Lancashire Family History & Heraldry Society

Chorley Branch Education Group
Talk Handout

Further Steps
Reconstitution
(Partly extracted from Genuki.org.uk)

Recommended (Beginners) Book on Reconstitution

Nuts and Bolts (Second Edition), Family History Problem Solving through Family Reconstitution Techniques by Andrew Todd

This widely-acclaimed book describes how you to research wide -- i.e. look for brothers, sisters, cousins and the many networks to which your ancestors belonged. As well as helping you to understand their lives more fully, these little-known techniques should help to progress the direct line.

'*This book is excellent for a beginner. From the first page it is packed full of tried and tested methods with added illustrations to explain the points. Just what I wanted*'

Mrs CM, France

Recommended (Advanced) Books and Articles on Record Linking (Reconstitution)

1. "News from the Cambridge Group for the History of Population and Social Structure: Automatic record linking for family recognition," *Local Population Studies*, vol. 40, pp.10-16, 1988.
   A summary account of the work of the Centre. States that they can now perform fully-automatic family reconstitution. "We have tested the record-linking algorithms extensively against manual reconstructions, and we have found that they provide comparable or superior results in a fraction of the time taken to link the records by hand.." Provides a brief but useful narrative description of the techniques, but no references to publications giving detailed accounts of the techniques.

2. P. Adman, S.W. Baskerville and K.F. Beedham, "Computer-Assisted Record Linkage: or How Best to Optimize Links Without Generating Errors," History and Computing, vol. 4, no. 1, pp.2-15, 1994.
   "Record linkage is arguably a more complex process than most of its practitioners realize. Once a historian decides to settle for a subset of the potential links that can be made between two sets of nominal records s/he is already involved in the process of sampling. Fine, if that is relevant to the enquiry at hand: but if the objective is to achieve the maximum number of 'true' links, then accurate judgement becomes paramount. The key assumption underlying this paper is that 'judgement' is the function of the historian, not the computer. It is our belief that no systematic algorithm, however sophisticated, can perform this task as well as can an experienced team of researchers equipped with an appropriate set of software tools. In this paper we develop this line of reasoning in the context of our work on early eighteenth-century elections, and describe the functions of a software package known as CARL - Computer-Assisted Record Linkage."

3. R. Barker, "Reconstituting the Family," Genealogists' Magazine, vol. 21, no. 9, 1985. Based on a talk given at the Society of Genealogists. Gives a fairly full account of manual techniques for performing family reconstitution; in describing the work of the Cambridge Group for the History of Population and Social Structures mentions use of a computer only in connection with the analysis of the results of such a manual reconstitution. Then describes how genealogists could use essentially the same manual system of coloured record slips to analyze all the data they collect on

surnames of interest, arguing: "This systematic way of attacking genealogy leaves much less margin for error, in my view. You cancel out many births by means of infant mortality - and know you've done so. You can slot births in because you recognise the pattern of birth interval. A five-year gap which should contain a baby may be filled from another parish. You can keep your finger on the variations of surname. There is, however, one other important advantage. Genealogy is a very self-indulgent form of research. It benefits, in the last instance, no-one but yourself, because no-one sees it. . . Don't stop being genealogists. Become *responsible* genealogists. Become family historians. If you use the Cambridge system, then you are not only helping yourself, you are making a useful contribution to historical research. Write-up your findings; deposit your slips in the record Office for others to borrow. Let the Record Office microfilm your reconstitutions. . ." (An editorial note refers to the Local Population Studies Society, formed by the Cambridge Group, as a source of further information.)

4. G. Bouchard, "The Processing of Ambiguous Links in Computerised Family Reconstructions," Historical Methods, vol. 19, no. 1, pp.9-19, 1986.
    Follows on from Bouchard (1980), and gives a detailed account of the methods used. Starts by pointing out the (often minimized) problem of deciding which records are candidates for matching, stating that in fact this, and hence the whole matching procedure should depend on the characteristics of the data being matched. States that in this case the data being used (Saguenay registers for 1842-1951) was of sufficiently high quality to use nonprobabilistic methods of record matching to achieve progressive linking, starting with the most secure situations (involving couples) and dealing later with clusters. "During the operations, the accumulation of information in the family records allows more and more difficult decisions to be made: thus each step is based on the preceding one . . . from one step to another, the algorithms for matching and decision making are modified according to the nature and quantity of the information available in the family records and the complexity of the situations to be dealt with."

