

Evaluating the 1881 Census Transcription: a Pilot Survey of Hertfordshire

Introduction

This survey arose from an invitation received to attend a workshop at the University of Essex held 17-18th September 1998, hosted by the staff of the History Data Service of the UK Data Archive, to discuss the uses of the 1881 census transcription which had been coordinated by the Genealogical Society of Utah (hereafter GSU) in alliance with the Federation of Family History Societies.¹ This invaluable resource, now held by the History Data Service, was being supplied on request to historians in datasets of various sizes to facilitate both teaching and research, and the main purpose of the meeting was to determine the needs of end-users with a view to discovering the most appropriate ways of developing and distributing the data.² Somewhat surprisingly, there appeared to be minimal concern about the quality of the transcription, despite the fact that there was significant anecdotal testimony that should have given historical researchers cause for concern. As the coordinator of an ongoing project to work with family historians on computerising the 1851 and 1891 Hertfordshire censuses, I had received numerous personal communications concerning the 1881

¹ The Genealogical Society of Utah is, of course, an arm of the Church of Jesus Christ of the Latter-day Saints.

² The data scrutinised here are: Genealogical Society of Utah, Federation of Family History Societies (1997), 1881 Census for England and Wales, the Channel Islands and the Isle of Man [computer file]. Colchester, Essex: History Data Service, UK Data Archive [distributor]. SN:3643. The data are available from the History Data Service as an enriched version (K. Schürer and M. Woollard (2002), 1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enriched Version) [computer file]. Genealogical Society of Utah, Federation of Family History Societies [original data producers]. Colchester, Essex: History Data Service, UK Data Archive [distributor]. SN:4519) and as a 5% sample (K. Schürer and M. Woollard (2002), National Sample of the 1881 Census of Great Britain. [computer file]. University of Essex, Genealogical Society of Utah [original data producers]. Colchester, Essex: History Data Service, UK Data Archive [distributor]. SN:4375). The enriched version is the one examined here.

transcript revealing a sequence of ‘horror stories’ regarding personal names so mangled as to be unidentifiable, non-existent addresses, missing information, mysterious place-names and various other shortcomings. Genealogical publications such as *Family Tree Magazine* move from the local to the general, including articles such as ‘1881 Index – Caution!’ by Geoff Riggs published in the January 1996 issue, as well as a chain of correspondence, usually highlighting individual mistakes, but sometimes suggesting that whole chunks of the census might be missing.³ Particular errors continue to be exposed, a recent example being the mysterious whereabouts of Karl Marx. A search by David Gatley of the CD-ROM supplied by the GSU apparently proved fruitless, until a personal communication revealed that in 1881 Marx lived next door to one Henry Goddard, the doorkeeper of the House of Lords, in Maitland Park Road in London. Goddard proved traceable, and next door to him lived one ‘Karl Wass’, aged 68, author of political economy, born in Germany. Further research revealed that this was indeed another example of poor transcription.⁴

These errors have been identified in the data distributed by the GSU despite the employment of a rigorous procedure for transcription and checking. That transcription involved the production of an ‘A’ transcription of the records by one individual, a ‘B’ transcription by a different transcriber, and a quality check by a third party who was to compare the two transcripts page by page, and consult the original where a discrepancy occurred to adjudicate on the correct entry. Where the photocopy relied

³ G.A. Eady, Letters, *Family Tree Magazine*, November 1996.

⁴ D.A. Gatley, ‘Poor Marx’, *Local Population Studies Society Newsletter*, 28 (March, 2001), pp. 7-8.

upon was unclear, both transcribers and checker has access to the data on microfilm, and were instructed to declare those cases in which they had recourse to the film copy. The number of hours spent by each transcriber and by the checker was recorded on a pro forma which was returned to the GSU who coordinated data input.⁵ What is being evaluated, therefore, is the end product of this process.

No transcript, particularly one of this size, is ever going to be perfect. The process of transcribing and computerising parts of the Hertfordshire censuses of 1851 and 1891 has given a personal insight into just how easy it is to make errors, especially when working from poor quality photocopies. It is for this reason that the Centre for Regional and Local History at the University of Hertfordshire employs a three stage checking process, which is very costly of time but deemed necessary if our data and the historical results derived from them are to carry confidence. In general, the family historians who do the vast majority of our transcribing and to whom we are deeply indebted, are excellent readers of nineteenth-century handwriting, and probably more experienced than many professional historians. They also often have an intimate familiarity with the local area, its place names, addresses, occupational composition and the range of family surnames to be found. But, it must be said, we have also discovered considerable variation in the quality of work sent in, and very occasionally have received transcriptions that could only be described as poor in the extreme.

⁵ Where film copies were particularly bad it was possible to submit the transcription to a team at the Public Record Office, which happened in 'perhaps three to five occasions' for Hertfordshire: personal communication from the coordinator of the Hertfordshire transcription.

