Male occupational data from English parish registers c.1700-1820 (processed summary data)

Note: Please note that there are two data collections of male occupational data from English parish registers deposited with AHDS History. This collection (SN 5398) is of processed summary data. The other, *Male occupational data from English parish registers c.1700-1820 (raw data)* **(SN 5397), is of the raw or unprocessed material as it was collected.**

Contents

This data collection consists of 23 files. In most cases each file pertains to one of England's ancient counties.[1] In every case the file contains male occupational data abstracted from virtually all the surviving Anglican baptism registers for that county for the period 1813-20.[2] A list of the counties covered and the number of baptism registers from which occupational data have been collected can be found in Table 1.

Table 1 Data abstractions for 1813-20

County	Number of Rose Act abstractions (1813-20)
Bedfordshire	129
Buckinghamshire	204
Cambridgeshire	171
Cheshire	132
Cornwall	206
Devonshire	460
Durham	90
Hertfordshire	132
Huntingdonshire	95
Lancashire	237
Leicestershire	256
Metropolis of London	169 #
Middlesex	53
Northamptonshire	295
Northumberland	110
Oxfordshire	232

---

[1] However, Yorkshire has been split into the three Ridings and the City and Ainsty of the city of York. Additionally, 'Metropolis of London' refers to the area defined as the Metropolis in the 1851 census and 'Middlesex refers to that part of the ancient county of Middlesex which did not fall within the 1851 census definition of the Metropolis.

[2] The small number of exceptions relate to cases where the incumbent has not deposited the records and the relevant archives do not hold microform copies or Bishops' Transcripts. For London alone the data derive from the single year 1817 rather than the longer period 1813-20.

Rutland	49
Westmorland	68
Worcestershire	206
City and Ainsty of the City of York	40
Yorkshire East Riding	197
Yorkshire North Riding	230
Yorkshire West Riding	303
Total	*4,064*

\# Data only recorded for 1817

In the case of the eight ancient counties listed in table 2 the files also contain occupational data abstracted from Anglican baptism registers for one, and sometimes two, eight-year periods between 1740 and 1799. This represents virtually all the surviving occupational data that met our data quality requirements (described below).[3]

These data are derived from the occupations of fathers recorded in Anglican parish registers at the baptism of a child. From 1813, under the terms of Rose's Act it was a legal requirement to record the fathers' occupation and practice almost universally conformed to the letter of the law. Hence virtually complete geographical coverage of Anglican baptism registers is possible for this period.

Table 2. Eighteenth century data abstractions

County	Parishes with c18th abstracted data
Bedfordshire	63
Cheshire	33
Durham	18
Lancashire	60
City and Ainsty of the City of York	35
Yorkshire East Riding	122
Yorkshire North Riding	96
Yorkshire West Riding	135
Total	*562*

Before 1813 the recording of occupations was not a legal requirement but some incumbents or parish clerks nevertheless chose to record occupations regularly. However, to locate suitable runs of data it was necessary to search all the surviving registers for the whole of the period between 1740 and 1799.[4] Table 2 indicates the number of baptism registers for which suitable runs of occupational data were located and abstracted.

Note

As noted above, there are two data collections (a processed and a raw version) of male occupational data from English parish registers. Most users simply seeking data on the occupational structure of particular areas will probably wish to use this collection of processed material. The other collection is likely to be of interest only to those

---

[3] Again the small number of exceptions relate to cases where the incumbent has not deposited the records and the relevant archives do not hold microform copies or Bishops' Transcripts.
[4] In the case of Bedfordshire the registers were searched for the period 1690 – 1799.

who wish to reconstruct the occupational structures of particular areas from the raw data for themselves or specifically to test or probe the way in which we have processed the data.

The processed summary material consists of counts of occupational descriptors by parish or chapelry whereas the unprocessed material consists of the individual uncounted events. It is important to stress that the raw data do not provide an entirely satisfactory picture of the male occupational structure. To produce the processed data the raw data had to be adjusted in four ways in addition to counting. Firstly, a number of codes used to increase the speed of data collection were removed. Secondly, the numbers of events were inflated for those parish registers that had been sampled by the sampling ratio (for details see fields 11 and 13 under *Structure* below). Thirdly, years in which occupational recording fell below 95 per cent were rejected (for details see field 14 under *Structure* below). Fourthly, abstractions were standardised to eight years in length by re-weighting them if they were shorter than eight years (for details see field 13 under *Structure* below).

