Family reconstitution in an urban context: some observations and methods

Gill Newton

University of Cambridge

ghn22@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Abstract

This paper concerns the application of methods for constructing family reconstitutions to early modern London parish register data, together with some observations on what types of demographic analysis may most successfully be performed in this context. It begins by considering the challenges presented by the urban environment, where population turnover is high and the spatial unit of analysis (the parish) will often have arbitrary boundaries not reinforced by topology. The steps of family reconstitution record linkage as best applied in an urban context are then defined and discussed. Ways of normalizing or shaping London parish register material into an appropriate input to those record linkage steps are described, concentrating in particular on name standardisation. Dates, ages, occupations and street addresses are also considered. The overall intention is to provide a practical guide to performing family reconstitutions of urban areas using computerised or computer-assisted methods.

Introduction

Family reconstitution can be thought of as a set of rules for linking historical parish records of baptism, marriage and burial into nuclear family groups, and a set of methods for measuring the demographic characteristics of these reconstituted families without bias. An overview of the evolution of the technique can be found in a companion piece to this paper recently published in *Local Population Studies*.[1] The overall aim is to arrive at a sample of the inhabitants of a parish for which it is possible to control for migration. Fortunately, this does not mean that those in the sample must have lived their whole lives in the parish of interest. What is important is that beginning and end dates can be set for the period in which each family must have been resident, according to the events they register, so that the size of the population at risk for any given demographic measure is known.

In this paper we shall begin by considering the formidable obstacles to demographic analysis posed by fluid urban populations, and what characteristics of parish baptism, burial and marriage registers are most useful in helping to overcome this. We consider how to evaluate supporting information that may be given, such as occupations or street addresses. We then describe the practical steps necessary to construct a family reconstitution through nominal record linkage, concentrating in detail on the pre-processing or normalization of input data that is necessary to achieve automated or semi-automated reconstitutions. This includes methods for standardising or classifying names, dates, ages, occupations and street addresses. Finally we evaluate the applicability and utility of these methods on different sizes of parish. Throughout the paper examples are drawn from family reconstitutions of large urban parishes, especially St Botolph Aldgate in the East End of London, and mainly over the period 1550 to 1750.

Measuring demographic characteristics in an urban population

The main obstacle to urban family reconstitution is migration. Cities typically experience exceptionally high levels of in-migration, and their inhabitants frequently change residence within the city boundary. The fluidity of urban populations impels caution in the construction of family reconstitutions of urban parishes, and caveats in the demographic analyses that may reasonably be performed.

Some city dwellers were born in the city in which they lived, and a section of the population persisted in one area for decades, but few adults in early modern London lived their entire life from birth to death in the same parish. Work opportunities, marriage, changing family size and changing rental values all provided inducements to move elsewhere at some point in the life course. This means that the prospects for tracing the full life cycle of most individuals in the baptism, marriage and burial registers of an urban parish are slim. However, the length of residence required for meaningful demographic analysis depends on the analysis being carried out. We can observe portions of inhabitants' lives as registered in the parish baptism, marriage and

---

[1] Newton G (2011): 'Recent developments in making family reconstitutions', *Local Population Studies*, 87, 84-89.

burial register. From the sum of those portions, an assessment of the characteristics of married inhabitants and their children can be made.

Some demographic analyses can be obtained from an urban family reconstitution with more confidence than others. Even where the absolute level of a measure is in some doubt, comparison between urban parishes can prove illuminating. Urban reconstitutions are best suited to the measure of mortality among the young. Measures of child and infant mortality and its seasonality require periods of residence of only a few years and can readily be obtained, whereas it is very difficult to obtain a representative sample for adult mortality and life expectancy.[2] Certain aspects of fertility and nuptiality may also be readily measured, such as the average interval between successive births and the age at first marriage. The latter must be approached with caution, since it rests upon record links that each span several decades, and can only be calculated for the London-born minority.[3] Details of how to obtain demographic measures from a family reconstitution have been described elsewhere and will not be reiterated here.[4]

Population movement in London and its effect on family reconstitution

Cities had high mortality rates in the early modern period, especially among children, and could not have sustained their vast populations without substantial levels of in-migration. London not only sustained its population throughout the period, but also grew rapidly, even at times when the national population was falling, although not all urban parishes grew at the same rate or the same time. Small, central London parishes such as those in the Cheapside area, where land and accommodation came at a premium, experienced stable numbers of inhabitants over much of the seventeenth century, even declining slightly after the last plague of 1665 and the Great Fire of 1666. In contrast, suburban parishes transformed during the seventeenth and eighteenth century into very densely populated areas, like Aldgate in the east, and the city end of Clerkenwell in the northwest. The flood of migrants into the suburbs creates problems in distinguishing between individuals in the parish register. It also means that most adult Londoners cannot be observed in the reconstitution from baptism onwards, but for a shorter portion of their life course, often beginning with their marriage or the birth of one of their children.

City dwellers changed residence frequently, albeit locally. Parishes within the walls of central London are very small, and in Cheapside any move of more than a couple of hundred yards would take a family into another parish and thus out of observation in the parish registers of the originating parish, and any family reconstitution based upon

---

[2] For a consideration of the difficulties of measuring adult mortality in family reconstitution studies in general, and how these may be overcome, see Appendix 6 pp. 581-600 in: Wrigley E A, Davies R S, Oeppen J E and Schofield R S (1997): English Population History from Family Reconstitution 1580-1837, Cambridge University Press, Cambridge.

[3] For a consideration of the difficulties of measuring age at first marriage in family reconstitution studies in general and the possible effect of high levels of population turnover on the observed age at marriage, see Wrigley E A (1994): 'The Effect of Migration on the Estimation of Marriage Age in Family Reconstitution Studies', *Population Studies*, 48, pp. 81-97.

[4] See Wrigley *et al* (1997) *op. cit.* and Wrigley E A (1996): 'Family reconstitution' in Wrigley E A (ed): *An Introduction to English Historical Demography from the Sixteenth to the Nineteenth Century*, Weidenfeld and Nicolson, London.

them. Extramural, suburban parishes usually encompass a much larger area, which reduces the potential for such cross-parish boundary moves. There may also be physical boundaries that reinforce the notional parish boundary and provide logistic constraints to local moves. The large suburb of Clerkenwell extends into the countryside, with the narrow Fleet River forming its western boundary. But the city end of the parish was by far the most densely settled, and here the parish boundary abuts the parish of St Sepulchre and cuts across both the main streets (Turnmill Street/Cow Cross and St John Street). Parts of the suburb of Aldgate are more strongly bounded in part by the physical features of London Wall and the wide River Thames, but its north and in particular its eastern boundary are not clearly delineated from other parishes.

Despite the lack of physical separation between parishes and the close proximity of other churches, in early modern London parish boundaries were recognised and adhered to by a surprising number of inhabitants. The majority of Londoners did register family baptisms and burials at their parish church. However, those who registered events outside the parish register may cause problems for family reconstitution, and the reasons for this are discussed below.

Religious nonconformity may result in families being absent from parish registers, or registering only burials. It is worth remembering that in urban areas nonconformity was often more widespread than elsewhere. In London, as well as nonconformist sects, there were the Stranger churches for the Dutch and French migrant communities, which allowed some groups to record their baptisms, burials and marriages wholly outside the parochial system. These communities of aliens were concentrated in particular parts of London, and Aldgate in particular had a sizeable community of Dutch and Flemish migrants, only some of whom used the parish church. While no family reconstitution based on Anglican parish registers can expect to capture the behaviour of nonconformists, records from other registration systems, such as those of the Quakers, can separately be used for family reconstitution studies to determine the demographic characteristics of those groups.[5]

Provided families do not register some of their baptisms, burials and marriage in the local parish and others elsewhere, the absence of particular groups from a parish register should not bias or affect the demographic characteristics of other families whose events are recorded therein. More problematic for family reconstitution are the range of opportunities cities offered for recording events extra-parochially that even Anglicans might use. In London, this was especially the case at marriage, but opportunities for burial outside the parish also existed.

Any marriage entails two parties who may come from different parishes, but the marriage will be recorded in only one parish register unless records of the banns that may have preceded it have survived. In a city comprising hundreds of parishes in close proximity to each other, the potential for the bride and groom to reside in different parishes is high. In sixteenth and seventeenth century London, considerable difficulty in observing inhabitants' marriages arises because of the popularity of marrying outside the home parish of either party. Before Hardwicke's Act in 1753,

---

[5] For family reconstitution work on the London Quakers see Chapter 4 pp. 131-161 in Landers J (1993): *Death and the metropolis: studies in the demographic history of London 1670-1830*, Cambridge University Press, Cambridge.

marriage licences enabling a couple to marry at the church of their choice were easily and cheaply obtainable. There was also a succession of 'clandestine' marriage centres where couples could marry outside their home parish without a licence. Clandestine marriage was at its most prevalent in London in the first half of the eighteenth century, when at times the majority of all Londoners married at the Fleet.[6]

The clandestine centres kept their own registers that identify the originating parish of both parties, but searching them or the registers of other parishes entails extra work. For example, data compiled for a family reconstitution of the suburb of Clerkenwell, which extends throughout the first half of the eighteenth century, included exhaustive searches of the registers of the Fleet and the earlier clandestine marriage centre of Holy Trinity Minories. Searches were also made for marriages licences of parishioners and of their marriages in all other London and neighbouring Middlesex parishes for which printed transcriptions of the marriage register were available. For the family reconstitution of the suburb of Aldgate, all marriage records of the earlier clandestine marriage centre of Holy Trinity Minories were input, since it forms an enclave within the parish.

The effect on family reconstitution of events belonging to families otherwise present in the parish registers being recorded elsewhere varies depending on the type of event that is missing and the demographic analysis being performed. While it is desirable to have the marriage records of as many parishioners as possible, marriage records are less essential than baptisms and burials to obtaining useful results from a family reconstitution, particularly if early age mortality is the main focus of investigations. Average age at first marriage can still be measured providing a large sample of parishioner marriages can be linked to other records, although there is a risk that such home parish marriages alone will not reflect the typical demographic behaviour of inhabitants. This can be mitigated by gathering additional records from other parish and extra-parochial sources. In constructing an urban family reconstitution, marriage records are also the least informative for the kinds of record linkage that can most confidently be made.

