

---

## DEBATES

---

### *Measurement of the Change in Populations Through Time: Capture-Recapture Analysis of Population for St. Lawrence Parish, Southampton, 1454 to 1610\**

T. B. James - N. A. Price

St. Andrews University, Fife, Scotland

« Migration is becoming the most important branch of demography, just as fertility has dominated the scene for the past 80 years or so, and mortality did before that ».<sup>1</sup> This recent statement is qualified by the admission that little is known about the subject of migration before 1850. The major problem with the study of migration is that the nature of the source material, unlike parish registers for fertility and mortality, virtually defies quantification. Information relating to migration is scattered through a wide range of differing sources. Even in cases where information can be gleaned on migration it is rare to obtain a picture of mobility even for single individuals during their lifetime. The mid-XIXth century census returns which furnish the first detailed information on origins give no indication of residence between a birthplace and residence at the time the census was taken.

Research into origins and mobility in pre-industrial society is in its infancy. Various sources furnish information on the subject: consistory court deposition books, deposition books from the proceedings in the High Court of Admiralty, probate material and property deeds. Quantification of such data is almost impossible, although attempts have been made to

---

\* In the preparation of this paper much valuable assistance and advice has been freely given by Professor R.M. Cormack, Department of Statistics, University of St. Andrews, and by Dr. N.G. Parker, Department of Modern History, University of St. Andrews.

<sup>1</sup> T.H. HOLLINGSWORTH, *Migration*, in « Historical Demography; Problems and Projects », prepared by Michael Drake (Open University Press, 1974), 170.

study individual sources in depth.<sup>2</sup> Little work has been completed on mobility and migration in individual communities; in the current study of Southampton as a whole from 1400 to 1600 there is considerable evidence of mobility, perhaps an exceptional amount considering the role of the town as an international port and communications centre. Although the idea that pre-industrial society was static dies hard,<sup>3</sup> overwhelming evidence is now being produced to indicate that considerable changes took place through migration and replacement in the composition of societies over time.

It is this aspect of mobility and migration, the change which took place in communities over time, which is analysed in this through a study of a series of nominal listings compiled for a single parish, for different purposes and naming varying proportions of the population. Study of nominal listings over time shows the continuous process of change amongst those listed.

In the crudest form some idea of the change which occurred can be gleaned from a study of surname change, for example 90% of all the surnames which appear in the parish listings of 1454 in Southampton had disappeared by the end of the century. What would seem to be required is a method whereby the change in the composition of parish communities could be measured with some accuracy.

Capture-recapture analysis, as described below, would appear to offer the opportunity of measuring the rate of change, or "birth" and survival, within a community by reference to a core, or sub-population, of individuals who appear regularly in listings compiled at different dates and for different purposes.

The tax listings considered here refer to the parish of St. Lawrence, Southampton over the period from 1454 to 1610. Of the listings which survive for the six parishes of Southampton those for St. Lawrence particularly lend themselves to analysis. St. Lawrence was the second smallest parish in medieval and early modern Southampton, containing some forty dwelling plots. As far as is known all the properties faced onto the High Street, although there may have been some dwellings behind the street frontage. The area of St. Lawrence was built up as were the parishes immediately bordering upon it, while it was bounded to the east by the town wall and ditch, which situation allowed little room for expansion, unlike the parish of All Saints which spread into the area outside the walls to the north of the town. On account of its enclosed position the population of St. Lawrence

---

<sup>2</sup> See for example P. CLARK, *The Migrant in Kentish towns 1580-1640*, in «Crisis and Order in English Towns», ed. P. Clark and P. Slack (London 1972), pp. 117-163. This study is based entirely on court deposition material.

<sup>3</sup> J. R. HALE, *Renaissance Europe, 1480-1520*, (London 1971), p. 32. It is postulated that the vast majority of people in both eastern and western Europe were born, married and died within sight of the same church or wood.

remained remarkably constant from the middle ages to the XIXth century. The earliest full count of heads taken for the purposes of the muster in 1596 gave the population of the parish as 370, out of a population total of 4,200. The census of 1851 listed a population of 364 out of a total population of 34,000, including a number of households which included large groups of employees living in, which would help explain the large average size of household which is obtained from dividing the total population of the parish by the number of households.<sup>4</sup> The constancy of the population size offers the opportunity to expand this study of turnover beyond the terminal date of 1610 chosen for this study, after which there is a considerable gap in the survival of population listings.

The parish of St. Lawrence was comparatively wealthy, consisting of the private dwellings of well-to-do townspeople containing servants, hostelry accommodation and the shops and workshops of merchants and artisans well placed on the main street of the town. It would seem likely that the tax listings therefore represent a large sample of the total heads of household in the parish.