Additionally it should be noted that the data for four counties (Buckinghamshire, Hertfordshire, London and Northamptonshire) exist in the processed form in this data collection but not in the unprocessed form in the other data collection. This is because these datasets, collected early in the project, used a different pen and paper-based form of original data collection. The remaining 19 datasets in this data collection were derived from the raw unprocessed data in the other data collection.

Provenance

Who created the data and why?

These data were collected as part of a research project run by Dr Leigh Shaw-Taylor and Professor E.A. Wrigley and funded by the Economic and Social Research Council: *Male occupational structure and economic growth in England 1750-1851* (RES-000-23-0131). The data were collected by a number of individuals employed on the project: Joseph Barker, Richard Churchley, Alec Corio, Dr Peter Kitson, Dr Amanda Jones, Dr Victoria Masten, Niraj Modha, Thomas Nutt, Dr Silvia Sovic, Stephen Thompson, Rebecca Tyler, Matthew Ward, Alison Warren and Matthew Westlake. Dr Peter Kitson undertook the quality control of the datasets to ensure that the data met with our requirement that the occupational recording should be 95 per cent or more complete. Dr Kitson then created the 23 datasets in this collection with assistance from Ms Gill Newton. All project members were affiliated to the Department of Geography at the University of Cambridge. The aim of this project was to reconstruct the evolution of England's male occupational structure from c.1750 to 1851. The underlying aim was to improve our understanding of the industrial revolution. The results of the project have not, at the time of writing, been published.

Further details can be found on the ESRC website where publications will be listed as they come out: http://www.esrc.ac.uk/. Much more detailed information, including unpublished papers, can be found on the project website:http://www-hpss.geog.cam.ac.uk/research/projects/occupations/

How were the data created?

A major element of the research project was the collection of male occupational data from Anglican baptism registers from the period c.1750-1820. We selected an eight

year period because this was on the one hand short enough to be viewed as a single moment in time and on the other hand long enough that any man in a fertile marriage would be likely to appear at least once in the baptism register. In a very few cases the new rules were not implemented immediately in 1813. In such cases the eight-year period begins a few years later. In other cases, where one or more years failed a quality test regarding the completeness of occupational recording (described below) extra years were collected.

The coverage is different for the two sub-periods period c.1750-1799 and 1813-1820 as follows:

1   1813-20

From 1st January 1813 it was a legal requirement under Rose's Act of 1812 for the occupations of fathers to be recorded at baptism. It is therefore possible to create parish level datasets of male occupations for every parish in a county from 1813 onwards. For the 23 counties in the datafiles the coverage of parishes and chapelries within the ancient counties is virtually complete (see footnote 2).

2   c.1750-1799

Before 1813 there was no legal requirement to record the fathers' occupation but some registers nevertheless did so. These were located by searching all surviving Anglican baptism registers for the nine counties enumerated in table 2 for the period 1740-1799 for suitable eight-year periods. For these counties virtually all surviving registers have been searched (see footnote 3).

For both periods it was a requirement that occupational recording was 95 per cent or more complete for the period of data abstraction. However, our measure of completeness was not quite straightforward.[5] Firstly, we excluded illegitimate births from our calculation. Secondly we excluded events for which the occupational description was illegible. Thirdly, for the pre 1800 period we discarded cases where the father was from outside the parish or chapelry, and when no occupational data was recorded. Finally, baptisms where no named parents were recorded were also excluded. We rejected runs of data where less than 95 per cent of legible legitimate baptisms of children did not record an occupation. Additionally, we also rejected runs of data which had both less than eight years of data and less than 200 events. To create runs of occupational recording which met these rules sometimes required collecting more than eight years of data and subsequently rejecting individual years. Each run of occupational recording has been termed an abstraction. This data collection excludes years that were rejected for failing to meet the data quality requirements. However, those rejected data are preserved in the related data collection, *Male occupational data from English parish registers c.1750-1820 (raw data)*.

Before data collection we created a complete list of all parishes and chapelries in each of the 23 ancient counties. These *Registration Units* were assigned a unique identifier and its archival location specified.