Excessive familiarity can also, on occasion, prove problematic, leading to over-confidence and an insistence on ‘correcting’ the errors (particularly in the spelling of surnames) made by the original census enumerator. This personal experience was endorsed by the coordinator of the 1881 transcription for Hertfordshire, who informed me that at least one instance of ‘correction’ had indeed occurred in this case, despite the clear instructions in bold letters as item number one of the summary instructions to ‘transcribe what you see – not what you think’. Fortunately, she took the trouble to change these amendments back.⁶

In the light of the foregoing, it seemed both appropriate and responsible to attempt to put the 1881 transcription to the test, rather than simply to take its general accuracy for granted. To this end an agreement was negotiated between the History Data Service and the Centre for Regional and Local History at the University of Hertfordshire whereby the History Data Service would provide a sum of £4,324 towards the cost of employing a research assistant, who would work with the present writer on a pilot evaluation of the 1881 census transcript for Hertfordshire. By coincidence the Hertfordshire Family and Population History Society, many years ago, donated to the Centre the original photocopies from which the transcript for this county was made, so we had the opportunity to conduct a true test of the quality of transcription, as well as information on who transcribed what.⁷

⁶ Personal communication.

⁷ For obvious reasons no names will be mentioned in the following discussion.

The sample

Our first task was to identify an appropriate sample to check. The photocopies of the Hertfordshire 1881 census fill three drawers of a filing cabinet, covering 10 Superintendent Registrar's Districts, 26 Registrar's Districts and 323 Enumeration Districts (hereafter ED), incorporating a total population of approximately 203,000.⁸ It was decided to take one ED for checking from each of the 26 Registrar's Districts, which would give appropriate geographical coverage of the county and provide a sample of 8% of the county's EDs.

Two other criteria informed the selection of districts, the first of which was the desire to reflect the proportions of the 1881 population living in urban and rural areas, just in case any particular problems affected the transcripts for one or the other.⁹ The proportions derived from the 1881 Census Report for Hertfordshire suggested that 10 of the 26 EDs in the sample should be urban. The second criteria was the desire to sample EDs that has been transcribed by individuals with different levels of experience and commitment. It was felt probable that those who had transcribed a substantial number of EDs would be likely to be more committed and experienced

⁸ W. Page (ed.), *Victoria County History of Hertfordshire*, Vol. 4 (London, 1914, repr. Folkestone 1971), p. 235. The figure given here, taken from the 1881 Census Report, is 203,069: we have not checked this total against the number of individuals recorded in the transcription.

⁹ Urban households were often more complex in the sense that more sub-tenancies existed, which might be a cause of confusion. Urban occupational descriptions were also, of course, more varied, while long distance migrants (with more obscure birth places) were more common in town than countryside. It was also at least possible that there was a difference in educational standards between urban and rural enumerators. For a preliminary analysis of enumerators in Cornwall, which suggests we might explore variations in the quality of the work produced by different enumerators, see T. Arkell, 'Identifying the census enumerators – Cornwall in 1851', *Local Population Studies*, 53 (Autumn, 1994), pp. 70-5.

than those who had transcribed a smaller number. Hence the 89 transcribers were ranked according to the number of EDs transcribed, which allowed the identification of nine major transcribers who had completed 10-15 EDs each (covering 33% of the total number of EDs), 13 moderate transcribers who had completed 5-9 EDs (covering 26% of the total) and 67 minor transcribers who had completed 1-4 EDs (covering 41% of the total). To reflect this range, it was decided to use one ED that had been transcribed by each of the nine major transcribers, and eight or nine from each of the other two categories.

A systematic sample of 78 EDs was taken, three from each Registrar's District, and from this sub-set 26 EDs were selected to accord with the desired urban/rural composition and an adequate representation of major, moderate and minor transcribers. The chosen EDs ranged in size from a population of just 112 to a population of 1,014, and together incorporated 15,654 individual records, or 7.7% of the population of the county. Of the population of the chosen sample, major transcribers had been responsible for 37%, moderate transcribers for 27% and minor transcribers for 36%; 48% lived in towns, and 52% in rural areas.

The checking procedure

The checking procedure was very simple, though complicated by the fact that the one key item of information that the GSU did not require to be recorded was the

enumeration district, which formed the basis for our sample selection.¹⁰ The data was supplied by the History Data Service in substantial blocks, organised by Public Record Office Piece Numbers, in a delimited text format. These proved extremely easy to convert to a series of Excel spreadsheets, a procedure which worked smoothly whether the data was described in the conversion process as either comma or tab-separated. At this time the data contained a number of variables that were redundant from the point of view of checking the transcription, such as county code names, a household identifier, and ages in days, and these were removed so that the data resembled the original.¹¹ Discrete EDs were identified by reference to first and last entries found in the original Census Enumerators' Books, and removed to separate files. The data was then laid out as it appeared in the Census Enumerators' Books, formatted for ease of reading and comparison, and printed out.¹²

Checking simply involved a careful comparison of the print-outs with the photocopies that has formed the basis of the original transcription. Our selection of transcribers (major, moderate and minor) was made, a little arbitrarily, on the basis of the 'A' transcribers, although where this information is recorded in the relevant boxes on the pro forma it does appear that 'A' transcribers usually spent longer on their work than 'B' transcribers, while checkers almost invariably spent far less than both. The pro formas clearly show that the prescribed procedures had generally been followed.

