censuses were structured can be found in a variety of sources. It is sufficient to note that the responsibility for undertaking the 1861 census lay with the General Register Office (GRO) in London. Information in the census abstracts was presented at a number of different levels:

- a) England and Wales as a whole;
- b) divisions or regions;
- c) registration counties;
- d) large towns and cities;
- e) registration districts;
- f) sub-registration districts;
- g) places (parishes, townships, chapelries, etc.).

The degree of information published varied at each level, being most detailed at national level and least detailed at the level of the individual place. Our interest,
however, lies in the registration districts, because it was at this level that information for this project was extracted from the census volumes. The registration districts were based upon the Poor Law Unions that had been created in the 1830s. Many of the Poor Law Unions were, in turn, based upon market towns, and included the surrounding parishes which were serviced by their weekly market and other facilities. As the Poor Law Commissioners themselves put it:

The limits of unions which we have found most convenient are those of a circle, taking a market town as a centre, and comprehending those surrounding parishes whose inhabitants are accustomed to resort to the same market. This arrangement was found highly convenient for the weekly attendances of the parish officers, and some portion of the guardians. Some auxiliaries to good management were derived from the town itself.  

The number of registration districts, however, grew during the nineteenth century by a process of sub-division and the re-drawing of boundaries. Whilst in 1851 there were 624 of them, by 1861 there were 635.

The tabular results from the 1861 census were published in two key volumes. Volume One contained details of the areas, populations and numbers of houses and buildings in each place (that is, in each parish, township, chapelry, etc.), and a summary table gave the same information for each registration district. Volume Two included the following tables for each registration district: the ages of males and females in five-year age groups; the numbers of boys and girls aged under five; the marital condition of males and females broken down by age group; the occupations of males and females aged 20 and over; the birthplaces of males and females aged under 20, and aged 20 and over; the numbers of blind, deaf and dumb people; and the inmates of workhouses, prisons, lunatic asylums and hospitals.

The Registrar General was also responsible for compiling national statistics of the numbers of births, marriages and deaths that had occurred in each Registration District. These statistics were compiled from the copies of birth, marriage and death certificates sent to the GRO every three months by superintendent registrars. These statistics were published in a series of abstract tables in the Registrar General's Annual Report. The abstract tables of births, marriages and deaths showed the numbers of legitimate births, illegitimate births, marriages and deaths occurring in each registration district in the preceding twelve months. The abstract marriage table included details of the religious denominations of marriages, the numbers of marriages occurring in each quarter, the marital status of brides and grooms (that is, whether they were single, widowed or divorced), the numbers of brides and grooms who were aged under 21 and the numbers 'signing' the marriage register with a mark. The latter can be used as a crude measure of literacy.
To coincide with the 1861 census, the Registrar General published a *Decennial Supplement* which included a series of tables showing the numbers and causes of deaths occurring in each registration district in the previous ten years.

The *Returns of the Poor Law Commissioners* were compiled once every six months from statistics sent to the Central Poor Law Commission by the Boards of Guardians of each Poor Law Union. These figures show the numbers of people in receipt of both indoor and outdoor relief and the numbers of lunatics in each Poor Law Union.

**The accuracy of the printed census abstracts**

At this stage the accuracy of the printed census abstracts will be examined. Inaccuracies in both the ages and birthplaces of individuals in the census enumerators' books (CEBs) have been reported by, amongst others, Anderson and Perkyns who have traced individuals between censuses. The fact that the errors occur for the census records of individuals is well known, however, with regard to this work, the fact that the 1861 census reports aggregated the ages of individuals into five-year age bands and reported the *counties of birth* of individuals rather than their specific birthplaces, the published tables are likely to be more accurate than one might suppose given the discrepancies that have been found in the CEBs. First, discrepancies in the ages or the birthplaces of individuals in the CEBs tend to be small. The ages of individuals rarely differ by more than one or two years from those expected, and where differences have been found in their birthplaces they more often than not report places within the same *county*. Second, errors will possibly cancel one another out to some extent. In 1851, the census authorities were reasonably confident about the accuracy of their tables relating to the ages of the population, although some concern was expressed regarding the accuracy of the ages of women in their early twenties.