Burials of parishioners that took place outside the parish are uncommon but nonetheless a possibility in urban areas. In the wealthier central areas, children might be sent out of the parish to nurse, and subsequently die and have their burial recorded in other parishes.[7] Parishioners could also be sent out of the parish after death. While inter-parochial movement of corpses was comparatively well registered, the burial usually appears only in the destination parish rather than the parish of origin.[8] Extra-parochial burial grounds existed, often originating as sixteenth and seventeenth century plague burial pits on land that was privately owned, such as Tindall's ground and Bunhill Fields. Burials in such grounds were not usually recorded in parish registers or in any extant records until the eighteenth century, when the extra-

---

[6] See Boulton J (1991): 'Itching after private marryings? Marriage customs in seventeenth-century London', *London Journal* 16, 1 (1991), 15-34.
[7] Newton G (2010): 'Infant Mortality Variations, Feeding Practices and Social Status in London between 1550 and 1750', *Social History of Medicine*, doi: 10.1093/shm/hkq042
[8] See pages 247-50 in Boulton J (1992): 'The Marriage Duty Act and parochial registration in London, 1695-1706', in Schurer K and Arkell T (eds) *Surveying the people: the interpretation and use of document sources for the study of population in the later seventeenth century*, Leopard's Head Press, Oxford.

parochial burial grounds became especially associated with religious nonconformity. Early eighteenth century burials records at Bunhill fields reveal several burials of families in Clerkenwell who recorded other events at the parish church. These burials, while few in number, were unpredictable, encompassing adults, children, high and low status alike.

Missing burials of children whose families are otherwise recording events in the parish register are a much more serious obstacle to family reconstitution than missing marriages. Burials that were recorded elsewhere while the family is in observation will erroneously alter the number of children who appear to have survived, and hence affect the observed rate of mortality. It is important to note that these consequences of missed burial records apply equally when the family reconstitution is being constructed, so that care must be taken that burial records, particularly those of parishioners' children, are not 'lost', but are correctly linked to baptism records during the process of record linkage.

Any baptisms that may be missing while a family is otherwise in observation will affect measures of fertility, but one rule employed when constructing the record linkage for family reconstitutions concerns the maximum interval between successive baptisms to the same mother. The application of this rule may help to detect missing baptisms and remove the suspect part of a family's history from observation.

The quality of London registers

The format of early modern registration systems is irregular, and the recording of any supporting identifying information other than names was left to the discretion of the clerks who kept the registers. There was no requirement for London clandestine marriage centres to record details such as the home parish of either party, for instance, or the groom's occupation. That they did so seems to have arisen as a pragmatic response to the sheer volume of marriages conducted in these centres. The clerks of very populous suburban parishes like Aldgate or Clerkenwell were similarly compelled to considerable engagement with their duties by the number and frequency of events they had to record. In such parishes, the clerks could not have been personally familiar with every parishioner, as they might have been in the tiny parishes of central London, or in smaller communities elsewhere in England. This probably induced a greater and earlier awareness of the difficulty of distinguishing between individuals by name alone.

Of course, there is a tension between a population sufficient to trigger enrichments in parochial registration and a population so large that the effective recording of its baptisms, marriages or deaths is an impossible task. This is particularly noticeable when the recording system is under exceptional strain, as in plague years and at other times of crisis. Large parishes may be prone to lapses in registration or a reduction to the tersest possible form, where the name of the subject alone is given. However, the early Aldgate parish registers are beautifully well kept, at times resembling an illuminated manuscript. One factor mitigating the temptation to succumb to 'short-form' registration in later decades may be the parish clerk's awareness of being part of a continuing tradition, so that parish registers which are well-kept in the early years of parochial registration (that is, before 1600) may have a greater tendency to remain

so later on. As the custodian of the parish's earlier registers, the clerk would have been well aware of earlier standards, and where previous registration had been exemplary, this evidence of care and toil perhaps served to reinforce the importance of his own work.

Suburban parishes often produce registers that include long runs of occupational information, ages at death or addresses, as well as consistent naming of both parents and other related persons in the baptism and burial register. It is useful to obtain an indication of the extent and timing of such information before embarking on record linkage, since this helps to ascertain which types of supporting identifying information are most widely available, and thus where efforts in standardising this information for record linkage may most profitably be directed. Figure 1 shows the proportion of baptisms and burials in each year where both parent forenames, occupations and addresses are given in Aldgate, together with the number of baptisms and burials per year as an indicator of the changing volume of events the parish clerks were recording. The recording of new types of identifying information sometimes appears to arise as a response to times of rapid population growth, when the difficulty of distinguishing individuals was presumably greatest.

Additional identifying information in the parish registers vastly improves the potential for making confident matches between individuals, especially where it is carried across more than one of the three event registers of baptisms, burials and marriages. The baptism register is of most central importance to family reconstitutions. These records typically give the most information, naming one and often both parents as well as the child being baptised. It is a great help to record linkage if both parents are named. Father's occupations are also valuable in distinguishing between individuals if they are given consistently for all or nearly all records. Street addresses are useful to a lesser degree.

The quality of the burial register is very important in determining how unambiguous record linkage will be. It is important to be able to distinguish between adults and children, and for the details of at least one parent to be given in child burials, to cross-compare with the parental details given at baptism. Burial records typically give varying amounts of information depending on whether the deceased is an independent adult or a child still dependent on, and probably living with, its parents. Age information is rare, but most children can be identified as such by the phrase 'son of' or 'daughter of' followed by the parent(s)'s details. Usually only the deceased themselves is mentioned if an adult male, but often the surviving husband's name is given if the deceased is a married female. The presence or absence of titles such as Mr or gent, or of an occupational descriptor, can further aid in the distinction between adults and children. As for baptisms, with the burials of children it is a great help to record linkage if both parents are named, occupations of the deceased or their parent or spouse are also valuable where given consistently. Street addresses are useful to a lesser degree.

In the marriage register consistent indications of whether or not the bride or groom are of the parish will greatly improve the reliability of record linkage. In later parish records the marriage register may mention the father and perhaps the mother of the bride and groom, but this is almost never the case in pre-1753 London parish records, where parish marriage registers typically give details of the bride and groom only.

From parish registers to reconstituted families

The potential links which may exist between the persons typically mentioned in London parish baptism, burial and marriage records are conceptualised in Figure 2. Some of these links are more robust than others, either because they take place over a shorter span of time in which the likelihood of leaving the parish or others of the same name entering the parish is consequently less, or because they involve other, related persons as well as the subject of an event and hence permit multiple sets of personal details to be taken into account when establishing whether a match between records is correct.

An important general principle of record linkage in family reconstitution is to make the strongest, most secure links first, so that the subject of an event is not accidentally removed from consideration for one type of linkage by establishing a weaker type of link that precludes it. For example, if a child died at an early age, it cannot also have gone on to marry. With London parish records, it would be unwise to try to find the baptisms of brides and grooms given in marriage records before establishing which baptisms can be matched confidently to child burials. This is not only because infant and child mortality is high in London, but also because the link between child baptism and burial is a two- or three-person link, featuring parent details as well as those of the child. It is therefore stronger than any link that can be made between a child's baptism and subsequent marriage, where only the details of the child themselves can be cross-compared.

A sequence of the twelve record linkage steps necessary to form a family reconstitution is summarised in Table 1 and described in detail under the next section heading below. The steps are given in the order that has proven the best fit for London parishes, proceeding from the more robust and probable types of linkage involving baptisms to less likely, weaker types of linkage involving burials and marriages. The order of these steps differs in several respects from the ten steps originally envisaged by Wrigley and Schofield, and there are two additional steps.[9] The corresponding Wrigley and Schofield step, where it exists, is given in the third column of the table for the purposes of comparison. The numbering of the Wrigley and Schofield steps is discontinuous because steps 3, 6, 8 and 10 consist of quality testing and revision of earlier steps. In the 12 steps identified below, this testing forms an integral part of the steps themselves. Demographic constraints on record linkage referred to in the steps that follow deriving from maximum life span, minimum interval between births and so forth are outlined in Wrigley and Schofield (1973) and stated in more detail in Appendix 4 of Wrigley *et al* (1997).[10]

Before record linkage can begin, every baptism, burial and marriage must have a unique identifying number. Every subject of an event and the parents of that subject, where mentioned in the same event, should also have a unique identifying number.

---

[9] For the original Wrigley and Schofield steps see pp. 68-73 in Wrigley E A and Schofield R S (1973):'Nominal record linkage by computer and the logic of family reconstitution' in Wrigley E A (ed) *Identifying People in the Past*, Edward Arnold, London.

[10] *Ibid.* pp. 73-75 and Wrigley and Schofield (1997) *op. cit.* pp. 574-577.

Since it involves alterations to the data, record linkage should be performed on a copy of the underlying parish records dataset. Names, occupations, age information and residences will need dictionaries or lookup tables. Widow(ers), stillborns, illegitimates, bachelors, spinsters, chrisoms, twins etc will need to be flagged. This and other preparatory work is described in the section headed 'Prerequisites for record linkage and normalization of input' below.

As mentioned above, baptisms are at the core of the family reconstitution. In practical terms, this means that proto-families are constructed primarily by adding information to the baptisms. An arbitrary unique number represents each proto-family, and it is the addition of this number to the baptism records that are deemed to fall within the same family that creates each proto-family in the first instance. Links to other records are usually represented by adding the identifying number of that record (and pertinent other information, such as the date of the event) to new data fields in the baptism records. In some cases the link applies to the subject of the baptism and hence is added to only one baptism record. In other cases the link is to the parents of a baptism record and hence is added to all baptisms that have been linked to a proto-family. It is useful to add a reciprocal link from other records to the baptism. This is because once a record has been linked it is usually removed from the pool of possible matches in later steps, and it is therefore convenient to be able to identify with maximum simplicity which records have already been linked.

Family reconstitution record linkage steps

(See also Table 1)

At **step 1**, baptisms are compared against each other. Baptisms with the same parent forename(s) and with dates indicative of plausible birth spacing are assigned the same identifying number in a new data field, forming proto-families. Twins are treated as though they were one baptism at this stage. An indication that the mother has been widowed implies that the baptism must be of the last-born child, as any subsequent children will have a different father and hence should form a separate proto-family.

