

How Good is the 1881 Census Transcription? The Results of a Pilot Evaluation for the County of Hertfordshire

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 It will, of course, be well known to readers of *Family Tree Magazine*, most of whom will probably have obtained their copy directly from the GSU. Somewhat surprisingly, there appeared to be minimal concern at this conference about the quality of the transcription, despite the fact that there was significant anecdotal testimony that should have given historical researchers cause for concern. These concerns have been voiced in the columns of various genealogical and family history society publications, not least in *Family Tree Magazine* itself, usually highlighting individual mistakes, but sometimes suggesting that whole chunks of the census might be missing.² 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 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,

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

² G.A. Eady, Letters, *Family Tree Magazine*, November 1996; G. Riggs, '1881 Index – Caution!', *Family Tree Magazine*, January 1996.

‘author of political economy’, born in Germany. Further research revealed that this was indeed another example of poor transcription.³

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 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.⁴ At the University of Hertfordshire we adopt a similar rigorous procedure in our transcriptions of the census data for 1851 and 1891 for the county, involving transcription and two subsequent checks.⁵ Such rigour notwithstanding, occasional errors still escape notice, as readers of our publications kindly inform us from time to time. No transcript, particularly one covering the entire country, is ever going to be perfect. The question is, just how good *in general* is the 1881 transcript, which forms such an important resource for both genealogists and historians? It was to establish this that the Centre for Regional and Local History at the University of Hertfordshire, with the help of financial support from the History Data Service at the University of

³ D.A. Gatley, ‘Poor Marx’, *Local Population Studies Society Newsletter*, 28 (March, 2001), pp. 7-8.

⁴ The Genealogical Society of Utah, *How to transcribe the 1881 British census* (Salt Lake City, 1988).

⁵ See N. Goose, *Population, economy and family structure in Hertfordshire in 1851. Vol. 1 the Berkhamsted region and Vol. 2 St Albans and its region* (Hatfield, 1996, 2000).

Essex, decided to conduct a pilot evaluation of the transcription for the county of 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 access to information on who had transcribed what.⁶

The population of Hertfordshire in 1881 was approximately 203,000, from which a sample was selected to reflect the geography of the county (one Enumeration District from each Registrar's District), the balance between town and countryside, and the commitment of the transcribers as measured by the number of Enumeration Districts (hereafter EDs) they had taken responsibility for.⁷ This systematic selection produced a sample of 26 EDs, forming 8% of the total in the county. The chosen districts 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 county's population. Of the population of the chosen sample, those whom we termed '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. Checking simply involved a careful, laborious comparison of the print-outs with the photocopies that has formed the basis of the original transcription. Where these photocopies were illegible,

⁶ For obvious reasons no names will be mentioned in the following discussion.

⁷ 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.

recourse was had to microfilm copies held at Hertfordshire Archives and Local Studies. Supplementary checks were made on the computerised data, by conducting quick visual on-screen 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. Identified errors were divided into two categories, serious or minor, on the basis of whether or not they distorted the information upon which an historian or genealogist would wish to rely.⁸

So, how good is the 1881 census transcription? The short answer is extremely good, far better than might have been expected in view of the anecdotal testimony to the contrary, though with a number of qualifications. Errors classified as 'serious' were found in a mere 244 records out of the total of 15,654, or just 1.6%. If we were to treat the information contained in each category of each record as a discrete piece of data, ignoring all blank entries, the transcription would look even more impressive, exhibiting serious errors in a mere 244 cases out of some 187,000 entries, a mere 0.1%. Minor errors were considerably more common, affecting (additionally) a total of 1,004 records or 6.4% 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 or genealogical data.

⁸ Fuller details, including full tabulation of results, will hopefully soon be available in an article submitted to the journal *Historical Research*.

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 0.5% of the total but, as we will shortly see, their distribution is significant. They appear usually to have been the result of the use of poor quality photocopies where a clerk's check marks made ages difficult to read, perhaps allied to over-confidence on the part of the transcriber. The reason for the relatively high (but still small) number of serious errors in the occupation data (0.3%) 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 main reason for the relatively high figure for minor errors under occupation (3.0%) is the failure to record information that was added in a different hand in single round brackets, as instructed by the GSU, and it may be that the instructions issued were not entirely clear in this respect. 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 to successive entries to which they do not pertain.

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'.⁹ It is quite clear in the case of the Hertfordshire sample evaluated here that considerable confusion arose from these 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.

On first impressions the number of errors made with respect to the disability column of the EDs is small, but this is only true when this number is expressed as a proportion of the total number of individuals in the sample. When expressed as a proportion of the total number 'at risk' (that is, those with a disability entered in the original documents), the error rate here comfortably exceeds that found under any other category of information. For the sample as a whole, in fully one-quarter of cases the disability is omitted completely, in another quarter the information is only partially recorded, producing an overall error rate of just over one half. No obvious explanation for this extraordinary error rate comes to mind, although it may be that partial recording (regarded here as a minor error) was again the product of unclear instructions.

⁹ Genealogical Society of Utah, *How to transcribe the 1881 British census*, p. 10.

Were there any significant variations between town and country, between category of transcriber, or between districts? Little in the way of urban-rural difference could be found, but it was interesting to note that major transcribers made proportionally more serious errors than did moderate transcribers, who in turn made more than minor transcribers, perhaps indicating that experience was outweighed by either over-confidence or 'transcription fatigue'. In the case of minor errors, however, the hierarchy is reversed, but in each case any differences lay with fairly narrow bounds. More worrying is the variation found between EDs. For serious errors the proportion of records affected varied from none at all to as many as 5.6%; and while in 16 of the 26 districts the error rate stood at 1.0% or less, in four it exceeded 3%. In three of these four districts this was due to poor transcription of ages where they had been difficult to read, in the other it was due largely to a failure to indicate where the term 'housekeeper' had been crossed through. The difficulty for the local historian or genealogist is, of course, that there is no way of knowing without checking which districts are particularly well transcribed, and which are less reliable.

The conclusion of this survey is that, in terms of overall quality, the Hertfordshire transcription of the 1881 census 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 or genealogist 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. The genealogist should be particularly encouraged by the results for surnames, for a mere 20 out of a total of 15,564 were found to be seriously mis-transcribed.¹⁰

At the more local level, that of the Enumeration District or parish, the results are more problematic. In two senses the Hertfordshire transcriptions were found to resemble the proverbial curate's egg: a number of transcriptions were considerably 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 at all for many districts. 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 rigorous procedures adopted for the transcription. 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 relied upon at all, and nor can the oblique strokes intended to distinguish separate houses and discrete families.

¹⁰ This result was foreseen in 1996 by Matthew Woollard: '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.

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 3.2%. For the local historian, 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.¹¹ Nor will the (relatively) small number of errors regarding 'housekeepers' seriously distort an analysis of occupational structure at parish or enumeration district level. For the genealogist, however, the variability in accuracy remains worrying, for individual level data is either right or it is wrong, and certain districts in the Hertfordshire sample potentially contain enough misinformation on ages to cause difficulties.

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, 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 general standard, and one that we can consult, with the few reservations rehearsed above, with considerable

¹¹ 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

confidence. One wonders if the 1901 index, transcribed through a Private Finance Initiative and employing large numbers of foreign nationals unfamiliar with census data, will prove as reliable.

© Nigel Goose
University of Hertfordshire
June 2002.