A more serious limitation of the birthplace data in the printed census volumes relates to the different ways in which the counties were defined in the nineteenth-century censuses. Generally, two different types of county were distinguished: *ancient counties* and *registration counties*. Ancient counties were simply the *civic* or *historic* counties into which England and Wales had for long been divided. In contrast, *registration counties* were defined by simply amalgamating together those registration districts that fell entirely or *mainly* within an *ancient county*. This difference is important because registration districts sometimes crossed county boundaries and were thus located partly within two or more *ancient counties*. Unfortunately the birthplace tables in the 1861 census abstracts record the *ancient counties* of birth of individuals living in each registration district. In consequence, in those instances in which a registration district crossed a county boundary the birthplace tables cannot be used to calculate the proportion of the population who had been born outside the county.

The accuracy of the occupational tables given in the 1861 census abstracts should also be considered. In this census occupations were classified using an extended schema based upon that devised for use in the 1851 census, and in total some
425 occupational categories were distinguished on the basis of the materials used by those employed in each category. Higgs has argued that this reflected both the concerns of the Registrar General regarding the health of the population and a belief, at this time, that many illnesses resulted from chemicals contained within the products and other materials with which people worked.\textsuperscript{12}

Several problems faced the Census Office in compiling the occupational tables presented in the published reports, mainly stemming from the inexact nature of the information given in the CEBs. These problems can be summarised under three main heads.\textsuperscript{13} First, job titles were sometimes only stated vaguely, with little or no information being given on either the industry of employment or the actual job undertaken. Second, it was often difficult to distinguish between dealers and makers. Third, although people were asked to say how many people — if any — they employed they did not always do so. However, neither this nor the previous concern would affect the actual numbers given in the occupational tables. More fundamentally, Higgs has questioned the completeness of the occupational tables relating to women, arguing that the occupations of many married women were not recorded in the census.\textsuperscript{14} Recently, however, Anderson has argued that the occupations of married women were nowhere near as likely to have been under-enumerated as Higgs has claimed. Using a 2 per cent sample of households for Lancashire drawn from the 1851 CEBs to examine the occupations of women disaggregated by age, marital and parental status and the occupations of husbands, Anderson has concluded that — although some under-enumeration of part-time occupations may have occurred — the occupations of the vast majority of women in full-time employment were almost certainly recorded.\textsuperscript{15} Moreover, it should also be added that the census authorities implicitly assumed that women married to farmers, innkeepers and husbands engaged in several other trades were assisting their husbands and classified them as such.

The least accurate data collected in the census was the information sought on disabilities (that is, the numbers of people who were blind and deaf and dumb). This question was poorly worded and the replies given have been little used. Higgs cites a study in Wales in which replies included 'unhealthy from birth', 'helpless' and 'not well'.\textsuperscript{16} Despite these limitations, it is likely that the tables presented in the census volumes provide a reasonably accurate picture of the distribution of the blind and deaf and dumb in the country as a whole, if not their absolute numbers.

Let us now consider the accuracy of the vital registration statistics. It is difficult to know how complete the registration of births was, especially in the decades immediately following the introduction of vital registration. Widespread illiteracy meant that many people were probably unaware of the registration process, and it was only in 1875 that the responsibility for registering births was transferred from registrars to parents. Teitelbaum has calculated that registration was 94 per cent complete in the period 1841–50.\textsuperscript{17} After this decade registration improved, and was almost complete by the 1880s. Teitelbaum did, however, find pronounced regional variations in his figures, the under-registration of births being most acute in Surrey, Sussex, Middlesex, Shropshire and North Wales. It
also seems probable that some illegitimate births would have gone unregistered.  

The process by which marriage took place, involving either a religious or civil ceremony in front of a minister or registrar and witnesses, makes it highly unlikely that many would have gone unregistered. Some of the information given on marriage certificates is, however, known to be subject to error. Occasionally minors falsified their ages when they were marrying without the consent of their parents. It should also be noted that the signatures on marriage certificates can only be regarded as, at best, a crude measure of literacy.