5. G. Bouchard, "Current Issues and New Prospects for Computerized Record Linkage in the Province of Quebec," Historical Methods, vol. 25, no. 2, pp.67-73, 1992.
    "Most of this article focuses on [work on a database of 665,000 baptism, marriage and death records for the period 1838-1971 in the Saguenay region of Quebec] in order to single out a few basic theoretical and methodological issues relating to automated record linkage. Some major lessons must be learned from our own accomplishments (and failures) and from the varying fates of the various well-known, large-scale automated record-linkage projects set up in the 1970s in the fields of history, demography and genetics. One of these lessons has to do with the close relationship between the nature and the quality of the data, the research goals, and the linkage strategies. The diversity of the pioneer projects has provided ample evidence of (1) the need to devise linkage systems consistent with the purposes of the research and the capacity of the data to support them; (2) the consequent contradiction that may arise between some ideal accuracy and efficiency levels; and (3) the difficult choice that has to be made in such circumstances."

6. G. Bouchard and C. Pouyez, "Name Variations and Computerised Record Linkage," Historical Methods, vol. 13, pp.119-125, 1980.
    An excellent account of some of the techniques developed, in connection with a large-scale parish register and census data linking project in Quebec, for matching data concerning couples (rather than individuals). Five different types of name variation are described: spelling variations, phonetic variations, double-names, double first names, and alternate first names, with details given on the techniques used for dealing with first two of these. The first was handled by a specially created

scheme of phonetic encoding for the French language, containing 64 rules - the article points out that Soundex is not really a phonetic encoding scheme, but rather just a crude sorting device. The second type was handled by a technique based mainly on Guth's scheme for assessing the extent to which two words contain the same letters in the same order. Another aspect of the scheme is that it builds up a dictionary of equivalent names, using separate criteria for isolated names, and names in the context of all the name data available for a couple. On one test, involving 2,000 records, the scheme succeeded in finding 98.5% of the possible links, whereas only two thirds of the possible links could be found if one tested simply for name identically.

7. C. Bourlet and J.-L. Minel. "A Declarative System for Setting Up a Prosoprographical Database," in History and Computing, ed. P. Denley and D. Hopkin, pp.186-191, Manchester Univ. Press, 1987.
    Brief description of a 70,000 item database of data from13th century French tax registers, and the use of a Prolog-based scheme for determining whether items refer to the same individual.

8. J. Carvalho. "Expert Systems and Community Reconstruction Studies," in History and Computing, II, ed. P. Denley, S. Fogelvik and C. Harvey, pp.97-102, Manchester Univ. Press, 1989.
    Brief description of a small-scale project that is to use a database, and a set of Prolog modules for record linking, life-story reconstruction, genealogical analysis and network reconstruction.

9. P.G. Cook, "Is Your John Cooke my John Cooke: introducing the "C-Vector" for finding common ancestors among databases.," Genealogical Computing, vol. 10, no. 1, pp.37-38, 1990.
    A heuristic for matching individuals in two lineage-linked databases, based on so-called C-vectors. Each vector corresponds to a separate individual, and lists in sequence the birth years of the individual, his/her parents and his/her grandparents, and the individual's sex, and name. (The scheme involves finding pairs of C-vectors which match on three or more known dates.)

10. P. Cooley, "Biographs for 19th Century Family Records," Computers in Genealogy, vol. 4, no. 3, pp.104-111, 1991.
    Shows how a biograph ("a potted life story on one individual", generated automatically from a PAF file) can be used together with some simple heuristics, to predict census returns, survivors, and electors, for checking against the actual lists.