In **step 2**, parents from the baptisms that comprise each proto-family are compared to the spouses from all marriages that precede the earliest baptism date. Details of the most plausible marriage are added to the proto-families. If no match is found no details are added, and in later analysis the family will be regarded as in observation from the date of the first child's baptism.

In **step 3**, the children identified in baptism records are searched for in all burial records of the same date or later that are known to be of children, either because parents are named or ages are given. Both the child's own details and those of its parents are used to identify matches. Only matches that imply an age at burial of less than 15 years are considered, unless the match is an exceptionally good one, in which case the age criterion may be raised to 25 years. Details of the most plausible match, if any, are added to the baptism record and this burial is removed from consideration in later steps. If the burial record indicates either parent has been widowed, this is regarded as the latest possible end date for that family, and any baptisms that postdate it are split into a new proto-family.

In **step 4**, both baptism parents and marriage spouses are taken together and matched against burials where a surviving spouse is identified. Details of the most plausible match, if any, are added to the baptism records of the proto-family. Where the match involves a baptism parent, any baptisms that postdate it are split into a new proto-family. Where two or more children of the same name have been baptised into the same family and one or more potential burial records potentially match, links that produce the shortest baptism to burial interval are favoured. This is a heuristic rule and is not infallible, but it represents the most likely case since deliberate same-naming of living siblings is uncommon, and the risk of dying among children broadly decreases with age.

Steps 1 to 4 represent the strongest types of record link that may be made. The record links that come after are weaker, and hence must fit in fully with the proto-families now established. This means that the full complement of potential reconstituted families has now been obtained. Subsequent steps will not alter those links that have already been made, and there will be no further splitting of proto-families. However, Step 5 may be considered almost as strong a type of record link if the marriage register gives information on the parents of those marrying.

Step 5 looks for the first marriages of sons and daughters born into a family, by comparing baptisms where no burial record has been found to non-widowed marriage partners that are known to originate from the parish, where this information is available in the marriage register, or at least to those not explicitly from another parish. If a marriage has also been matched to subsequent baptisms, the interval from parental baptism to prospective children's baptisms must be plausible. The calculated time interval from parental baptism to any implied date of burial for that parent must represent a plausible age. In circumstances where many parishioners married elsewhere, it may be desirable to implement a subsequent search for matches straight from the baptism of a male to any subsequent reappearance as the father in baptism records, where no intervening marriage is in evidence. Women cannot be traced in this manner because of their change of surname on marriage.

Step 6 continues the process of linking baptisms to corresponding burials. Those baptisms not already linked to a burial or to a subsequent marriage are matched against burials where it is not known whether the deceased is a child or an adult. For this stage only matches that imply an age at burial of less than 15 years are considered plausible.

Step 7 makes allowance for children who died before their baptism was registered, using primarily the burial parent details to determine whether a link with a proto-family should be considered. This step encompasses both those known to be dead on delivery (stillborns and abortive children), and those who lived a short while, who are often but not always identified as chrisom burials. The first group are straightforward, but the latter is more difficult to deal with as it can be difficult to distinguish from other child burials. As well as chrisoms, a child buried without a forename usually falls into the category of those who lived only a short while. However, so may a child that is named, since during the chrisom period the child might have a name assigned but not yet entered into the baptism register. To reduce the uncertainty, it is imperative that the other types of baptism to child burial record

described above place before this step. Where the buried child is not a chrisom or forename-less, the remaining consideration for linking it to a proto-family is that it must fit in with events and dates already established, and in particular it must slot into the known sequence of baptisms in a proto-family without introducing an implausible birth interval. Child burials that satisfy these conditions are assigned dummy baptism records with a date of baptism nominally equal to the date of burial, and linked to the proto-families.

In **step 8** parents and marriage partners are linked to their own burials, taking into consideration only those burials known or thought to be of adults. Where the parent or marriage partner's date of baptism is known from earlier record linkage, the implied age at burial must be plausible, or else the interval between the earliest baptism and burial must accord with a plausible maximum fertile period in the case of women, or lifespan in the case of men. Adult status is derived either from stated age information or because an occupation or title (such as Mr or Mrs) is given, and in the absence of a parent being named. The presence of occupations or titles are heuristic rules for identifying adult status only, and clearly some types of occupational descriptor will identify children or young adults. While titles may be applied to children, they are much more usually ascribed to adults with responsibilities in the parish.

Step 9 establishes a link between subsequent marriages of the same bride or groom, matching between baptism parents or marriage partners and their later remarriages. To reduce the potential for mistaken identity, these links are only attempted where one parent or spouse is known to have died. The match is thus from the surviving partner the wife or husband of marriages that postdate their partner's burial. Any marital status information given in the marriage register must accord with their widowed status, so matches cannot be to known bachelors or spinsters.

Step 10 continues to establish remarriage links for surviving spouses, but for cases where the remarriage took place elsewhere and only subsequent baptisms can attest to the continued presence of the male partner. Women's remarriages cannot be investigated in this manner because their surname change on marriage renders them untraceable without an entry in the marriage register. As in step 9, these links are only attempted where the wife or mother is known to have died. The link is from the surviving husband in a marriage, or father in a proto-family, to a father identifiable only through the subsequent baptisms (and burials) of a proto-family. The minimum implied duration of both marriages must be consistent with a plausible maximum lifespan.

Step 11 completes the search for burials of those who reached adulthood and married, linking from marriage partners or baptism parents where no burial record has yet been traced to burial records of individuals whose status as an adult or a child is unclear. These links are usually very weak. In addition to according in full with a plausible maximum lifespan or maximum fertile period, as in step 8, new burial links must also not render invalid any remarriage links identified in steps 9 and 10.

Step 12 accounts for individuals who reached adulthood but died without marrying, linking between the child in a baptism record to subsequent burial records of known or probable adults. These links are usually very weak and must fit in with all prior links and a maximum plausible lifespan. Burial records indicating the deceased was

married may not be included (but titles such as Mrs are no reliable indicator of marital status before the 19th century).

Ways of proceeding with record linkage for family reconstitution

The steps of record linkage in reconstituting families described above can be conducted by three main methods:

1. Manual method

In the first method, events are copied from the parish register onto paper slips and the reconstitution is constructed by hand, sorting and re-sorting the slips to match up the individuals belonging to each family and entering their details onto a Family Reconstitution Form. The forms can later be digitised to facilitate analysis.

2. Computer-assisted method

The second method is to use a relational database to store the records, adding lookup tables to standardise crucial parts of the input, such as names, and to match the records against each other using SQL (Standard Query Language) queries, intervening manually to select the best record links from a large number of possibilities at each stage. There may be scope for reducing the amount of manual intervention needed by using Bayesian logic to score the relative probability of record links.

3. Automated method

The third approach is to adopt a fully automated or computerised approach, preparing lookup tables to standardise all parts of the input (names, dates, occupations, residences and so forth), and processing machine-readable transcriptions of parish registers (which may be stored in a database as in the second method) by algorithm to form reconstituted families. This third method is not necessarily the best in all circumstances, for it depends absolutely on the quality of the lookup tables and the comprehensiveness of matching rules. These are no trivial task to construct and may omit some of the surrounding context.

When choosing suitable methods for constructing a family reconstitution, scale is an important consideration, as is the time available and the intended purpose of the reconstitution. The main limiting factor in an automated solution is the comprehensiveness and flexibility of the record linkage algorithms, and the preparatory work that shapes the input into a suitable form, since the computer processing power and memory requirements to handle large numbers of records swiftly are easily obtained.

For smaller communities, the quality and completeness of fully automated record linkage is likely to fall short of what may be obtained by manual intervention. An important pragmatic consideration is, then, how much manual intervention may reasonably be attempted. A parish with less than a hundred baptisms per year and a population of a few hundred persons or even a few thousand will lend itself to the

entire output of each stage of record linkage being scrutinised by eye, but once there are several hundred baptisms per year and the population exceeds ten thousand or so this becomes unacceptably time consuming at best, and indeed virtually impossible for the largest parishes. In most cases, the second approach outlined above of using a database to assist in record linkage and refining the results by hand will provide a good solution, but in the largest urban parishes it is too time-consuming.

As a rule, for the smallest parishes, the first or second method outlined above will take similar amounts of time and produce similar results. For medium-sized and large parishes and the second method will probably be more efficient, and for the largest parishes only the third method is likely to be practicable.

The assertion that fully computerised record linkage for family reconstitution is not always the best approach requires further emphasis and explanation. In record linkage, when matching up records by hand we see the information on a case-by-case basis. At a glance, discrepancies between records or contextual disparities can be seen and a match prevented. Each prevention is simple, even trivial, to apply, but the reasoning behind it may be a complex interplay of factors. Overall we make use of a considerable amount of internalised knowledge, such as the usual phonetic behaviour of English and its orthographic representation when deciding whether two names refer to the same person. This sort of context-dependent knowledge is extremely difficult to codify fully in a computer program. Record linkage for family reconstitution is further complicated by variation according to historical and evidential circumstance, since different customs prevail at different times and English parish records tend to go through unpredictable phases of recording different degrees of detail, largely at the discretion of the parish clerks. On the other hand, in record linkage by hand it is very difficult to maintain absolute consistency – to treat each potential match exactly equally. A computerised solution will always be consistent, but care must be taken that it is consistently correct rather than consistently flawed.

A useful cautionary maxim when generating any sort of dataset by computer is that of 'garbage in, garbage out'. It means that flawed input will lead to meaningless results. Computers operate on large numbers of records at once, which both magnifies the effect of errors and leads to a difficulty in detecting them. A program can generate vast amounts of output in very little time, often more output than can or will be checked by human eye. Incompatibilities, particularly ones applying in certain circumstances only, will pass undetected unless they can be envisaged upfront and measures are taken to prevent them, or they are tested for and resolved later on. Resolving problems at a later stage is generally a much more time-consuming process than preventing them from occurring in the first place.

One important additional consideration is that a partly or fully computerised solution to record linkage has a great advantage over manual methods in representing records, as well as in manipulating them for analysis. It is possible to create multiple ways of looking at records, and this allows easy visual inspection of the record links that have been made. In a Microsoft Access database, Forms can readily be constructed as a user interface not only for data inputting, but to view reconstituted families at a glance and to inspect the transcriptions of parish register events that underpin them. Like hyperlinks between web pages, clickable buttons can take the user from persons in a reconstituted family on one Form to the separate Forms representing baptism, burial

and marriage records that are implicated. This is a particularly useful tool for expository work while the record linkage is still under construction, as well as providing a convenient way of representing the output.