Because a death certificate was required for the disposal of a corpse, the registration of deaths was far more complete than that of births. In the first decades of registration more than 98 per cent of deaths appear to have been registered, most of the rest being those of young infants. Problems occur over the accuracy of the information relating to the causes of death. It was only in 1875 that doctors were required to issue certificates detailing causes of deaths, and before this date it is probable that some were incorrectly recorded.

Transcription and database creation

In undertaking this project, information relating to each of the 635 registration districts in England and Wales was keyed into a computer. The following information was included from the census abstracts:

a) the area of each district, numbers of inhabited and uninhabited dwellings, population totals and numbers of males and females;

b) the ages and marital status of the population;

c) population totals;

d) occupations of males and females aged 20 and over;

e) birthplaces of males and females;

f) disabilities and the institutionalised population.

The information included from the Registrar General's Annual Report for 1861, related to:

a) the numbers of births, marriages and deaths in each district;

b) marriage statistics (religion, marital status at time of marriage, literacy, etc.).

In addition, details of the numbers of deaths, causes of deaths and ages at death (both male and female) were extracted from the Registrar General's Decennial Supplement for 1861. Finally, the Poor Law statistics which have been entered into the databases are those for July 1861 which were submitted to Parliament approximately three months after that year's census. These show:

a) the numbers in receipt of both indoor and outdoor relief;

b) the numbers of lunatics, in each Poor Law Union.
Accuracy of the database

Throughout the exercise, extreme care was taken to ensure accurate transcription of the data. This was done by summing the totals for each table both vertically – to ensure that the total in each column was equal to the sum at the head or foot – and horizontally – to ensure that the row totals summed up to the total figures for each county. Such a checking procedure was adopted to overcome the problem of compensating errors which cancel one another out.

Accuracy of the compiled sources

An examination of the accuracy of the data contained in the various sources used for this study leads to the overall conclusion that the compilers of the census abstracts achieved a remarkably high degree of accuracy. Very few errors were found in the column totals, and those which were found could usually be traced to printing errors in which the relevant cells in the table were left blank. These were easy to spot and have been corrected. A similar level of accuracy was also found in the Registrar General's Annual Report for 1861. Unfortunately, however, the Registrar General's Decennial Supplement containing details of the ages at and causes of death was found to be somewhat less accurate.

Value of the database to local historians

First, the database brings together material for 1861 which is difficult and time-consuming to obtain because paper copies of the records are not held in all libraries. Moreover, the data have been arranged in an electronic retrieval system in such a way as to make it relatively simple to both read with most database packages, such as Microsoft Works and Access, DBase and Paradox, and to manipulate and generate statistics (e.g. to calculate percentages, proportions, death rates, birth rates, etc.) using a spreadsheet package such as Excel and other statistical packages such as SECOS and the Statistical Package for the Social Sciences (SPSS).

Second, the database is a useful descriptive tool which allows the researcher to place his or her area of interest into its regional or national context, so making it possible to say to what extent area X might differ from area Y or be atypical of either the region or the country as a whole according to any number of factors, including inward migration, occupational structure, illegitimacy rate, etc. Again, this might take a simple form in which areas are compared on the basis of just one or two factors, or more complex statistical tests might be undertaken based upon factor analysis. For example, Stephens, in his study of literacy of the mid-nineteenth century, used the marriage tables in the Registrar General’s Annual Report to look at how the level of literacy varied between registration districts in this period.21

Another example of how a particular study area might be placed in its national context is given in Figure 1, which shows the numbers of females to every 1,000 males aged 15–29 years in Shrewsbury and the surrounding registration districts.
From this we see that, whilst in Shrewsbury there was a surplus of young women, in Atcham and the surrounding districts there was a shortage of them.