11. P. Cooley, "Generation and Usage of Machine-Readable files from the GRO Indexes," Computers in Genealogy, vol. 4, no. 5, pp.190-199, 1992.
    Describes some techniques that have been developed for automatic enhancement of the output obtained from using a scanner to scan St. Catherine's House indexes, and describes the sort of test (making use of names and dates) that could be used to check whether two registrations refer to the same individual.

12. C. Davey. Reconstructing Local Population History: the Hatfield and Bobbingworth districts of Essex, 1550-1880, Cambridge Ph.D., 1990.

13. C. Davey and A.S. Jarvis, "Microcomputers for Microhistory: a database approach to the reconstruction of small English populations," History and Computing, vol. 2, no. 3, pp.187-193, 1990.
    Describes a scheme, based on the use of a relational database, for assisting local amateur historians to perform family reconstitutions. Baptisms, Marriages and Burials are kept in three base tables. "For every marriage, or implied marriage, all

associated baptisms, burials and remarriages are found. Links between entities are made by matching the character strings that represent names. In addition, certain logical, biological and social conventions [Wrigley, 1966] are observed which limit the number of possible links to only probable links." Possible links are kept in separate tables. Uses Soundex, plus a list of special cases, for matching names. The user is then provided on request with all the relevant information on screen in order that he/she can perform resolution of competing links manually.

14. G. De Brou and M. Olsen, "The Guth Algorithm and the Nominal Record Linkage of Multi-Ethnic Populations," Historical Methods, vol. 19, pp.20-24, 1986.
   Discussion of experiments comparing the letter-by-letter comparative Guth surname matching algorithm (Guth 1976) against language-specific surname matching algorithms, as Soundex, Henry and FONEM. "Although relatively succesful for the particular projects for which they were designed, each language-specific system suffers from a major drawback: Its effective application is limited to the language for which it was originally created. . . . [Guth's] algorithm does not depend on recognition of phonetic similarity . . . Our study suggests that Guth is correct to claim that her algorithm has an important advantage over other systems: It is able to identify variant spellings through the position of letters in names. Because of this, the Guth algorithm, unlike the language-specific programs, is well-suited to the linking of a mult-ethnic population. The algorithm, however, is not perfect. It can produce incorrect matches between surnames that bear little resemblance to one another. This problem becomes particularly acute when comparing short names, where one or two incorrect vowels can produce an incorrect match. One way of overcoming this would be to include a new function that would assign a level of 'confidence' in the link according to the number of letters that occur in the same positions of the names being compared."

15. G.J.A. Guth, "Surname Spellings and Computerized Record Linkage," Historical Methods Newsletter, vol. 10, pp.10-19, 1976.
   Provides a full description of a surname matching ("name-recognition") algorithm that takes account of letter ordering, rather than being phonetic in character, and its use in a project to link various sets of early 18th century Norwich records: poll books, land tax assessments, window-tax assessments and registers of freemen. Details are given of how linking was performed first of all within individual sets of data, then taking pairs of sets of data at a time. "The limited number of identifiers available in the Norwich data did not justify establishing an elaborate weighting scheme for making links automatically. However some percent of the 'true links' were formed by the program, enabling the historian to select the remaining links clerically."

16. J. Hitchon, "Russell Soundex Code: a BBC Basic program," Computers in Genealogy, vol. 1, no. 5, pp.122-123, 1983.

17. C.J. Jardine and A.D.J. MacFarlane. "Computer Input of Historical Records for Multi-Source Record Linkage," in Proc. 7th Int. Economic History Conf., pp.71-8, Edinburgh 2, 1978.
   A fascinating account of the means by which completely transcribed documents were marked up in order to make the individuals, property, etc., and the relationships that they refer to evident, ready for computer input.