Prerequisites for record linkage and normalization of input

In any family reconstitution that is to be undertaken partly or wholly by computer, there is substantial preparatory work that must be done well for the output of reconstituted families to be useable. Most important of all is the manner in which names and dates are treated. When linking English historical records, names and dates are the most ubiquitous pieces of information given that help to identify individuals. Occupations, ages and addresses are important too, but it is almost unheard of for one or more of these to be present throughout a pre-19th century parish register. An additional problem with addresses is that they are prone to frequent change over the life-course of an individual, and the same is true to a lesser degree for occupations.

The parts of historical records that identify people that may be relevant in a family reconstitution are:

- names
- dates
- ages
- occupations
- residences

The most important identifying information comes from names and dates. Some information on age can be inferred, such as whether the individual is a child or an adult, but more exact, numerical ages will be useful where present in the parish registers. Likewise occupations and residence information will be useful where this information is given.

The first task in preparation for record linkage is to normalize, or standardise, these crucial pieces of identifying information to make comparison across records possible. None of the information will take the same format in every record of the same individual (nor will it always be present), but some types of information are subject to more variation than others. Dates, and to some extent ages, lend themselves to relatively simple numerical representations, whereas the textual information of names, occupations and residences is much less clear-cut. It is time-consuming to standardise this sort of information well, although some potential shortcuts exist.

When capturing data from London parish registers we have preferred to retain as much of the original source as possible but in a highly structured format that separates out different types of information into different data fields. Some types of information such as dates can safely be represented in a standard format during data capture (and translated into other calendars as necessary afterwards). Certain circumstances that are critically important for record linkage that should be flagged up during data inputting, as described below. However, to avoid inconsistencies or prejudicing the end result it is preferable to defer standardisation of names, occupations, ages, and residences until after all the data have been gathered.

Flags

The circumstances that are flagged up during data inputting are as follows: date uncertainty, out of register order, name alias, abortive, stillborn, chrisom, infant, child, junior, senior, adult, aged, of this parish, stranger/alien, householder, inmate, lodger, twin, illegitimate, foundling, parish child, bachelor, spinster, widow/er. The flag is applied by appending one or more standardised textual conditions to the individual or event concerned, with further detail supplied in a comment where necessary.

Dates

Taking the easiest information to standardise first, we begin with dates. If these have been entered in a standard format, only the date uncertainty flags require further work. It may also be convenient to convert the dates to the Gregorian calendar for ease of manipulation. For uncertain dates, the period of uncertainty is determined and represented numerically within the event record. Minimum and maximum possible dates for the event can usually be established from other, surrounding events in the parish register. A new numeric field is added to represent the uncertainty, and is assigned a positive or negative integer value that gives the maximum possible deviation in days from the stated event date. The event date itself should represent either the earliest or the latest possible day on which the event must have been recorded.

Since dates are very important, yet highly susceptible to typographical errors that may alter them substantially, it is useful to perform a check that the dates conform to the expected sequence. In order to perform the check the input data must have been entered in the same sequence as the parish register and possess a numbering scheme that reflects this. The check is carried out by using an algorithm to detect where the day, month and year parts of each date are not chronological sequence when ordered as in the original parish register, except where the register is noted to have been out of order. Erroneous dates are corrected manually. Large jumps in time from one event to the next are also checked to make sure they genuinely reflect the source.

Ages

The goal of standardising ages is to represent all types of age information numerically in a standard unit of measurement, for which years (and fractions thereof) are usually a convenient choice. Age standardisation is performed by constructing an age dictionary or lookup table that pairs each age descriptor from the original data with two standardised values that give the minimum and maximum age in years that the descriptor is deemed to represent.

It is a relatively straightforward matter to standardise ages that are given as a quantity of time, whether in words as "tenn Yers ould" in roman numerals as "ii weeks" or in Arabic numerals as "quarters 3", converting the units into decimal fractions of years as necessary. However, there is inherent uncertainty in age information, whether it arises from the way we describe age, ambiguities in the intended units, or inaccurate recall. The first of these causes of uncertainty refers to the fact that ages usually represent the number of *completed* years (or other unit) that an individual has

survived, not the age to which they are closest. A child described as 10 years old may thus be only a day away from their 11th birthday. This can be accommodated by representing each standardised age descriptor in two separate fields containing a minimum and maximum value, so that the 10 year old child may be represented as being at least 10 years old but less than 11 years, for instance. We can also account for possible vagueness in remembered ages in this manner, and so the bounds for the 10-year-old child might become at minimum 8 years but less than 12 years. When representing potentially inaccurate age reporting in this way, it is best to make less allowance for error in the youngest ages than in older ages where the likelihood of reporting the age wrongly is greater. Allowance may also be made for greater inaccuracy on ages ending in zero, or other values over-represented in the data, where age heaping is detected.

Table 2 shows a sample from an age lookup table. Age information that is vague, implicit, or expressed only relatively can be represented numerically by setting *probable* upper and lower bounds. Some age information of this type is flagged up during inputting, such as infant or parish child, as described above. This type of information becomes much more convenient to use in record linkage when it can be represented as lying between minimum and maximum numeric values. For example, a woman described as "aged" is probably no younger than 50 years and no older than 100 years. A "youth" is probably no younger than 10 years and no older than 20 years.

Names

Name standardisation is the single most important factor in improving the success of record linkage. Written English is highly irregular and changeable in the way it represents the sounds of the spoken language. An almost infinite variety of spellings were usual for names, particularly in the 17th and 18th centuries. Reducing this variety to standard name forms that can be used to cross-compare historical records is a time-consuming process, and it cannot reliably be performed entirely by algorithm. Using the method that will be described below, about 2000 distinct names can be standardised per day. Some subjective judgement is necessary, but objectivity and traceability can be kept at the core of the process by using an algorithm to guide the initial stages and by fully recording every subsequent amendment that is made.

The forenames and surnames of a parish population have different characteristics, and there are many more surnames than forenames. Figure 3 shows the number of distinctly spelled forenames and surnames in London parishes of different sizes. Figure 4 shows the marked contrast between the distribution of the relatively small number of different forenames in London parish registers, and the much larger number of different surnames. Forenames do not differ markedly in other English communities, but different types of community produce quite different surname distributions. Table 3 shows surname popularity in several different settlement types. In isolated rural communities parish records can be heavily concentrated around only a few surnames, but in London surnames are much more diverse. While this is a great aid to record linkage, it also means that the names take longer to standardise.

 Forenames and surnames are treated separately during normalization, in two sets of name dictionaries or lookup tables. Forenames are generally easier to standardise than

surnames because there are far fewer distinct names. Names in London parish registers before 1750 usually consist of one forename and one surname. However, multi-part forenames or surnames are occasionally given, and ways of dealing with these are discussed below. Titles are not consistently given in parish registers but may be useful for implicit age information, especially in the burial register.

The starting point for forename and surname dictionaries are lists of every distinct name spelling found in the parish registers, together with the number of times it is used. The forename dictionary additionally includes the sex(es) to which a name can belong. We derive sex in the first instance from the use of specific words in the source material, such as son, daughter, man, woman and so on. The sex may be inferred from the name itself, but we deem it safer to do so only where necessary. It is worth remembering that some names have changed sex over time, such as Christian or Bennet (both originally female but later male), while others may very occasionally be used for the opposite sex than expected – there are a small number of females named Peter in St Botolph Aldgate, for instance.

Once the names have been aggregated, codes representing the sound of each name are attached to the dictionary using a modified version of the Double Metaphone algorithm.[11] This contains additional rules to accommodate Early Modern orthography, such as the interchangeability of v, u and w. It also produces a number representing the count of syllables in the name. While Double Metaphone performs better than Soundex, it is far from infallible. Sorting the names by these codes and the syllable count produces rudimentary groupings of like-sounding names. Each list is then refined by hand, and this is the time-consuming part of the processs.

Table 4 shows a sample of a surname table. For surnames, the refinements performed by hand assign each name to at least one and up to two standardised forms. Borrowing the classification terminology of Linnaeus, the first standard form may be designated its species and the optional second form its genus. When used for record linkage, the system operates as a hierarchy of preference, with matches between species favoured above matches on genera. Matching is also permitted between the species of one name and the genus of another, or *vice versa*, to add flexibility to the system. A match between names that share the same Double Metaphone code only is not sufficient for a link to be made, but it is useful nonetheless to conceive of the codes as a third, higher level in the hierarchy that can be used to improve efficiency (that is, to optimise algorithmic performance) when searching for matches.

The spelling of the standard form chosen to represent a name's species or genus may reflect either the most common usage in the data, or the one with which the standardiser is most familiar – often these are one and the same. The former is less prone to misjudgement, but the latter is easier to remember and so may aid consistency. For example, many surnames are also English place names, and if the standardiser possesses a reasonable knowledge of place names it may be convenient to use the modern place name spelling as the standard, where applicable.

---

[11] The implementation used is that developed in Newton G H (2005): 'Creating a customisable name matchinig algorithm for historical computing by refactoring' (unpublished Anglia Polytechnic University MSc thesis), but the algorithm was originally conceived by Lawrence Philips and described in Philips L (2000): 'The Double Metaphone Search Algorithm', *C/C++ User's Journal*, 18:6.

This method is essentially a data-driven approach, where surnames will be grouped chiefly with reference to each other rather than to some external standard. The main advantage is that no further knowledge need be acquired or applied, and the method stands a reasonable chance of bringing together names the parish clerks of a specific place and time heard as equivalent. However, inevitably some names will not be standardised correctly, and in some respects the method is limiting because it makes the name dictionary for a parish specific to it alone. It will not necessarily be applicable to other locations, because different gradations from one name to the next may prompt different standardisation decisions.

The possibility of names being standardised incorrectly can be mitigated for small numbers of problematic names by facilitating rapid lookups between surnames isolated in the name dictionary and the records in which they appear. This will permit wider context to be taken into account for some names during name standardisation, but it is too time-consuming to be undertaken for all names in a large parish. In effect, the alternative method of name handling described above, where forenames alone are used to form all possible matches and the correct ones are picked out later, is a logical extension of this more context-dependent way of handling names.