These differences might be explained by the migration of young women from the surrounding countryside into Shrewsbury – which was a market town with a relatively large middle class – to work as domestic servants. A similar situation has been found in a number of other market towns in this period and, indeed, was noted by Ravenstein who studied the processes of migration in the late nineteenth century.22

Third, it is hoped that the database will encourage the standardisation of coding schemes and practices amongst those using the CEBs. Although a considerable amount of worthwhile and important work is being undertaken on the CEBs, researchers tend to work in isolation from one another and often develop coding schemes for occupations and birthplaces which, although, no doubt, excellent for their own purposes, are not readily comparable with those used by other researchers working elsewhere in Britain.23 Much useful work has been done on life and conditions in the nineteenth century in spite of this problem. But it has meant that much recoding and re-ordering of census data has been necessary for comparisons to be drawn between studies undertaken in different parts of the country. One of the advantages of the 1861 database is that the occupational categories have been classified using a modified version of the Booth-Armstrong categories. This schema was originally devised by Booth in the 1880s and it is widely used by other researchers following the work of Armstrong, who advocated its use in an article published in 1972.24 The main advantages in using this schema relate to the facts that: it is widely used by other researchers; a
coding book exists which greatly simplifies its use; and both Booth and Armstrong have published tables which allow researchers to view their findings in a national framework. One limitation of this schema is that farmers’ wives, innkeepers’ wives, and other wives assumed to be assisting their husbands are re-classified as non-productive workers. This has been reversed in this project.

More importantly, however, birthplaces can be coded according to registration district. To facilitate this, a small file has been created for use as a linked or relational database. This file contains details of a grid reference for a single point within each registration district, together with details of the proportions of men and women employed in a number of key industries. The advantage of using such a linked database to code birthplaces lies in the ability which it gives to append details of his or her birthplace to each individual’s census entry in the CEBs. This has two principal applications:

a) To calculate the distances travelled by migrants into a particular area. Since the linked database contains details both of the registration county and the grid references, it is very simple to calculate how far incomers into a study area have moved.

b) An examination of the main socio-economic characteristics of the birthplaces of migrants. The database could be used, for example, to study the extent to which migrants to a large ‘iron and steel’ town came from other iron or steel towns, or the extent to which ‘long-distance’ migrants tended to originate from towns and cities, or domestic servants from country districts.

Unfortunately, the linked database cannot be used to study migration within a single registration district. This is important because each registration district contained on average approximately 28 parishes, and it seems likely that a considerable amount of migration in the mid-Victorian period have taken place within registration districts. Nevertheless, the database can be used to study migration between registration districts, as has demonstrated in a recent paper on the study of migration into Hanley in 1851.

A note of caution is necessary here, as the information on birthplaces contained within CEBs does not usually give information on the Registration District of birth – the researcher will have to allocate appropriate codes to individuals’ birthplaces for themselves.

Fourth, the 1861 database is of use in aiding the processes of sampling when undertaking large-scale studies based on the CEBs. Studies based on random or systematic samples of CEBs are always subject to a degree of error, and the database is of value insofar as it provides the researcher with a means of calculating the degree to which the sample is representative of the population as a whole. For example, were a registration district to have a population of 20,000 of whom 1,000 were aged under five years and 2,300 were employed as coal miners, a representative sample of one-in-ten people or households should contain approximately 100 children under five and 230 coal miners.
Fifth, the database might be used to seek explanations for phenomena at the macro level of analysis. This refers to looking at a variation in a factor or phenomenon (such as the rate of illegitimacy or the numbers dying from infectious diseases) and trying to explain it in terms of other factors which vary between registration districts. This was the approach taken by Woods and Woodward who looked at life expectancy and infant mortality for the early 1860s, using a subset of information taken from the census and the Registrar General’s Annual Reports. They found that deaths from infectious diseases and infant mortality tended to be higher in the more urban and densely-populated districts of London, Liverpool, Birmingham and the other large cities than they were in the countryside. Reference was made earlier to work by Stephens, who studied literacy in the mid-nineteenth century by using the numbers in the Registrar General’s Annual Reports who ‘signed’ the marriage register with an ‘X’. The census database allows us to expand upon his approach by looking at those factors in each registration district which might be related to the level of literacy. One such factor, shown in Figure 2, is the number of children per teacher in each registration district. From this Figure we see that there is a marked and highly significant relationship between the two factors, such that where there were
relatively few children per teacher literacy amongst brides was high and where there were many children per teacher literacy amongst brides was low.