18. S. King, "Record Linkage in a Protoindustrial Community," History and Computing, vol. 4, no. 1, pp.27-33, 1994.

19. A. Macfarlane, S. Harrison and C. Jardine. Reconstructing Historical Communities, Cambridge, Cambridge Univ. Press, 1977, 222 p. ISBN 0 521 21796 2
This fascinating book describes the development and use of a sophisticated manual technique for creating a set of inter-related indexes to data obtained from a variety of documents in order to perform a "total reconstitution" of a given small community. The technique is illustrated mainly by reference to a fourteen-year long study of the history of the villages of Earls Colne in Essex, and Kirby Lonsdale in Cumbria, covering the period 1500-1750. The twelve document types dealt with in detail are: Anglican parish registers, manorial rentals, court baron land transfers, frankpledge listings, court leek cases, church court cases, quarter sessions cases, assize depositions, hearth tax records, wills, probate inventories, and population listings. For each of these document types, details of the indexing techniques used are described. In the case of Earls Colne, whose population was about 1200, the parish registers alone resulted in about 30,000 index cards, and the total for the twelve sets of records was nearly 140,000 cards. The analyses that the integrated set of indexes make feasible are very impressive. Far more complete family reconstructions can be produced than would be possible using just the parish register data. (One example shows a family tree which includes 25 marriages only four of which could be traced from parish register entries alone.) Similarly, extremely detailed historical accounts of the ownership and land and property are made feasible. (For Earls Colne it is "possible to know who owned every one of the approximately 650 separate parcels of property, land or housing, at any point in the last four hundred years.") Moreover, much can be learnt about the accuracy and the completeness of the data contained in the various types of document, and indeed about what some of the documents really mean, so that the author claims that one can learn far more about a typical English community during the period covered, both on an individual basis, and as a whole, than had hitherto been thought possible. The final chapter, incidentally, contains a careful analysis of how the manual techniques described could be aided by computer processing (albeit written in terms of then-current large mainframe computers). The task of doing a complete family reconstitution project manually for a village of 1000 inhabitants over a period of 300 years, based solely on parish registers, is estimated at 1,500 hours. A "total reconstitution" of the type described of such a community is estimated as needing 10 to 20 man-years if performed manually, or about 3 to 4 man-years if aided by computer.

20. F. Nault and B. Desjardins. "Computers and Historical Demography: the reconstitution of the early Québec population," in History and Computing, II, ed. P. Denley, S. Fogelvik and C. Harvey, pp.143-148, Manchester Univ. Press, 1989. Description of a very large family reconstitution project; it is stated that the method used, though "often incorrect at the individual level . . . yields results that are statistically valid."

21. H.B. Newcombe and J.M. Kennedy, "Record Linkage: making maximum use of the discriminating power of identifying information," Comm. ACM, vol. 5, pp.563-565, 1962.

22. H. Rhodri Davies, "Automated Record Linkage of Census Enumerator's Books and Registration Data: Obstacles, Challenges and Solutions," History and Computing, vol. 4, no. 1, pp.16-26, 1994.
"In this paper the need for individual-level longitudinal data to study nineteenth century fertility and migration is highlighted and reasons why so little work has been carried out in this area using record linkage techniques are given. After hopefully alleviating the fears about data quality we discuss a computer package which can be used to speed up the linkage process."

23. R.S. Schofield, "The Standardisation of Names and the Automatic Linking of Historical Records," Annales de démographie historique, 1972.

24. K. Schürer. "Historical Demography, Social Structure and the Computer," in History and Computing, ed. P. Denley and D. Hopkin, pp.33-45, Manchester Univ. Press, 1987.
A general overview of the work of the Cambridge Group on computer-based record linking of parish register records, and of successive census records. Mention is made of the use of Soundex, and of standardized coding (e.g. of occupations), in establishing links, and of the programs for finding logical inconsistencies in the data. (35 refs.)

25. K. Schürer, J. Oeppen and R. Scofield. "Theory and Methodology: an example from historical demography," in History and Computing, II, ed. P. Denley, S. Fogelvik and C. Harvey, pp.130-142, Manchester Univ. Press, 1989.
Summary of the "Cambridge" approach to family reconstruction, which is described as based on two principles: "First an agnosticism about the number and identity of individuals whose life histories are represented by the events recorded in the registers leads us to form all possible links, and to take note of cases where records cannot be linked together. Second, a belief that in situations of uncertainty it is better to prefer links in which the information content of the records gives one the greatest confidence leads us to adopt a hierarchical, sequential approach to the resolution of ambiguities." It argues that this approach to deleting competing links is to be preferred to that by Skolnick, of searching "for the combination of links with the lowest aggregate total confidence score that would need to be deleted to resolve the network . . . [because] there can be a very large number of possible combinations of links that could be deleted."