For four counties, Buckinghamshire, Hertfordshire, London and Northamptonshire, for which data collection was effected early in the project, research assistants collected the data using a pen and paper method and then aggregated the data onto a

---

[5] The account which follows is somewhat simplified. In due course we will publish a full account of the data collection rules.

single spreadsheet for each county. Dr Kitson created the four pertinent files in this deposit from those four spreadsheets.

For the other 19 counties research assistants visited the archives concerned with laptop computers. All the 1813-20 data were then entered direct into an Excel spreadsheet usually from microfilms of the original registers. Searches of the baptism registers for the period 1740-1799 were mostly undertaken using the microfilm or microfiche copies. For data abstraction the microform sources were usually too poor in quality for efficient abstraction. However, in most cases it proved possible to take digital photographs of the original source material. Abstraction then took place in Cambridge with the image on one computer screen and the data entry was performed on another computer, usually a laptop, into Excel files.

Dr Kitson designed the Microsoft Excel data entry form (an example of which is included below) and the research assistants passed the data to him and he then implemented the necessary data quality tests. Dr Kitson with the assistance from Gill Newton then created the 19 county level datasets that form the related collection, *Male occupational data from English parish registers c.1750-1820 (raw data).* Dr Kitson then created the remaining 19 files in the current data collection by implementing the data-quality requirements set out above and converting the individual events into counts of events.

Example of a Parish Register Abstraction Form

What sources were used to create the data collection?

The data collection was created from 4,064 Anglican baptism registers. Full source details are provided in the parallel data collection: *Male occupational data from English parish registers c.1750-1820 (raw data).*

Structure

Table 3 indicates, for each county, the name of the data file, the number of records in that file and the size of the file in kilobytes.

Table 3   Data Files

County	Name of file	Number of records	Size of file in Kbytes
Bedfordshire	BED County table.txt	3,633	364
Buckinghamshire	BUC County table.txt	3,608	363
Cambridgeshire	CAM County table.txt	3,596	388
Cheshire	CHR County table.txt	7,160	739
Cornwall	COR County table.txt	5,731	571
Devonshire	DEV County table.txt	11,926	1243
Durham	DUR County table.txt	4,790	495
Hertfordshire	HER County table.txt	3,550	369
Huntingdonshire	HNT County table.txt	1,646	172
Lancashire	LAN County table.txt	15,260	1920
Leicestershire	LEI County table.txt	5,151	542
London County	LON County table.txt	9,229	616
Middlesex	MID County table.txt	2,620	259
Northamptonshire	NHA County table.txt	5,093	520
Northumberland	NHB County table.txt	4,927	511
Oxfordshire	OXF County table.txt	4,824	502
Rutland	RUT County table.txt	722	70
Westmorland	WML County table.txt	1,548	156
Worcestershire	WOR County table.txt	5,379	572
City and Ainsty of the City of York	YCA County table.txt	3,232	402
Yorkshire East Riding	YER County table.txt	7,077	724
Yorkshire North Riding	YNR County table.txt	7,471	741
Yorkshire West Riding	YWR County table.txt	17,706	1795
Total		*135,879*	

Each file contains 16 fields as follows (the data type is indicated in brackets following the field name):

1   *ID (integer)*

   This is a unique ID for each line in the file.

2   *porID (string)*

   This field is a unique ID for each line of occupational data representing a single occupation during a single 'period of occupational recording' (see 4 below) for a single 'registration unit' (see 5 below).

3   *ruID (string)*

   This is a unique identifier (at national level) for each registering unit. The three-letter prefix refers to the ancient county.

4   *por (string)*

   This stands for 'period of occupational recording.' It takes one of three values: '1', '2' or 'rose.' Ideally we wanted one tabulation starting around 1750 (indicated by '1'), one starting around 1780 (indicated by '2') and one for 1813-20 (indicated by 'rose' – which alludes to the operation of Rose's act discussed above). In practice we took the nearest available set of years of good occupational recording. Bedfordshire is different and for this county alone por 1 runs from 1698 to 1749 while por 2 runs from 1750 to 1799. The

exact dates are recorded in the related data collection, *Male occupational data from English parish registers c.1750-1820 (raw data).*

5   *runame (string)*

This is the name of the entity to which the baptism register belongs. This may be either an ancient Anglican parish (AP), an Anglican chapelry (CH) or a non parochial Anglican unit (NP). For ease of reference these have been termed Registration Units. Each name used is unique at the county level but not necessarily so at national level.