Finally, and directly related to the previous point, a limitation of much work done on the CEBs concerns the factors that lead people to study particular places. This is important, because all too often researchers will work on a particular area either because it happens to be the place in which they were born or in which they live, without any clear idea as to the potential value of the data and without the intent of testing a particular theory or hypothesis. One of the major advantages of the 1861 database is that not only can it be used to generate hypotheses, but it also directs researchers to particular areas of the country where those hypotheses might be profitably tested at the more local level by using a variety of sources. For example, the database might be used to identify areas of high and low literacy with a view to undertaking detailed local studies of precisely what factors (the numbers of schools, religion, child employment, etc.) might account for the observed differences.

Conclusion

In conclusion, this paper has looked at the creation of a machine-readable version of the 1861 census abstracts, vital registration statistics and the returns of the Poor Law Commissioners. It has been argued that the database has many potential uses for both the academic researcher and the local historian. In particular, researchers might find it to be a useful tool in the testing of hypotheses and in the directing of research to particular localities. It also contains a great deal of information which will be of use to local historians, and it is hoped that it will help encourage the standardisation of coding practices. A small linked database has also been created which provided grid reference coding and other summary information for registration districts.

In order to enable other researchers to make use of the 1861 census database, a subset of 114 variables has been made available through the Local Population History Book Club. The variables include details of the population, age structure, occupations, birthplaces and causes of death of both males and females in each of the 635 registration districts into which England and Wales were divided in 1861. Also included with the database is a User Guide to the 1861 Census Data Base in booklet form and a copy of the linked database.

An enlarged version of the database, including information from the 1831, 1851 and 1871 censuses and Pigot's County Atlas, together with a mapping facility, will hopefully be available on CD ROM sometime in 1997 or 1998. Readers might, however, like to purchase a pilot computer program including variables and a map (see Figure 1) for the counties of Lancashire, Cheshire, Staffordshire and Shropshire, entitled SECOS–3: An Integrated Approach to the Study of British Nineteenth-Century History, price £6.00. Cheques for this second publication should be made payable to Staffordshire University and sent to the author, c/o The School of Social Sciences, Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE. 
NOTES


3. Divisions were roughly analogous to the standard regions of the late-twentieth century. Each consisted of a number of counties, but London and Yorkshire constituted divisions themselves.

4. However, detailed statistics of ages, populations, etc., were given for a select number of large towns and cities.


6. *Population tables. Vol. I. Numbers and distribution of the People, with index to Names of Places*, BPP 1862 L, 911–. *Population tables. Vol. II. Ages, civil condition, occupations and birthplaces; and inmates of workhouses etc.* (in two parts), BPP 1863 LIII, pt I, 265–, pt II, 1–. Other reports relating to the 1861 census of England and Wales include; *Tables of population and houses enumerated in England and Wales and in islands in British seas on 8 April, 1861*, BPP 1861 L, 855–, *General Report; with appendix of tables*, BPP 1863 LIII, 1–.

7. The term ‘lunatic’ was that used by the Poor Law Commissioners.


10. Tables showing the counties of birth of individuals living in each ancient county are, however, included in the census abstracts.


23. See Mills and Schürer, ‘Employment and occupations’.


27. For example, farmers’ wives have been classified as agricultural workers.


30. This refers to the total number of children aged under fifteen years in the area to every teacher, rather than to the number of scholars or pupils at school per teacher.

31. Details of the Local Population History Book Club are given in the *News from the Local Population Studies Society* of this issue of the journal.

32. Thanks are due to Leonie Simpson who helped create the database, and Peter Franklin who kindly read through various drafts of this paper and whose comments helped to improve it.