¹⁰ The other key omission was schedule numbers.

¹¹ The household identifier, county codes and ages in days had been added by the GSU.

¹² It was not possible for us to check totals against the published census reports (to detect missing or doubly transcribed records), as the published reports only give parish totals and not totals for individual EDs. Both problems will, however, have been detected during the checking process.

What we have conducted, therefore, is a check on transcriptions that had been prepared according to a fairly rigorous procedure, which should themselves have been already checked, and which had been made in accordance with clearly laid out, printed instructions.¹³ It is possible, of course, that mistakes were made at the point of data entry rather than transcription, despite the fact that the GSU adopted a double entry procedure that should have obviated such errors. Where it is possible to propose that an error was probably made at the data entry stage rather than at the point of transcription, this is noted in the following discussion.

The checking process was laborious, and proved impossible to complete for some EDs from the paper copies alone due to the poor quality of the photocopied pages. A particular problem was posed by the check marks entered on the enumerators' books as they were processed at the General Register Office, which in some instances completely obscure the original data, particularly that relating to age. In these cases resort was had to the microfilms held at Hertfordshire Archives and Local Studies.¹⁴ Supplementary checks were made on the computerised data, by conducting quick visual searches to ensure that each individual's sex corresponded with appropriate entries in the name, relationship and condition columns, that ages conformed to expected parameters and that the list of counties contained no rogue entries.

¹³ The Genealogical Society of Utah, *How to transcribe the 1881 British census* (Salt Lake City, 1988).

¹⁴ We will return to the significance of such instances below.

Identified errors were divided into two categories, serious and minor, on the basis of whether or not they distorted the information upon which an historian would wish to rely. Serious errors included surnames that had been mis-read to the extent of being unidentifiable, incorrect occupations, ages, sex, condition, relationship to head of the household, clear errors in county or place of birth (often the most difficult category to read and adjudicate upon) and omission of information on disability.¹⁵ Minor errors included mis-read surnames that remained readily identifiable (such as 'Honshin' for 'Honchin', 'Phipp' for 'Phipps') errors in forenames, small errors in addresses, exclusion of additional information in another hand which was supposed to be entered in single brackets (notably 'dom' next to a nurse or gardener), exclusion of double brackets which were supposed to be entered against information which was crossed through, minor mis-spellings of place names which nevertheless remained identifiable, other small typographical errors, exclusion of 'British Subject' against those born overseas, and failure fully to transcribe an identified disability (such as 'deaf' for 'partially deaf').¹⁶

¹⁵ For many purposes incorrect surnames would not pose a problem for historians, but for those wishing to trace individuals from census to census, or conducting other forms of nominal linkage, such errors might prove an insuperable handicap. Serious errors under occupation were dominated by misrepresentation of 'housekeepers', discussed below. Other errors under this head include the transposition of a qualifying description from one individual to the next (hence we have a 'Muffin and crumpet maker', followed by a 'Baker straw plaiter', when the term baker is a qualification of the first entry; also the term straw manufacturer being applied erroneously to a straw plaiter). Complete misreading is rare, although two entries 'Scholar' were transcribed as 'Gardener'.

¹⁶ The additional information entered in another hand, commonly against occupations, provide clarification as to how the General Register Office classified occupations. Other examples include 'teacher' or 'dom' next to 'governess', 'M' (for market) next to 'gardener' and 'SMS' (Subsidiary Medical Service) next to 'nurse'. For a further discussion see M. Woollard, *The classification of occupations in the 1881 census of England and Wales* (Colchester, 1999), pp. 10-11.

In one case the failure to identify entries that had been crossed through by enclosing them in double brackets was treated as a serious error, and this concerned the description 'Housekeeper' under occupation. This term, when it had been employed by the enumerator (itself an erratic occurrence) was often, though not always, crossed through, presumably by staff at the General Register Office during the processing of the data. It is clear that some enumerators had entered this term against women who 'did the housework', rather than confining it to those who were employed as domestic servants. The confusion among the enumerators themselves is thus compounded in such instances, and the database will provide misleading information on numbers of domestic servants in general, and housekeepers in particular. Only in one parish did the exact role of such individuals remain clear, for there – rather eccentrically – the enumerator chose to identify these 'housewives' with terms such as 'Assists household', 'Household manager' and 'Attends to home affairs', and as these terms are much less likely to prove misleading the failure to provide double brackets where they are crossed through was here regarded as a minor error.¹⁷

In this instance the cause of confusion emanated from lack of a clear understanding on the part of the census enumerators, but there is one instance in which the instructions issued by the GSU are themselves the cause of confusion, which is in relation to the use of single brackets around transcribed data. As noted previously, data entered in a different hand was supposed to be enclosed in single round brackets, according to