Forename dictionaries do not need to be refined in as much detail as those for surnames because there are fewer distinct forenames. They tend to fall naturally into relatively discrete groups, especially once sex is taken into account. For most names the goal is therefore only to correct the Double Metaphone algorithm groupings and not to refine them further. It does not matter whether the phonetic codes strictly represent the sound of each name so long as name spellings that are deemed to represent the same name end up in the same code group. To correct the groupings it is best not to edit the codes by hand, but rather to assign those names deemed to be in the wrong code group a regroup name and subsequently to run the algorithm again on these regroup names. The regroup name should reflect the spelling of a correctly grouped name that the wrongly grouped forename ought to match with. This makes it easy to trace the corrections that have been made and to see the reasoning behind the change. Table 5 shows an example of a forename table.

Since the top 10 forenames can easily account for three quarters of the entire male or female population (see Figure 3), it is prudent to identify the most popular forenames and to be especially careful in their treatment. For very popular names with close variants, such as Jane/Joan/Joanna, more refined grouping is added using the same method as that described for surnames above. Identifying the relative popularity of names within the full set of forenames represented in the dictionary can be achieved using the name frequencies. From these the probability of a particular forename applying to an instance of an individual may be calculated, and stored as a decimal value between 0 and 1 in the name dictionary. During record linkage, these name probabilities can help to rank the potential likelihood of matches, especially where record linkage is attempted using forenames alone. The probabilities are of course only meaningful if there is relatively little change in name popularity over time (although probabilities could be calculated for specific periods).

Taken alone, name probabilities are often not very discriminatory, but the combined probability of a couple's two forenames will be much smaller. A potential record link between an individual named John in one record and another record with an

individual named John is clearly very weak indeed, whereas a potential link between two records of couples named John and Mary is better but still far from conclusive (John and Mary are both very common names). A potential link between two records with couples named John and Dulsabella (a very uncommon female forename) is much stronger. It may be possible to accept potential links where the combined forename probability falls beneath a certain threshold without further reference to names, providing of course that all other identifying information in the records is consistent.

In the forename dictionary, special care must be taken with abbreviations and multiple names. Abbreviated names or nicknames will often be incorrectly grouped by algorithm because their spelling does not suggest a sound similar to the forename they are derived from. They must therefore be corrected by hand, by assigning each a 'working name' as described above. Forenames and middle names, or two-part forenames such as "John Baptist", can be handled by adding each distinctly spelled multiple forename to the dictionary multiple times. Each occurrence of a multi-part name in the dictionary will be assigned a different 'working name', so that all permutations of the forename parts can be represented and allowed to match. It is convenient to add a numeric value to such entries to represent any preferential treatment of some permutations over others. For example, "Tryphena Tara" might be added to the dictionary three times, firstly with the working name "Tryphena Tara" and a score of 10, secondly with the working name "Tryphena" and a score of 5, and lastly with the working name "Tara" scoring 1. A similar method serves for surnames where an alias is given.

The method of handling names described above assumes that all names will be fully standardised before record linkage begins. It is important to note that there are alternatives to this approach that have proven viable with smaller London parishes of fewer than around 10,000 inhabitants, using the second approach to constructing family reconstitution described above. One possibility is to concentrate only on fixing under-grouping of names without introducing further refinements. Another is to standardise forenames only, and to ignore surnames when forming potential record links. However, these methods do not eliminate the time-consuming nature of handling names, but rather defer it to later in the process of record linkage. Picking out the best name matches from the mass of potential links that result may be assisted by using string similarity algorithms to rule out the least likely. Human judgement will nonetheless be needed to assess the relative compatibility of many names.

Occupations

In pre-19th century parish registers, the only reason for occupation to be stated is to aid the parish clerk in identifying persons, and this usually results in quite specific occupational descriptors being given. In an urban area, the diversity of available work, and hence occupations, is much greater than is usual for a rural area. An urban parish register is therefore less likely to use vague descriptors such as "labourer". This specificity and diversity of urban occupational descriptors is very useful in discriminating between individuals. Of course, individuals can change occupation over time, but the likelihood of changing occupation is uneven throughout the life-course. During the years of family formation that family reconstitution primarily

seeks to capture, the occupation of the father is relatively unchanging. Changes of occupation that do occur are frequently linked to migration, which would in any case remove an individual from observation in the parish records. It is also extremely rare for a father to be accorded more than one descriptor in the same record.

Table 6 shows a sample occupational lookup table. The starting point for the standardisation and classification of occupations, as with names, is a list of each distinctly spelled descriptor. The first stage in standardising is to normalize each into its modern spelling. The resulting, smaller list of distinct descriptors can then be classified into a hierarchical scheme, so that each standardised occupation falls within a category, each category fits into a subgroup of potentially related occupations, and the subgroups themselves fall into an overarching occupational groups. It is possible to deploy a predefined hierarchical occupational classification scheme to this end, such as the Primary, Secondary, Tertiary scheme. However, a more customised classification that closely tailored to the data may prove superior for the purpose of cross-matching occupational descriptors between individuals, but it will also take longer to apply.

In deploying the hierarchical classification of standardised occupational descriptors in record linkage, we have chosen to prevent no potential link on the basis of an occupational mismatch. Rather, potential record matches are preferred if they have the same occupation, category, subgroup or group, in that order of preference. Mismatches on occupation are not deemed sufficient evidence to reject a record link for two main reasons. The more straightforward and general reason is that individuals *can* change occupation, even if they are not particularly likely to do so while in observation in an urban family reconstitution. The second reason is specific to urban environments and concerns the role of guilds. In London, guild membership is an important qualification throughout life, but in the early modern period it does not necessarily determine the branch of employment. A conflict between guild qualification and regular occupation is sometimes explicitly acknowledged in one record, as in "citizen & mercer of London but by trade a bearebruer", "citizen and yernnemonger [ironmonger] of london by trade a tippler" or more tersely "free Diar but vittaler", for example. However, where the guild alone, or the trade alone, is given in records of the same individual, there is potential for apparent mismatch.

Residences

Residence information that allows parishioners to be distinguished from non-parishioners is the type most widely given in parish registers. It requires relatively little preparation for record linkage. In London parish registers, street addresses are also relatively commonly stated. These have proven less useful for record linkage than might be expected, although this is partly because of the way addresses are described in early modern London. This is discussed further below. However, street addresses are inherently less reliable for identifying people than other personal information, because moving house is more likely and on average takes place more frequently than changing name or occupation.

Distinguishing parishioners from non-parishioners is particularly important in marriage records. Both the bride and groom may be resident elsewhere, and this is especially common in London before 1753. In order to be able to remove such

21

marriages from consideration for some stages of record linkage, it is convenient to divide the brides and grooms into three categories: those who are parishioners, those who are not, and those for whom no certain information is available. Those who are parishioners are often described as "of this parish", and as these are flagged up during data inputting, no further standardisation is required. Other persons may have their home parish, town or county given, or even a street address. To account for these, a dictionary or lookup table may be constructed that contains each distinctly expressed residence and assigns it to one of the three categories of parishioner, non-parishioner or indeterminate. (Further standardisation that brings together different instances of the same foreign parish or place may be useful for other purposes, but for record linkage this is sufficient.)

In the burial register, information on whether the subject or their parents is a stranger or alien, householder, or lodger is often given. All of these conditions are flagged up during inputting and thus no further standardisation is required, unless there is additional information similar to that typical in marriages, in which case the same three-part categorisation system can be used for parishioners, non-parishioners and indeterminates. The householder, lodger and stranger conditions can help to distinguish longer-term residents from those merely passing through the parish. They may be used to prefer some record links over others, but not to exclude records for consideration during record linkage. 'Householder' usually implies that the person is a parishioner and has a good chance of appearing in other parish records, whereas stranger and lodger suggest a lower likelihood of appearing in other records. Similarly, many (but not all) persons referred to as strangers, aliens and lodgers are newly arrived in the parish, and thus are less likely to match correctly with other records.

If the burial or baptism register contains any further information, the same three-way categorisation of parishioners, non-parishioners and indeterminates may be used as for marriages.

Street address information is probably only worth standardising for record linkage purposes if it is given in both the burial and the baptism register, and even then it may provide little additional discriminatory power. As discussed above, it was especially common for Londoners to move house frequently but locally, staying within the parish. Inhabitants might stay in one place for longer if they had more invested in a specific building, such as a shop or workshop, but these tended to be on the main thoroughfares, and main street addresses are the least discriminatory. The older London streets had numerous alleys leading off main thoroughfares, and courts in turn leading off those. Being larger, the streets tend to have more residents, and this is exacerbated by people who actually lived in an alley or court but sometimes gave, or were assigned, the nearest main street as their address.

Even without standardisation, street addresses can still be used to prefer matches between records where the address details are identical, and the presence of a street address usually implies that the person is a parishioner. If standardisation is to be undertaken, it can be done by creating a dictionary or lookup table of street address descriptors. This is an exercise best done in consultation with contemporary maps, since some streets go by alternative names or change name over time. However, streets described in the parish register and those depicted on a map tend not to

correspond exactly. The starting point for standardisation should be the street addresses in the parish register since for record linkage purposes the main concern is to bring together those that are equivalent.


Concluding remarks

Urban and, especially, suburban parishes present special challenges to the techniques of family reconstitution. Not all of these can be overcome in every respect, but some are mitigated by the very large parish population from which a sample of reconstituted families may be drawn, and the generally rich quality of parochial registration in cities. Urban family reconstitutions are particularly well suited to the study of infant and child mortality. For other demographic measures, even where the results must be treated with caution, there is good potential for comparison with other urban parishes to determine the range of likely or possible values, and to determine the variability of population characteristics within an extraordinarily diverse environment. Once the limitations are understood, the potential rewards from urban family reconstitutions are considerable, since cities are the economic hubs of a country and the behaviour of their inhabitants is of very great importance to the growth and development trajectory of the nation as a whole.