26. M. Skolnick. "The Resolution of Ambiguities in Record Linkage," in Identifying People in the Past, ed. E. A. Wrigley, pp.102-127, London, Edward Arnold, 1973.
The method described is based on the estimation of maximum likelihoods, and applied to the linking of records from medieval Italian parish registers, for purposes of genetic research. "The method [uses] the decision making techniques developed in artificial intelligence projects . . . The artificial intelligence approach consists of building a family of related solutions, developing a method of keeping the family of reasonable solutions small, and selecting the best solution with minimum effort and maximum accuracy. . . The frequencies of the nominal identifiers in the records to be linked forms the basis of the likelihoods. If one has the distribution of each forename and surname by parish, by time period and by type of record, one can calculate the probability that a link between two records will be made by chance. Thus if a record has much missing data, and the only data which is compatible consists of common names, the probability of a match, made at random, being compatible is quite large. If there are many identifiers, and some of them are rare names, the probability of a compatible match occurring is quite low, and can be estimated. . . Likelihoods are also computed from age error distributions which are formed from records whose links are most certain. . . A heuristic computer program, LINK, is being developed for the resolution of ambiguities in record linking. It is being constructed in a manner similar to the heuristic DENDRAL computer program."

27. Squire, "Expert Systems in Genealogy," Computer in Genealogy, vol. 3, no. 8, 1990.
Briefly argues the potential utility of expert systems for various genealogical applications.

28. N.C. Stevenson. Genealogical Evidence; a guide to the standard of proof relating to pedigrees, ancestry, heirship and family history, Laguna Hills CA, Aegean Park Press,

1979, 233 p.
An excellent account by a lawyer and genealogist of the standards of proof that should be sought in establishing an individual's ancestry, whether for producing a documented genealogy, or for legal purposes. Though aimed at an American readership, and dealing mainly with American records and laws, it is well worthy of study by genealogists in other countries, particularly the UK given the links (both of emigration, and of legal heritage) between Britain and the USA. Quotes from many legal judgements, but nevertheless is very readable. Provides careful and thought-provoking analyses of the strengths and weaknesses, as genealogical evidence, of various types of official records (vital, court, land and census records), church and family bible records, newspaper files, monuments, etc., and published and private genealogies and genealogical directories - a number of which are roundly criticised, Burke's Peerage in particular. To quote from the introduction: " ... there are some who believe that the rules of evidence in our legal system and in effect in our court proceedings are too technical and not completely practical for genealogical, historical and biographical research. This belief is not valid." This is perhaps shown best by the excellent chapter on "Rules of Evidence Applied to Genealogy". The book is thus a great antidote to the motherhood statements about the need for careful research found in many guides to genealogy.

29. E.R. Swart, "A Computer Simulation of the Ineradicable Uncertainty in Genealogical Research," Family History, no. 118, pp.389-396, 1989.

30. M. Thaller. "Methods and Techniques of Historical Computation," in History and Computing, ed. P. Denley and D. Hopkin, pp.147-156, Manchester Univ. Press, 1987.
A general discussion of desirable characteristics and capabilities of database systems used for various types of "historical computation", including record linkage.

31. J.E. Vetter, J.R. Gonzalez and M.P. Gutman, "Computer-Assisted Record Linkage Using a Relational Database System," History and Computing, vol. 4, no. 1, pp.34-51, 1994.
"Our intent is not to present a theoretically or methodologically correct approach to nominal record linkage, but to describe the evolution of a semi-automated linkage technique particularly suited to the needs and resources of a university demographic research setting."

32. J.D. Willigan and K.A. Lynch. Sources and Methods of Historical Demography, New York, Academic Press, 1982.
A scholarly treatise on the whole subject of historical demography, but containing only a brief section on techniques for family reconstruction. This describes, as an example method, that of the Montreal group.