¹⁷ The eccentricity of this helpful enumerator extends further: in the case of Sarah Salt, a 64-year old widow, we learn under the occupation heading that she had 'A few pounds in bank but depending principally on children'.

summary transcription instructions, but the instructions in the full booklet also read ‘Use round brackets to indicate that information on the transcription form has been transcribed differently than how it appears in the census schedule, or may be subject to other interpretation’, while round brackets used by the enumerators themselves were also to be retained.¹⁸ Round brackets could thus have appeared in the original enumerator’s book, could indicate changes that the transcriber has made for the purposes of clarification or could denote additional information entered in another hand.¹⁹ Not surprisingly, many minor errors – particularly under occupation – occur due to confusion over when to use round brackets, the most common error being the failure to use them at all, or their erratic use.

There is another category of information that appears to have caused confusion, and this is the double and single strokes that appear in the census enumerators’ books between names to delineate discrete houses, and discrete families that share the same house. They are commonly rendered as \\ (for houses) and \ (for families) in the original documents, and were to be rendered as // and / by transcribers in a separate column. If they had been omitted by the enumerator, they were to be entered by the transcriber, for ‘As schedule numbers are not transcribed, it is essential to indicate by the appropriate strokes the commencement of a new household or a new family, even if these strokes do not appear on the census schedule’.²⁰ It is quite clear in the case of the Hertfordshire sample evaluated here that considerable confusion arose from these

¹⁸ Genealogical Society of Utah, *How to transcribe the 1881 British census*, p. 4.

¹⁹ A common usage of round brackets is where the transcriber has attempted to clarify the relationship to head of household, notably by the addition of ‘(Head)’ after ‘Lodger’.

instructions, for these strokes are entered so erratically that it was decided there was little point in systematically analysing errors under this head. Historians wishing to identify discrete houses or to conduct analyses of family and household structure cannot rely upon these symbols as entered in the 1881 transcription, for Hertfordshire at least, any more than they can regularly rely upon them in the original enumerations. For analysis of the structure of the family and household, recourse must be had to the protocol suggested long ago by Michael Anderson, whereby a new household is deemed to commence with every entry 'Head' in the 'Relation to head of household' column, whether or not this corresponds to a new schedule number or is identified by single or double strokes.²¹

The pilot evaluation

In the tabulated data which follows, every effort has been made to prevent possible identification of individual transcribers. Hence no names are given, and EDs are identified only by our own selected number, not by the parish or town to which they belong. Errors are expressed as proportions of total numbers of entries, a procedure that is only entirely appropriate when data is regularly recorded against each individual, as indeed it is for most categories (name, age, sex and so on). By no means every individual, however, had a recorded occupation, and hence a separate analysis

²⁰ Genealogical Society of Utah, *How to transcribe the 1881 British census*, p. 10.

²¹ M. Anderson, 'Standard tabulation procedures for the Census Enumerators' Books 1851-1891', in E.A. Wrigley (ed.), *Nineteenth-century society: essays in the use of quantitative methods for the study of social data* (Cambridge, 1972), pp. 136-7. Note also the exceptions listed on p. 137, which have informed the household analysis conducted for Hertfordshire from the 1851 census: N. Goose, *Population, economy and family structure in Hertfordshire in 1851. Vol. 2 St Albans and its region* (Hatfield, 2000), p. 146.

Table 1 : Number and Percentage of Errors in Hertfordshire 1881 Census Sample by Category

	Serious		Minor		Total		No. Records
	No.	%	No.	%	No.	%	
Address	15	0.10	305	1.95	320	2.04	15654
Surname	20	0.13	46	0.29	66	0.42	15654
Forename	0	0.00	43	0.27	43	0.27	15654
Relation	13	0.08	12	0.08	25	0.16	15654
Condition	6	0.04	13	0.08	19	0.12	15654
Sex	20	0.13	0	0.00	20	0.13	15654
Age	71	0.45	3	0.02	74	0.47	15654
Occupation	50	0.32	463	2.96	513	3.28	15654
County	3	0.02	45	0.29	48	0.31	15654
Place	14	0.09	36	0.23	50	0.32	15654
Disability	25	0.16	27	0.17	52	0.33	15654
Other	9	0.06	11	0.07	20	0.13	15654
Total	244	1.56	1004	6.41	1248	7.97	15654

Note: Total for Serious and Total errors does not equal the sum of the columns as two records exhibit errors in two categories.

is conducted below which relates the number of errors only to those for which this information is actually given. It is even more important to correct the analysis in this way for information on disability, for relatively few individuals had a recorded disability which could be the subject of transcriber (or inputter) error. Separate analysis of errors under this head are presented in Table 5 .