Demographic analysis from family reconstitution depends in the first instance upon constructing reconstituted families from parish baptism, burial and marriage registers. The main intention of this paper has been to provide a step-by-step description of how to go about the record linkage that underpins family reconstitution, with particular reference to the most useful methods for large urban parishes. It is hoped that sufficient detail on both data standardisation and the steps of record linkage have been provided here that those with non-specialist technical skills may find a way to proceed. No one method has been prescribed for all circumstances, and a generalised programmatic implementation of record linkage for urban family reconstitutions that will work 'out of the box' for any urban parish is beyond the scope of this paper. Indeed, it has been suggested that this is not task that lends itself to a fully scaleable solution, and that communities of different size and growth rates will benefit from differently implemented methods. However, the importance of standardising the highly variable types of input that parish registers provide cannot be over-stated, in any programmatic or computer-assisted approach.

Figure 1: Comprehensiveness and timing of identifying information in the baptism and burial registers of St Botolph Aldgate

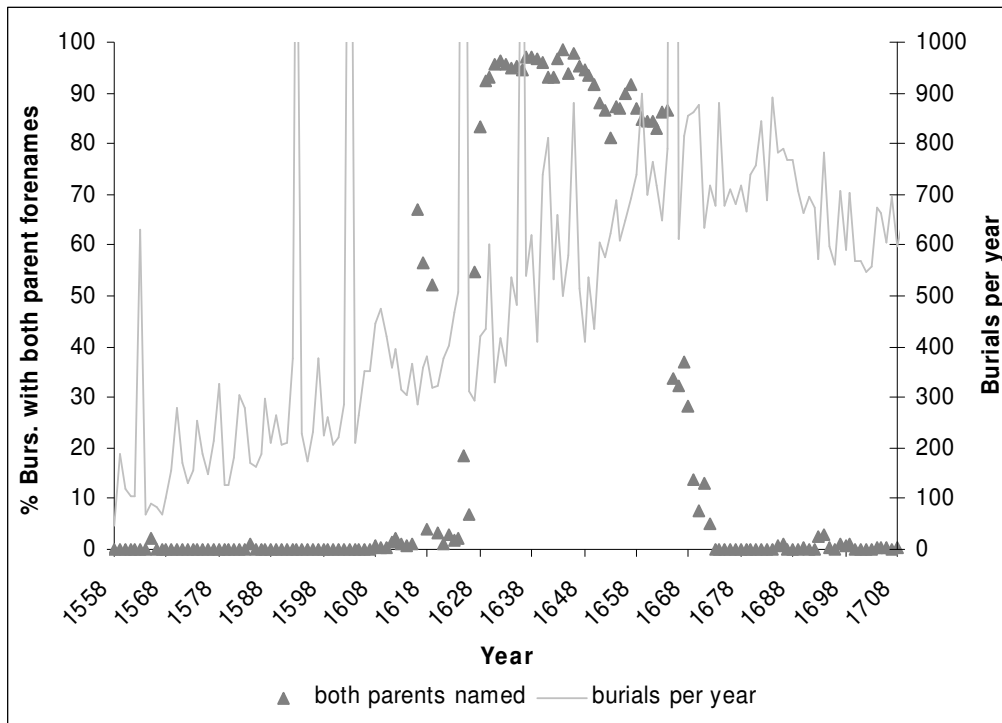(a) Proportion of baptisms where both parents' forenames are stated
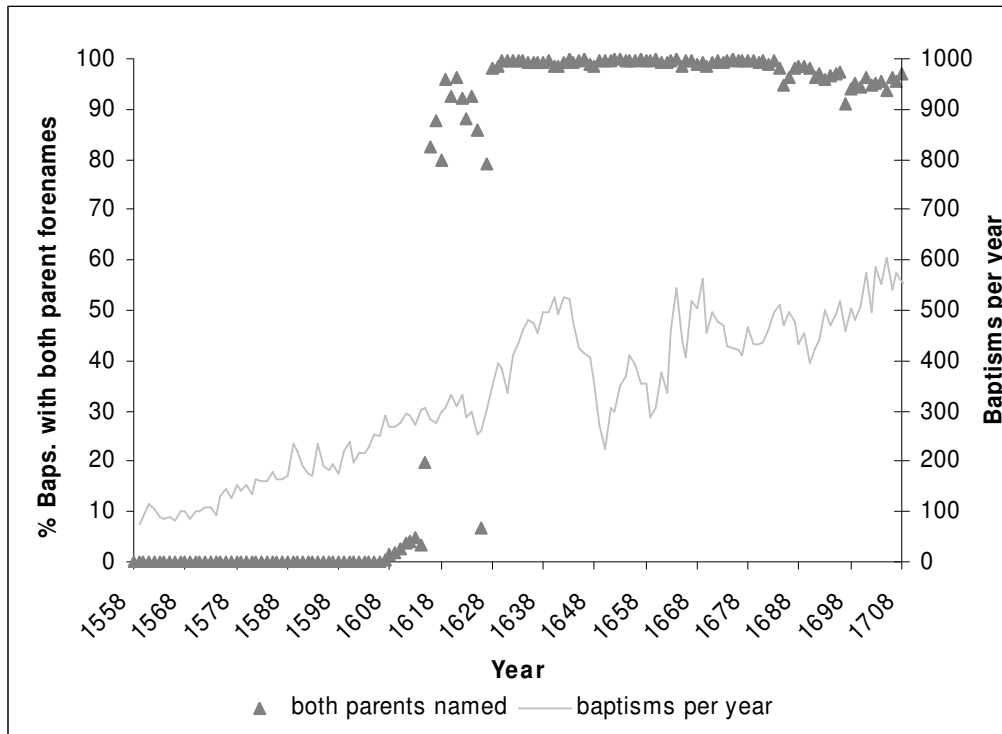
(b) Proportion of baptisms where father's occupation or address is stated
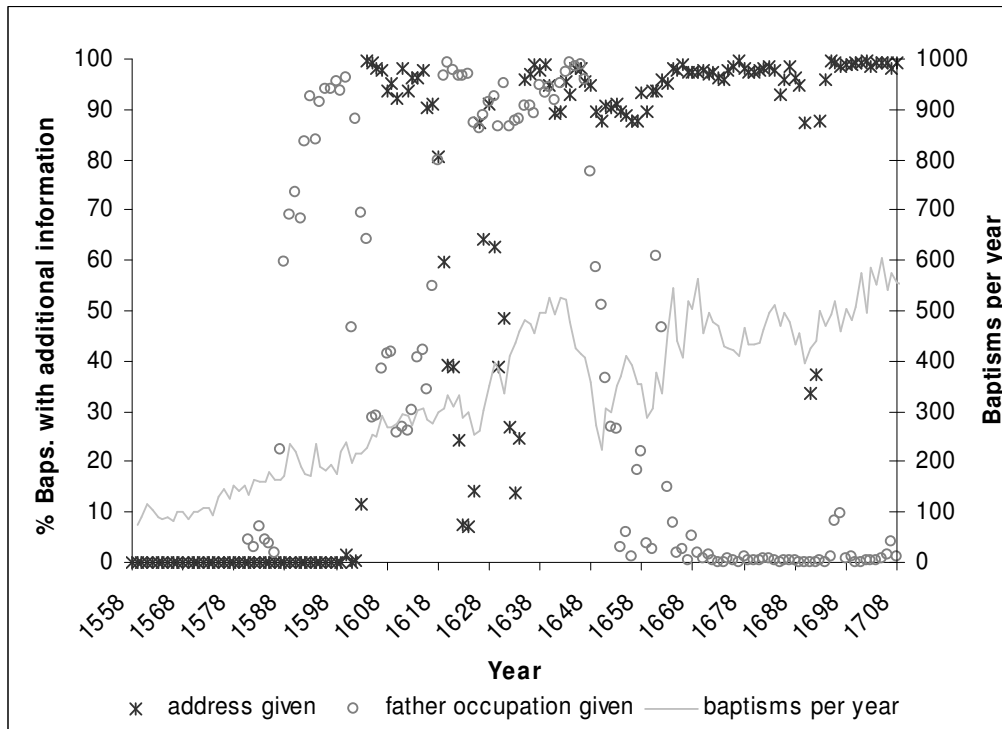
(c) Proportion of burials where both parents' forenames are stated

(d) Proportion of burials where father's occupation or address is stated
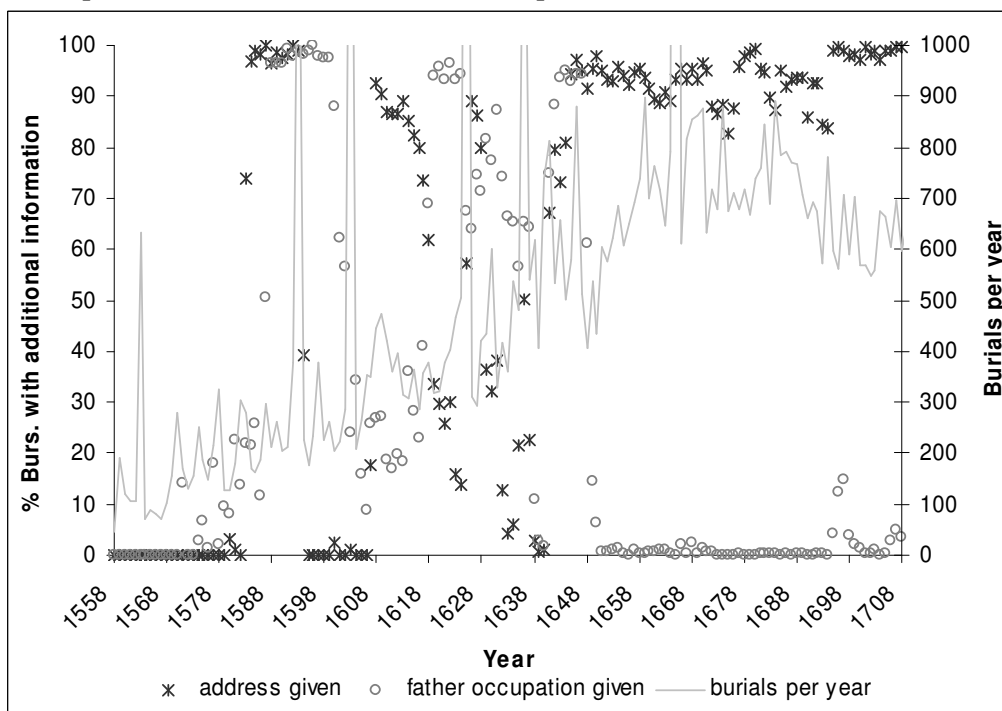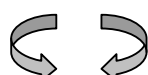
Note: In order to present a convenient, matching scale for baptisms and burials in each of these graphs, in (c) and (d) burials in the plague years 1593, 1603, 1625, 1636 and 1665 have been allowed to exceed the top of the *y*-axis. The total number of burials in each of these years was 1465, 1949, 2484, 1533 and 4548 respectively.