33. I. Winchester, "The Linkage of Historical Records by Man and Computer: techniques and problems," Journal of Interdisciplinary History, vol. 1, pp.107-124, 1970.

34. I. Winchester. "A Brief Survey of the Algorithmic, Mathematical and Philosophical Literature relevant to historical record linkage," in Identifying People in the Past, ed. E. A. Wrigley, pp.128-154, London, Edward Arnold, 1973.
"This survey consists of a discussion of the chief problems of record linkage which are relevant to historical data, followed by a select bibliography [of 54 items]. The discussion takes the form of a critical survey of literature about record linkage." An excellent account.

35. E.A. Wrigley, (Ed.). Family Reconstruction, An Introduction to English Historical Demography. 1966, 96-159 p.
Contains a very detailed account of a manual technique, using Family Reconstitution

Forms (FRF), for organizing parish register record linking pioneered by L. Henry. (The method to be used for resolving ambiguities is described in less detail, and would appear to be fairly subjective.)

36. E.A. Wrigley, (Ed.). Identifying People in the Past, London, Edward Arnold, 1973, 159 p.
From the introduction: "In recent years there has been a spate of historical studies involving nominal record linkage on a scale which requires the linkage rules to be set out formally and in detail. Most of these studies have been based on parish registers (or a comparable source of genealogical information), or on nineteenth century census schedules. And at the same time there has been a marked tendency to abandon manual methods in favour of computers. The six chapters which follow represent an attempt to describe some of the methods currently in use, and to discuss the problems and opportunities of record linkage work."

37. E.A. Wrigley and R.S. Schofield. "Nominal Record Linkage by Computer and the Logic of Family Reconstruction," in Identifying People in the Past, ed. E. A. Wrigley, pp.64-101, London, Edward Arnold, 1973.
Detailed account of the record linkage techniques used by the Cambridge group, including a section on Matchscoring and an Appendix on Name Spelling. The section on demographic constraints upon record linkage gives the following rules/heuristics:

"1. Age at death is never greater than 100 years, unless age information in the burial record overrides this rule.

2. At the birth of a child the mother is never less than 15 or more than 50 years old, nor the father less than 15 or more than 75.

3. No two successive birth events to the same mother occur in less than 10 months and no three successive birth events in less than 22 months.

4. The interval between the end of a marriage and the remarriage of the surviving spouse must be less than 20 years.

5. First marriages (for both sexes) occur only when the bride or the groom is above 15 years of age and less than 50, unless age information in the marriage record overrides the rule.

6. All brides and grooms are less than 75 years of age at marriage unless age information in the marriage record overrides the rule.

7. Whenever an age is given in either one or both of the records involved in a possible link the difference between the dates of the two records must be compatible with the age information given.

8. Where occupations are stated in both records involved in a possible link and they mismatch in a manner which is thought to be incompatible even with the most extreme assumptions about lifetime occupational mobility, no link is made (for example labourer/vicar).

9. In addition to the requirement that the names of the principal on two records should agree before a link is made, there must also be no disagreement about the names of any relatives named in both records (for example, if the names of both parents are recorded at the baptism of a child, and again when he is buried, they must not disagree if a BAP-BUR link is to be made)."

38. E.A. Wrigley and R.S. Schofield, (Ed.). The Population History of England, Cambridge Univ. Press, 1989.

Lancashire Family History & Heraldry Society

Chorley Research Centre at Astley Hall Farmhouse

Opening times
First & Third Saturdays Noon—4:00pm
Monday, Wednesday & Friday 10:00am—4:00pm

Last appointment 3:00pm

Booking advisable—Tel. 01257 231 600 (When centre is open),
Tel 01257 262 028 (When centre is closed)
Or
Book in line at - Chorley Family History Research Centre Website – www.cfhrc.com

Research Enquires - chorleyresearch@lffhs.org.uk

Chorley Branch Website - www.lfhhschorleybranch.com

LFHHS
President – Steve Williams
Society Website – www.lfhhs.org.uk
Reg. Charity Number - 513437