The overall results for the entire sample dataset of 15,654 records are presented in Table 1, and broken down into errors made by category of information. The most obvious feature of Table 1 is the *very* small number of serious errors that were identified. Serious errors were found in a mere 244 records out of the total of 15,654,

or just 1.56%. If we were to treat the information contained in each category of each record as a discrete piece of data, ignoring all blank entries in the original, the transcription would look even more impressive, exhibiting serious errors in a mere 244 cases out of some 187,000 entries, a mere 0.13%. Minor errors were considerably more common, affecting (additionally) a total of 1,004 records or 6.41% of the total. Combining serious and minor errors produces a figure of 1,248, approaching 8% of all records. It is, however, the serious errors that should be our main focus, given that they have been defined to include all mistakes that would produce incorrect historical data, whereas the minor errors are likely to be of only marginal interest to most historians, or of no interest at all. From an overall perspective, therefore, the Hertfordshire pilot evaluation must give us considerable confidence in the quality of the 1881 transcription, far more than one might have expected in view of the anecdotal evidence discussed above.

Two categories of information stand out as having relatively high levels of serious error, age and occupation, whereas for minor errors it is occupation and address that feature particularly prominently. Errors in age recording only amount to 0.45% of the total but, as we will shortly see, their distribution is significant, and they appear to have been the result of the use of poor quality photocopies where a clerk's check marks made ages difficult to read, allied to over-confidence on the part of the transcriber. Only two of the three EDs that were particularly difficult to decipher in this respect were, according to their pro formas, checked against the microfilm, while one records that it had been 'checked against [the] film but very feint in places'. What

is difficult to understand is the perpetuation of the errors in the other parish that was reportedly checked against the microfilm, for resort to the microfilm copies held at Hertfordshire Archives and Local Studies made the errors wholly transparent in the great majority of cases, the light of the microfilm reader usually rendering the data perfectly clear despite the clerk's check marks that had obscured it on the photocopies.

The reason for the relatively high (but still small) number of serious errors in the occupation data has already been discussed, and resulted largely from the confusion that affected both enumerator and transcriber alike concerning how to render the term 'housekeeper' when it had been crossed out in the enumerators' books. The reason for the relatively high figure for minor errors has also been rehearsed, and this time it is the failure to record information that was added in a different hand in single round brackets that is largely to blame. Under address it seems that problems may have occurred at the inputting stage, for many of the minor errors discovered here appear to be the result of addresses being copied over successive entries to which they do not pertain. Again, such mistakes are unlikely to worry most historians, whose interest will lie in the analysis of the economic, social and demographic features of this data. Under no heading, therefore, do serious errors affect as much as half of a percentage point of the total sample of records, while minor errors only affect in excess of 0.3% of the sample for two categories of information, and these are unlikely significantly to affect the value of the data for the purposes of historical analysis.

Table 2 : Number and Percentages of Errors in Hertfordshire 1881 Census Sample by Sub-Groups

	Serious		Minor		Total		No. Records
	No.	%	No.	%	No.	%	
Transcriber							
Major	116	2.02	255	4.44	371	6.46	5743
Moderate	67	1.59	256	6.09	323	7.68	4207
Minor	61	1.07	493	8.64	554	9.71	5704
Urban/Rural							
Urban	97	1.28	541	7.13	638	8.41	7585
Rural	147	1.82	463	5.74	610	7.56	8069
Total	244	1.56	1004	6.41	1248	7.97	15654

In Table 2 the total number of errors are broken down into two sub-sets, to provide information on two criteria that informed the selection of the sample of EDs: proportions urban and rural, and major/moderate/minor transcribers. No substantial differences emerged from comparison of urban and rural transcriptions: the former exhibit slightly fewer serious errors but slightly higher proportions of minor errors, to produce a roughly equivalent overall error rate. There were clearly no circumstances that adversely affected either the urban or the rural transcriptions to a disproportionate degree. Examination of the small differences that do exist by category of information suggests that they arose by chance, the higher proportion of serious errors in rural parishes reflecting the fact that they included both parishes where the ‘housekeeper’ problem was most marked, and two of the three that posed particular difficulties for interpretation of ages. A higher error rate in urban parishes with regard to place of

birth may have been the product of the higher levels of long distance migration commonly found in towns, rendering place names less familiar.

Turning to major, moderate and minor transcribers, it is interesting to find that major transcribers made proportionally more serious errors than did moderate transcribers, who in turn made more than minor transcribers, which is the reverse of what would be expected if experience and commitment were key influences on the quality of transcription. They clearly were not, for major transcribers made more serious errors than minor transcribers under every category except sex and occupation. This is unlikely to have been a chance result, and might possibly reflect excessive confidence, or perhaps 'transcription fatigue' as they laboured on through successive EDs. It is also possible that minor transcribers were able to concentrate upon those localities with which they were wholly familiar, whereas major transcribers were confronted with less familiar data. For minor errors the hierarchy is reversed, with minor transcribers making most and major transcribers least. The majority of these small errors occurred under the categories of address and occupation. In the case of address, the number of minor errors is greatly inflated by minor typographic or data input errors that affected a number of consecutive records in two parishes, and cannot be systematically explained. With regard to occupation, this result may well reflect more passing familiarity among minor transcribers with the transcription rules relating to information added to the enumerations in a different hand. In the light of these differences, it would appear that the decision to use volume of transcription as one of the criteria for selecting the sample of EDs to be checked is justified.