Figure 2: Potential links between individuals mentioned in baptisms, burials and marriages in a typical London parish (dashed boxes indicate persons less commonly mentioned)
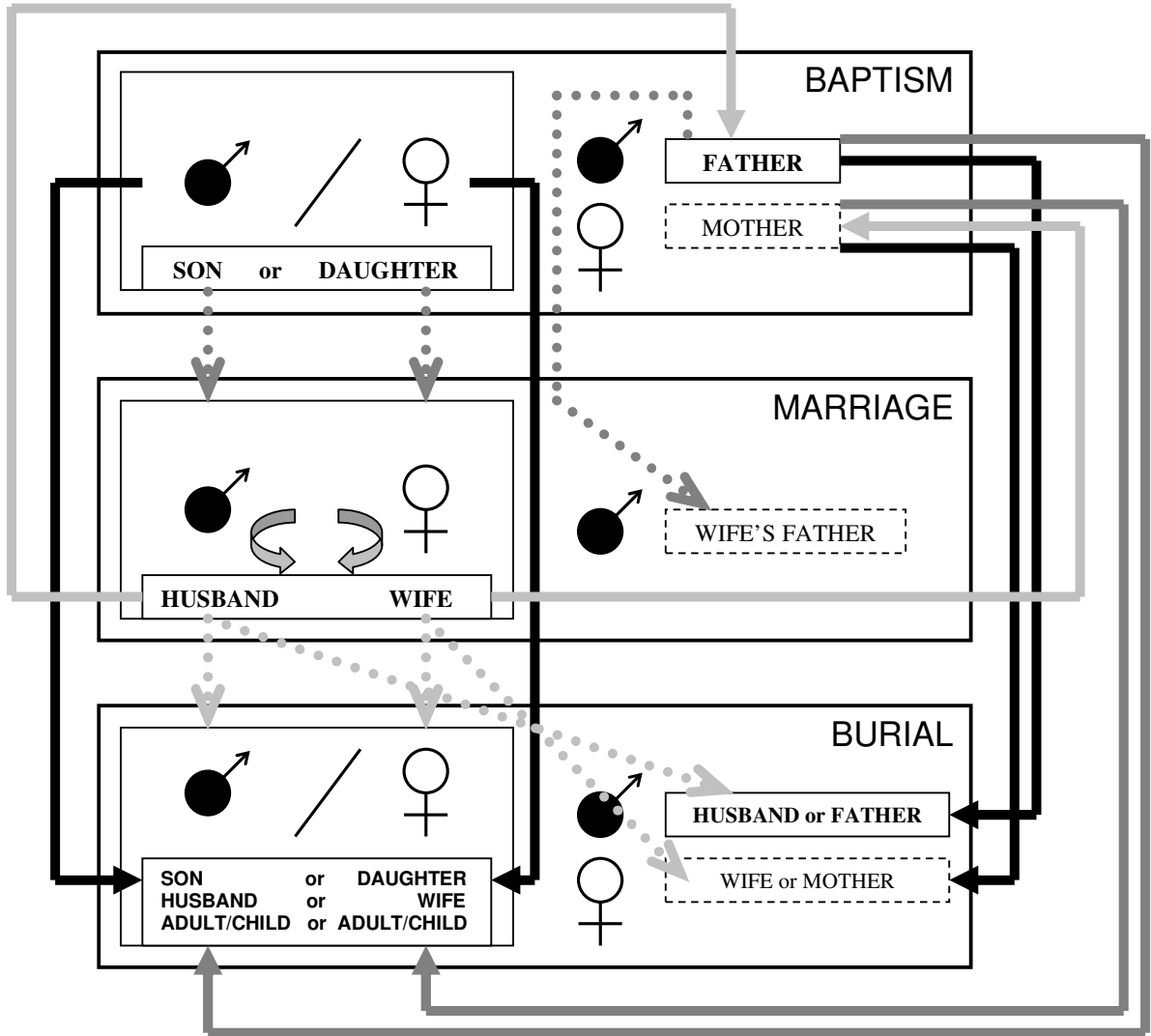
Table 1: The sequence of record linkage steps in family reconstitution as applied to London parishes

	Linking from	Linking to	Wrigley and Schofield Step
1	BAP parent(s)	BAP parent(s)	
2	BAP parent(s)	MAR partners	1. BAP-MAR [parents]
3	BAP child and parent(s)	BUR known child and parent(s)	5. BAP/MAR-BUR
4	MAR partner / BAP parent	BUR known husband/wife and spouse	2. MAR-BUR surviving spouse
5	BAP child not linked to bur	MAR non-stranger bachelor/spinster or status unknown	4. BAP-MAR
6	BAP child not linked to bur or mar	BUR adult status unknown	5. BAP/MAR-BUR
7	BUR known child and parent(s) not linked to bap	FAMILY parent(s) and child spacing	
8	MAR partner / BAP parent not linked to bur	BUR known adult	5. BAP/MAR-BUR
9	MAR partner / BAP parent where other partner linked to bur	MAR non-stranger widow/widower or bride/groom marital status	7 and 9. MAR-MAR
10	MAR groom / BAP father where other partner linked to bur	FAMILY	
11	MAR partner / BAP parent not linked to mar or fam	BUR adult status unknown	5. BAP/MAR-BUR
12	BAP not linked to bur or mar	BUR known adult not already linked	5. BAP/MAR-BUR

Source: For Wrigley and Schofield steps see pp. 69-70 in Wrigley E A and Schofield R S (1973):'Nominal record linkage by computer and the logic of family reconstitution' in Wrigley E A (ed) *Identifying People in the Past*, Edward Arnold, London.

Table 2: Age lookup table sample

Age	years	min	max
a Chrisome	0.08	0	0.1
an old man		55	80
1	1	0.5	2.5
1 1/2	1.5	1	2
1 month	0.08	0	0.25
1 quarter	0.25	0.1	0.5
12 yeres	12	10	15
126 yeres	126	70	116
14 weeks	0.26	0.1	0.4
15 monethes	1.25	1	1.5
aged 18	18	16	21
Aged 26 years	26	23	29
ladd		10	20
litle Child		0	7
man		20	70
man-child		0	1
weeks 8	0.15	0.08	0.25
yer 35	35	31	39
42	42	37	47
56	56	50	70
60	60	50	80
74	74	60	80
yer 80, old	80	60	90

Figure 3: Distinct name spellings in London parish registers

Legend:
- ☐ Forename strings
- ■ Surname strings
- ○ Number of baptisms in 1700 x 100

Y-axis: 0, 10000, 20000, 30000, 40000, 50000

X-axis categories: Cheapside (1538-1722), Clerkenwell (1558-1752), Aldgate (1550-1710)

Figure 4: Forename and surname distribution in Aldgate, London 1550-1710[12] (after name standardisation)

(a) Male forenames, showing the 10 most popular as shaded bars
(John, Thomas, William, Richard, Robert, James, Edward, Henry, George, Samuel)
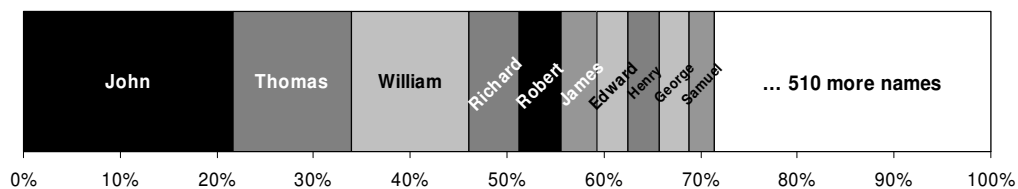
Source: 26,833 male children baptised in St Botolph Aldgate or Holy Trinity Minories, London. 55 children whose sex cannot be determined are omitted.

(b) Female forenames, showing the 10 most popular as shaded bars
(Elizabeth, Mary, Ann, Sarah, Susanna, Jane, Margaret, Catherine, Alice, Rebecca)
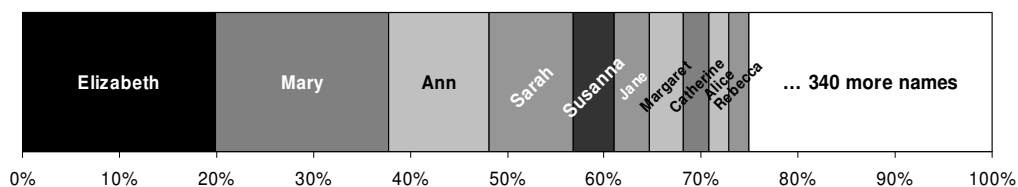
Source: 25,765 female children baptised in St Botolph Aldgate or Holy Trinity Minories, London. 55 children whose sex cannot be determined are omitted.

(c) Surnames, showing the 10 most popular as shaded bars
(Smith, Johnson, Clark, Jones, Brown, Green, Wright, Cook, Taylor, Davies)
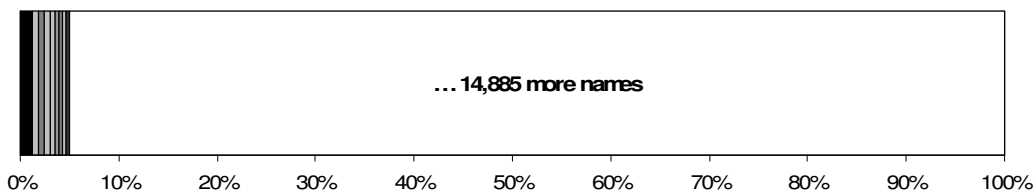
Source: 188,710 surname instances from the parish registers of St Botolph Aldgate and Holy Trinity Minories, London.

---

[12] The most popular male and female forenames are relatively invariant over time in this period. For decadal tabulations of the ten most popular names in a sample of English parishes see Appendix C pp. 191-201 in Smith-Bannister S (1997): *Names and Naming Patterns in England 1538-1700*, Oxford, Clarendon Press.