6   *utype (string)*

This records whether the registration unit is an ancient Anglican parish (AP), an Anglican Chapelry (CH) or a non-parochial Anglican unit (NP).

7   *parish (string)*

This is the name of the parish to which the registration unit belongs. Where the registration unit is a parish this is identical to *runame* (5).

8   *window (string)*

This records the range or ranges of years covered by the 'por' in the format 'yyyy-yyyy' for a single run of years or the format 'yyyy-yyyy/yyyy-yyyy' where the por is made up of more than one series of years.

9   *numevents (integer)*

This records the total number of events (baptisms) for the abstraction.

10   *pcs (decimal – 2 decimal places)*

This records the 'percentage completeness score'. This is a measure of the incompleteness of occupational recording. Crudely it refers to the percentage of eligible events with no occupational descriptor. For a precise definition see field 14 below.

11   *sr (integer)*

Where Registration Units recorded over 2,000 baptism in the period 1813-20 samples were normally taken. This field records the sampling ratio 'sr'. This is the ratio of the number of events in the register to the number of events recorded in the data file. In the vast majority of cases 'sr' is 1 because the register was not sampled.

12   *numyears (integer)*

This is the number of years of data in the 'por.'

13   *iratio (decimal – 2 decimal places)*

This field records the inflation ratio. In cases where the registration unit was sampled (see field 11) or where a period of less than eight years' data was abstracted it is necessary to re-weight the data so that such registrations units are not under-represented when data are aggregated. The re-weighting factor has been termed the inflation ratio and is recorded in this field. It is calculated by the formula *iratio = 8/(sr*numyears)*. The *iratio* is used to calculate the values in adjn (16 below).

14   *occ (string)*

This field records the occupation ascribed to the father. In the case of illegitimate children no father was normally recorded. Such cases can normally be identified and in such cases 'illegitimate' has been entered in this field. No attempt was made to retain original spellings except in cases where it was not clear to the research assistant either what the correct spelling should be or what the occupation was.

A total of five non-occupational text strings can be found in this field. In the original data collection the prefixes 'aa' were used in some cases to speed up data collection. The prefixes have all been removed from this data collection.

illegitimate

> This means that the baptism was judged to be illegitimate. In the overwhelming majority of cases where the birth was illegitimate, no occupation was recorded because no father was recorded. On the rare occasions where occupation was recorded we did not abstract it in order to maintain consistency in the data.

other parish - no occupation

> Here, no occupation for the father was recorded at baptism but the father was recorded as being resident in another parish.

no occupation specified

> Here, no occupation was recorded for the father at baptism.

baptism without named parents

> This indicates that neither parents was named at the baptism. In consequences there was therefore no occupation for the father. In most cases this occurred because the individual being baptised was an adult. Where this individual was male an occupation was occasionally recorded, but for the sake of consistency we did not abstract such occupations. The total number of cases was very small.

unreadable

> This means that the entry was unreadable.

These particular codes were standardised so that the computer script that performed the quality control tests could readily identify these cases. The reason for recording these items was that we did not wish to use data where the occupational recording was less than 95 per cent complete. However, this was defined not as 95 per cent of all entries but as 95 per cent of those with an occupation or for which no occupation was specified ('no occupation specified'). In other words entries where the baptism was illegitimate, or where no occupation was specified but the father came from another parish, or where no parents were named or where the entry was unreadable were excluded from the calculation. The percentage without occupational recording, defined in this way is termed the percentage completeness score (pcs). The pcs is recorded in field 10.

The value of the percentage completeness score (pcs) is computed as follows:

$$\text{pcs} = 100 \times \frac{s}{N - (b + i + o + u)}$$

Where:

N is the total number of baptisms;
s is the total number of baptisms where no occupation was specified;
b is the total number of baptisms without named parents;
i is the total number of baptisms of illegitimate children;
o is the total number of baptisms of children from a different parish and no specified occupation;
u is the total number of baptisms where the occupation was unreadable.

15 *n (integer)*

This is a count of the number of baptisms with the occupation recorded in field 14 in the 'ru' during the 'por.'

16 *adjn (decimal – 2 decimal places)*

This is 'n' (see field 15) multiplied by 'iratio' (field 13). Whenever data from more than one 'ru' are being used this adjusted 'n' should be used not the 'n' recorded in field 15. For the explanation see field 13 above.