Table 3 : Number and Percentages of Errors in Hertfordshire 1881 Census Sample by Enumeration District

ED	Serious		Minor		Total		No. Records
	No.	%	No.	%	No.	%	
1	39	5.31	40	5.44	79	10.75	735
2	5	1.84	29	10.66	34	12.50	272
3	11	1.18	14	1.50	25	2.68	933
4	22	3.64	36	5.96	58	9.60	604
5	17	2.02	62	7.37	79	9.39	841
6	4	0.67	28	4.70	32	5.37	596
7	2	0.34	25	4.29	27	4.63	583
8	6	0.79	23	3.01	29	3.80	763
9	1	0.46	12	5.53	13	5.99	217
10	0	0.00	10	2.49	10	2.49	401
11	5	0.75	50	7.50	55	8.25	667
12	2	1.00	7	3.50	9	4.50	200
13	2	0.53	5	1.32	7	1.85	378
14	36	4.25	57	6.73	93	10.98	847
15	0	0.00	3	2.68	3	2.68	112
16	6	0.68	43	4.89	49	5.57	880
17	15	2.14	193	27.57	208	29.71	700
18	4	1.35	8	2.69	12	4.04	297
19	2	0.20	12	1.22	14	1.42	985
20	39	3.85	43	4.24	82	8.09	1014
21	4	0.65	22	3.60	26	4.26	611
22	0	0.00	17	7.08	17	7.08	240
23	9	1.12	126	15.73	135	16.85	801
24	3	0.64	69	14.81	72	15.45	466
25	10	0.99	61	6.03	71	7.02	1011
26	0	0.00	9	1.80	9	1.80	500
	244	1.56	1004	6.41	1248	7.97	15654

So far we have encountered no major worries about the 1881 census transcription: overall the error rate was low, and although one can identify fluctuations between categories of information and between major and minor transcribers, these lay within

fairly narrow bands. Table 3, which presents the data for errors arranged by ED, gives more cause for concern, for the transcription quality across the 26 EDs varied quite widely. Starting with total errors, the figures range from an error rate of just 1.42% in the case of ED 19, to a massive 29.71% for ED 17. For six of the 26 EDs the overall error rate exceeds 10%. In every ED, however, the majority of these were minor errors, and will be unlikely to worry the historian unduly. Furthermore, the excessively high number of minor errors in a few districts can often be explained: hence ED 17 was affected by two typographical/inputting errors where an address was mistakenly carried over to numerous consecutive entries, as well as by repeated omission of both single and double brackets in the occupation column. ED 23 was similarly affected by the former problem, ED 24 by the latter. More worrying are the variations in the proportions of records affected by serious errors. While in a majority of EDs (16 out of 26) the serious error rate stood at 1.00% or less, EDs 4, 20, 14 and 1 exhibited rates of 3.64%, 3.85%, 4.25% and 5.31% respectively. For these four districts, therefore, between 1 in every 19 and 1 in every 27 records contained a mis-transcription (or input error) in one variable, serious enough to distort the historical information provided. Of course, whether or not such error rates should be regarded as 'high' may be debatable: but if they are, then fully 15% of the transcriptions of this carefully selected sample of 26 EDs for Hertfordshire contained a high level of errors.

Table 4 offers a closer look at these four cases by breaking down serious errors into categories. The transcriber of ED 1 was clearly unfortunate in that a serious error occurred in an address, whereby the description 'Grocer and pork butcher shop' ran

Table 4 : Serious Errors in Four Enumeration Districts by Category

District	1	4	14	20	Total
Address	14	1	0	0	15
Surname	6	2	2	1	11
Forename	0	0	0	0	0
Relation	0	0	1	5	6
Condition	0	0	0	0	0
Sex	0	3	0	3	6
Age	13	4	27	19	63
Occupation	4	11	2	2	19
County	1	0	0	0	1
Place	0	0	0	6	6
Disability	0	0	3	2	5
Other	1	1	1	1	4
Total	39	22	36	39	136
Population	735	604	847	1014	3200

over incorrectly into 13 records, and this was most likely an inputting rather than a transcription error. Nevertheless, even if this were to be removed from consideration, the serious error rate remains as high as 3.40% in this district. What is quite clear is the impact of errors in the transcription of ages, which featured prominently in three out of the four EDs. As discussed above, these errors resulted from reliance upon poor quality photocopies, over-confidence and possibly also a failure to check carefully enough against the microfilm.²² In the other ED it is errors under occupation that dominate, and ten of the 11 discovered here involved the failure to identify housewives who had been incorrectly recorded as ‘housekeeper’, where the offending label had been crossed through. This was therefore an error in interpretation of the

²² See above, pp. 14-15.

rules rather than an error of transcription. ED 20, besides the large number of errors with regard to age, produced errors in eight out of the 12 categories included in Table 4. Although there are some indications of carelessness in this transcription (for example 'visitor' transcribed as 'other'), it is again the case that the photocopy relied upon was extremely difficult to decipher in places. Here the 'A' transcriber spent 13.5 hours, the 'B' transcriber 12.5 hours and – most unusually – the checker spent even longer at 14.5 hours. But although this transcript was apparently 'checked against film', many of the errors were, mysteriously, not identified.