Table 3: Surname popularity in different types of English settlement

Aldgate 1550-1710 suburban London	Clerkenwell 1558-1752 suburban London	Cheapside 1538-1722 central London	Banbury, Oxon 1558-1837 large market town	Colyton, Devon 1538-1650 small market town	South Cambs 1539-1640 rural cluster	Colne, Lancs 1599-1653 rural, isolated
Smith	Smith	Smith	Smith	*Newton*	*Prime*	*Hartley*
Johnson	Jones	Brown	Taylor	Clarke	Taylor	*Hargreaves*
Jones	Brown	Taylor	Bull	*Marwood*	*Fuller*	Smith
Clark	Johnson	*Cage*	Baker	*Seward*	*Rayment*	*Emmett*
Brown	Williams	Allen	*Gardner*	*Stocker*	*Newman*	Robinson
% top 3.5	4.9	5.3	4.6	6.2	6.5	24.2
% top 1.2	0.9	1.2	1.4	1.8	1.6	9.9

Names are shaded relative to how popular they are in Aldgate. Names printed in black constitute at least 1 in 1000 names and rank in the top 90 names in Aldgate. Names printed in dark grey italic constitute at least 1 in 10,000 names and rank between 91st and 1475th in Aldgate. Names printed in light grey italic constitute at least 1 in 100,000 names and rank below 1475th in Aldgate.

Sources: Algate, Clerkenwell, and Cheapside: transcriptions of parish registers. Banbury: Cambridge Group family reconstitution based on parish registers. South Cambridgeshire cluster and Colne: Table 4 p.28 in Watson R (1975): 'A study of surname distribution in a group of Cambridgeshire parishes, 1538-1840', *Local Population Studies*, 15, pp. 23-32.

Table 4: Surname lookup table sample (complete phonetic group PLT /PLTS)

Surname	n	dmet1	syllcount	species	genus
Ballett	2	PLT	2	Ballatt	Ballatt
Ballatt	5	PLT	2	Ballatt	Ballatt
Ballet	1	PLT	2	Ballatt	Ballatt
Balitt	2	PLT	2	Ballatt	Ballatt
Belwoode	1	PLT	2	Bellwood	
Bellwood	2	PLT	2	Bellwood	
Bild	1	PLT	1	Bild	Bold
byld	1	PLT	1	Bild	Bold
billet	1	PLT	2	Billet	Ballatt
Billett	1	PLT	2	Billet	Ballatt
billetes	1	PLTS	2	Billet	Ballatt
Blade	1	PLT	1	Blade	Blatt
Blades	1	PLTS	1	Blades	Blade
Blatts	1	PLTS	1	Blades	Blatt
Blaides	2	PLTS	1	Blades	Blade
Blatt	1	PLT	1	Blatt	Platt
Bletsoe	1	PLTS	2	Bletsoe	
Beltsoe	1	PLTS	2	Bletsoe	
Blotsoe	1	PLTS	2	Bletsoe	
Bluitt	1	PLT	1	Blewitt	
Blewett	4	PLT	1	Blewitt	
Blout	1	PLT	1	Blewitt	Blatt
Blewet	1	PLT	1	Blewitt	
Blewit	2	PLT	1	Blewitt	
Blewitt	5	PLT	1	Blewitt	
Pluet	1	PLT	1	Blewitt	Platt
Blid	2	PLT	1	Blit	Bild
Blite	1	PLT	1	Blit	Blatt
Blit	1	PLT	1	Blit	Blatt
Bald	1	PLT	1	Bold	Bolt
Bould	1	PLT	1	Bold	Bolt
Bolde	1	PLT	1	Bold	Bolt
Bold	2	PLT	1	Bold	Bolt
Belt	1	PLT	1	Bolt	Bild
Boult	26	PLT	1	Bolt	Bolt
Bolte	7	PLT	1	Bolt	Bolt
Bolt	34	PLT	1	Bolt	Bolt
Bowlt	1	PLT	1	Bolt	Bolt
Boulte	9	PLT	1	Bolt	Bolt
Balte	1	PLT	1	Bolt	Bolt
Bolt	1	PLT	1	Bolt	Bolt
Balto	1	PLT	2	Bolto	Bolt
Boulto	1	PLT	2	Bolto	Bolt
Palladaye	1	PLT	3	Palladay	
Palladye	2	PLT	3	Palladay	
Palladay	3	PLT	3	Palladay	
Palleday	1	PLT	3	Palladay	
Palladey	1	PLT	3	Palladay	
Palleida	1	PLT	3	Palladay	
pallady	1	PLT	3	Palladay	
Pallatt	1	PLT	2	Pawlett	Ballatt
paillete	1	PLT	2	Pawlett	Pawlett
pailete	1	PLT	2	Pawlett	Pawlett
Paulett	6	PLT	2	Pawlett	Pawlett
Powlett	2	PLT	2	Pawlett	Pawlett
Paillett		PLT	2	Pawlett	Pawlett

Table 5: Forename lookup table sample (part of AN, ALKSNT and FRNSS groups)

Forename	n	sex	dmet1	regroup	newdmet	species	motherprob	fatherprob
Anna	262	f	AN		AN	Anna	0.00161247	
Ana	1	f	AN		AN	Anna	0.00161247	
Anah	1	f	AN		AN	Anna	0.00161247	
Ann	4212	f	AN		AN	Ann	0.1226552	
Anne	2951	f	AN		AN	Ann	0.1226552	
An	2683	f	AN		AN	Ann	0.1226552	
Anne	139	f	AN		AN	Ann	0.1226552	
Ane	5	f	AN		AN	Ann	0.1226552	
Anne	1	f	AN		AN	Ann	0.1226552	
Owen	68	m	AN		AN	Owen		0.00109853
Ouen	6	m	AFN	Owen	AN	Owen		0.00109853
Owin	3	m	AN		AN	Owen		0.00109853
owen	1	m	AN		AN	Owen		0.00109853
Owine	1	m	AN		AN	Owen		0.00109853
Eyoine	1	m	AN		AN	Owen		0.00109853
Iohn	4	m	AN	John	JN	John		0.20748976
Elexandor	1	m	ALKSNT		ALKSNT			0.00420736
Elexsander	1	m	ALKSNT		ALKSNT			0.00420736
Alexander	305	m	ALKSNT		ALKSNT			0.00420736
Alleckcandor	1	m	ALKKNT	Alexander	ALKSNT			0.00420736
Alxander	1	m	ALKSNT		ALKSNT			0.00420736
Alexandor	1	m	ALKSNT		ALKSNT			0.00420736
Elexander	7	m	ALKSNT		ALKSNT			0.00420736
Elexsandur	1	m	ALKSNT		ALKSNT			0.00420736
Allexander	52	m	ALKSNT		ALKSNT			0.00420736
Allexannder	2	m	ALKSNT		ALKSNT			0.00420736
Alectsander	1	m	ALKTSN	Alexander	ALKSNT			0.00420736
Elixander	1	m	ALKSNT		ALKSNT			0.00420736
elyssannder	1	m	ALSNTR	Alexander	ALKSNT			0.00420736
Alexand	1	m	ALKSNT		ALKSNT			0.00420736
Alaxander	1	m	ALKSNT		ALKSNT			0.00420736
Ellexander	2	m	ALKSNT		ALKSNT			0.00420736
Afexannder	2	m	ALKSNT		ALKSNT			0.00420736
ffraunces	1	m	FRNSS		FRNSS			0.01632411
ffrancys	2	f	FRNSS		FRNSS		0.01675177	
ffrancis	206	x	FRNSS		FRNSS		0.01675177	0.01632411
Frauncys	24	x	FRNSS		FRNSS		0.01675177	0.01632411
Frances	891	x	FRNSS		FRNSS		0.01675177	0.01632411
ffran:	1	m	FRN	Francis	FRNSS			0.01632411
Franscis	1	f	FRNSS		FRNSS		0.01675177	
Franncis	211	x	FRNSS		FRNSS		0.01675177	0.01632411
Frannces	8	x	FRNSS		FRNSS		0.01675177	0.01632411
frances	25	x	FRNSS		FRNSS		0.01675177	0.01632411
ffrances	152	x	FRNSS		FRNSS		0.01675177	0.01632411
Francys	7	x	FRNSS		FRNSS		0.01675177	0.01632411
francis	45	x	FRNSS		FRNSS		0.01675177	0.01632411
Francis	1240	x	FRNSS		FRNSS		0.01675177	0.01632411
Fransces	2	f	FRNSS		FRNSS		0.01675177	
Francis	2	m	FRNSS		FRNSS			0.01632411
Ffrauncys	27	x	FRNSS		FRNSS		0.01675177	0.01632411

Table 6: Occupations lookup table sample

Group	Subgroup	Category	OccDescriptor	n
administration	gentlemen, officials	gentleman	a devonshire Gentleman	1
administration	gentlemen, officials	custom house man	Customhouse man	7
administration	gentlemen, officials	doctor of medicine	doctor of Phisique	5
administration	gentlemen, officials	doctor of law	doctor of the Cyvill Lawe	2
administration	school	schoolmaster	schoolemaister	27
transport and military	shipping	hoy master	Hoyman	5
transport and military	shipping	lighterman	Lighterman	21
transport and military	shipping	sailor	Seaman	590
transport and military	shipping	sailor	Sailer	431
transport and military	conveyance	carman	Carman	415
transport and military	conveyance	drayman	drayman	207
food, drink, housekeeping and hospitality	brewing	beer brewer	Beerebrewer	17
food, drink, housekeeping and hospitality	brewing	ale brewer	Ale brewer	4
food, drink, housekeeping and hospitality	brewing	brewer	Brewer	188
food, drink, housekeeping and hospitality	brewing	brewer's servant	Brewers servant	154
food, drink, housekeeping and hospitality	innkeeping	innkeeper	Inholder	8
food, drink, housekeeping and hospitality	innkeeping	tapster	Tipler	82
manufacture and selling on	construction and ship outfitting	carpenter	Carpenter	625
manufacture and selling on	construction and ship outfitting	ship's carpenter	Shipcarpenter	3
manufacture and selling on	weaponry	gun maker	Gunsmith	325
manufacture and selling on	weaponry	gun maker	gonnemaker	72
manufacture and selling on	weaponry	gun maker	gonnefounder	11
manufacture and selling on	weaponry	crossbow maker	Cross-bowmaker	3
manufacture and selling on	weaponry	armourer	Cittizen & Armorer	2