It was noted above that while the manner of representing error rates employed here, as a proportion of the total number of records, is appropriate for most categories of information, it is not entirely appropriate for those categories where information is found irregularly – notably under occupation and disability. Occupations can be dealt with quite quickly, for an exercise in producing a revised error rate, which excluded all individuals for whom no occupation was recorded, produced only slightly higher proportions than those revealed in Table 1 above.²³ For serious errors the overall proportion rises from 0.32% to 0.43% and for minor errors from 2.96% to 3.98%, producing a revised total of 4.41% compared with 3.28%. Unsurprisingly, the increased proportions for individual EDs, particularly with regard to serious errors, were similarly modest.

²³ See p. 13.

Table 5 : Number and Percentage of Errors Under Disability

ED	No. Records	No. Disabled	Errors					
			Serious		Minor		Total	
			No.	%	No.	%	No.	%
1	735	2	0	0.00	0	0.00	0	0.00
2	272	1	0	0.00	1	100.00	1	100.00
3	933	15	4	26.67	5	33.33	9	60.00
4	604	2	0	0.00	0	0.00	0	0.00
5	841	6	3	50.00	2	33.33	5	83.33
6	596	4	0	0.00	2	50.00	2	50.00
7	583	3	0	0.00	1	33.33	1	33.33
8	763	4	1	25.00	0	0.00	1	25.00
9	217	2	0	0.00	0	0.00	0	0.00
10	401	0	0	0.00	0	0.00	0	0.00
11	667	4	0	0.00	1	25.00	1	25.00
12	200	2	2	100.00	0	0.00	2	100.00
13	378	2	0	0.00	0	0.00	0	0.00
14	847	6	3	50.00	1	16.67	4	66.67
15	112	0	0	0.00	0	0.00	0	0.00
16	880	0	0	0.00	0	0.00	0	0.00
17	700	22	5	22.73	9	40.91	14	63.64
18	297	3	2	66.67	0	0.00	2	66.67
19	985	5	0	0.00	2	40.00	2	40.00
20	1014	6	2	33.33	1	16.67	3	50.00
21	611	3	1	33.33	0	0.00	1	33.33
22	240	0	0	0.00	0	0.00	0	0.00
23	801	3	2	66.67	0	0.00	2	66.67
24	466	4	0	0.00	0	0.00	0	0.00
25	1011	2	0	0.00	2	100.00	2	100.00
26	500	1	0	0.00	0	0.00	0	0.00
	15654	102	25	24.51	27	26.47	52	50.98

The situation regarding disability is, however, a wholly different matter. The instructions to transcribers on what they were to do under this heading were not entirely clear, for the first sentence could be inconsistent with the remainder of the instruction: 'Transcribe the information exactly as it is given on the census. If the person has not been identified as being (1) Deaf-and-dumb, (2) Blind, (3) Imbecile or

Idiot, or (4) Lunatic, leave the box blank... Information should be written in this box only when a written description (i.e. blind) appears on the census schedule'.²⁴ In one instance the enumerator has entered a number rather than a description in this column, and thus our transcriber could be forgiven for omitting this information (though it is still omitted). In other cases the information entered differed from the categories listed in these instructions (for example, 'paralysed', or 'subject to fits'), and again strict adherence to the transcription instructions would explain why such cases were not recorded. But there are many other instances where there is no apparent explanation for the failure to record instances of blindness or the other disabilities listed above. In comparison to the total number of records the proportionate significance of these errors of either transcription or data input appear small, but their true significance is revealed in Table 5 where the number of errors is expressed as a proportion of the number recorded as disabled. In this case serious errors include all those cases where the information is simply missing from the transcription, while minor errors represent those cases where only part of the information is recorded (for instance, 'blind' rather than 'blind from birth', 'deaf' instead of 'partly deaf'). What Table 5 shows is that, for this Hertfordshire sample of districts, the information on disability to be found in the 1881 census transcription is wholly unreliable. In fully one-quarter of cases it is omitted completely, in another quarter the information is only partially recorded, producing an overall error rate of just over one half. Entries in the disability column in the original documents were found in 22 of these 26 districts: in 10 of these 22 one or

²⁴ Genealogical Society of Utah, *How to transcribe the 1881 British census*, p. 13.

more cases of disability were wholly omitted in the transcriptions, and hence just 12 – a bare majority – regularly record this information (if often partially). This is a pity, for the information on disability is possibly the most under-utilised of all the social data included in the census.²⁵ On the basis of the evidence presented here, however, this shortcoming could not be rectified with the aid of the 1881 transcript.²⁶

Conclusion

In conclusion, the limitations of this pilot evaluation must be rehearsed. It is based upon a sample of just 26 EDs, albeit a quite carefully selected sample. It is also based upon just one county, a small and essentially rural/small town county towards the south of England. Here the census transcription was organised by a highly efficient and committed family history society, and coordinated by an individual who is known to be both dedicated and meticulous. It seems clear from the documentation available in the form of the pro formas submitted to the GSU that the careful procedures laid down by them were generally followed. The transcription for Hertfordshire may, therefore, be a particularly good example, rather than a representative one. It must also be acknowledged that the error rate detected here will be a minimum one. As the experience of proof reading repeatedly demonstrates, it is impossible to find every

²⁵ D. Mills and K. Schürer (eds), *Local communities in the Victorian census enumerators' books* (Oxford, 1996), p. 9. One explanation of this is the fact that contemporary authorities placed little faith in the accuracy of information on disability by the 1880s: *1881 Census Report*, Parliamentary Papers 1883, LXXX, pp. 69, 71. A rare exception is E.A. Benjamin, 'Human afflictions: a study of the north Ceredigion census returns, 1851-1871', *Ceredigion*, vol. 10 (1985), pp. 155-60. Features such as the fact that 15 of the 22 with disabilities in ED 17 were to be found in the workhouse are particularly worthy of further examination.

²⁶ Preliminary testing by the History Data Service supports the conclusion reached here, with some 30% of the number of deaf, dumb and blind given in the published census report for Wales being absent from the database: personal communication, Matthew Woollard.

error no matter how much care is taken. It is doubtless the case that there are more errors in the Hertfordshire transcriptions than have been identified here.

These caveats entered, it is clear that – in terms of overall quality – the Hertfordshire transcription is remarkably good, and better than might have been expected given the more general anecdotal evidence on the 1881 transcript to the contrary. Again in overall terms, serious errors that would worry the historian are quite rare, small both in number and as a proportion of the total number of records affected, and virtually insignificant when broken down by category of information. This bodes well for analysis of this data at national, county and sub-district level currently in train at the University of Essex. The genealogist too should be encouraged by the results for surnames, for a mere 20 out of a total of 15,564 were found to be seriously mis-transcribed.²⁷

At the more local level, that of the enumeration district or parish, the results are more problematic.²⁸ In two senses the Hertfordshire transcriptions were found to resemble the proverbial curate's egg: a number of transcriptions were far worse than others, largely due to relatively poor transcription of ages, while one particular category of information – that on disability – was so badly recorded as to be of no use for many

²⁷ Credit must be given to the commentator who ventured the view in 1996 that 'the error rate for the completed project will be so small as to be entirely insignificant when considering the total Census': M. Woollard, 'Creating a machine-readable version of the 1881 census of England and Wales', in C. Harvey and J. Press, eds, *Databases in historical research* (London, 1996), pp. 98-101. Such a view can now be endorsed with some firm evidence to support it, although one must again allow the possibility of geographical variability.

²⁸ Analysis at parish level is also being conducted at the University of Essex.

parishes. Although these instances in relation to age are only found where photocopies used by transcribers were particularly poor, this variation remains surprising in view of the procedures adopted for the transcription, the amount of time apparently devoted to checking and the access that the transcribers had to microfilm copies. The problem the local historian faces as a result is simply that there is no way of knowing which parishes have been very carefully transcribed, and which will be subject to larger numbers of errors. The same problem faces the genealogist, for whom accurate information on age is probably second in importance only to that on surname. Information on disability cannot be used, and nor can the oblique strokes intended to distinguish separate houses and discrete families.

It would be wrong, however, to end on a negative note. Even in the enumeration district with the highest proportion of mis-transcribed ages the figure stands at only 3.19%. Given that some of these errors will cancel each other out, this level of error is unlikely seriously to distort a calculation of the district's age structure.²⁹ Nor will the (relatively) small number of errors regarding 'housekeepers' seriously distort an analysis of occupational structure at parish or enumeration district level; and even if it might in some cases impact upon a detailed study of domestic service, it must be remembered that the enumerators themselves dealt with such instances erratically.

The 1881 census transcription was a labour of massive proportions, conducted over a period of ten years, involving millions of (wo)man hours of labour that, if costed,

²⁹ Nor, of course, can we expect ages to have been wholly accurately recorded, no matter how well transcribed: Mills and Schürer (eds), *Local communities*, pp. 228-45

would run into many millions of pounds. It was a remarkable achievement, and while we must be grateful to the GSU for providing the initial impetus, the administration and for facilitating the data inputting, the true heroes are the thousands of volunteer family historians whose dedication and skill, in both transcription and data input, made the enterprise possible. The pilot evaluation of their work described here suggests that their dedication and skill has provided us, both historians and genealogists, with a transcription of an extremely high standard, and one that we can consult, with the few reservations rehearsed above, with considerable confidence.

© Nigel Goose
University of Hertfordshire
May 2002.