Thirty-five tax listings survive for the period from 1454 to 1610. They contain in the majority of cases some 30 to 40 names. Capture-recapture analysis permits the examination of all the listings surviving for the parish, subject to certain assumptions which are set out below.<sup>5</sup> It is possible to examine lists which contain differing numbers of names, and which were compiled for different purposes provided that a consistent section of the population is listed.<sup>6</sup>

It is this capability to take listings, and look at them in terms of a rigid statistical model (although there are problems in so doing) which is the important contribution which capture-recapture analysis can make to the study of historical demography, in particular in the investigation of population turnover.

Due to certain inadequacies inherent in the present data, we make little claim for the importance of the historical conclusions which may be drawn from this particular study. We are convinced, however, that the method is of considerable potential value in demographic studies. Provided the assumptions<sup>7</sup> hold good, the larger the population of towns under review, the more reliable will be the results of such studies.

### 1. *The data*

Three problems were encountered in the preparation of the data. Firstly the lists were compiled for different purposes, comprising parliamentary

---

<sup>4</sup> PRO, HO 107/1669; SRO SC 3/1/1.

<sup>5</sup> See below, pp. 194-5.

<sup>6</sup> The references to the lists are found in Appendix 2, table 1 below.

<sup>7</sup> See below, 194-5.

taxes, local taxes and muster lists. The muster lists were not fiscal taxes like the others, but rather listings of males between the ages of sixteen and sixty able to bear arms. However, identification of the group who appear in the tax lists was possible because the muster lists delineated individual households under their head and also because the heads of household who were too old or otherwise unable to bear arms were listed, being responsible for providing arms and producing those who were able to bear arms from their households. When the different listings were set side by side, notwithstanding fluctuations in the number of people listed on different occasions, it was apparent that they formed a readily comparable data set.

The second problem was to edit out of the listings various sub-populations which did not appear consistently, to leave the male heads of household. Three sub-populations were removed: servants, widows and aliens. The most numerous group was the servants, a term which included sons, household servants and apprentices, who appeared in greatest profusion in the muster lists although occasionally servants were mentioned in tax lists.

In no case did a servant appear as a head of household. Widows were removed as their characteristics as heads of household were inconsistent with the group of male heads — for example a widow might have remarried and remained in the population but have disappeared as a head of household, or being of a similar age to her husband a widow might be expected to die soon after her husband, thus inflating the number of people who appeared briefly. Aliens tended to pay a poll tax rather than qualify for taxation as the possessors of goods of a specific value, and were in general excluded from the muster. However, it was noted that when a householding widow or alien was eliminated from the data, this effectively reduced the actual population by one. This is only relevant in formulating the conclusions where allowance is made, at one point, for the fact that the total population is approximately stable over the whole period because of the fixed location of the parish. The widows and aliens were small groups compared with the servants who appeared in the muster, and so any adverse effect of their withdrawal from the data, would be negligible. This editing reduced the total names by 27% leaving 396 male heads of household who appeared over 160 years.

The third and greatest problem in data preparation was the accurate dating of the listings, which survived haphazardly through the period. Any survival rate, or "birth" (in the sense of appearing for the first time) rate, could be estimated with some certainty only if a date could be attributed to each list.

Since the earliest list (1454) and the second list (1500) were sufficiently far apart for no individual to appear in both lists, the list for 1454 was not relevant to the analysis. After 1500 the listings were more frequent and so required where possible a month in addition to a year for dating, as the

survival rate ( $\hat{O}$ ) and the immigrant rate ( $\hat{B}$ ) were to be expressed on an annual basis as far as possible. An error of twelve months in dating could cause serious discrepancies in the numerical results. Moreover, occasionally more than one list survived for the same year, three lists surviving for 1593. Meticulous care was taken to ensure that all the lists were in the right order. In considering lists where dating was given only by year, and not by month, in the absence of any more accurate dating information, such as the death and probate of an individual who had been taxed previously, the mid-point of the period within which the list was made was taken. Clearly a margin of error exists in such cases.

The analysis of the data and the conclusions reached are set out in the succeeding sections.

## 2. Analysis of the data.

The principle of capture-recapture analysis and the equations which produce the required estimates are stated in R. M. Cormack, *The Logic of Capture-Recapture Estimates*, « *Biometrics* », June 1972. It is not intended for the purposes of this paper to reiterate the mathematical arguments in full. The essential point is that the following analysis is based on incomplete data, and therefore any answers which may be derived from the analysis are *estimates* which are subject to a certain degree of error.

The aim of statistical analysis is to estimate features of a population which are not directly observable (called parameters) from observed data. One's faith, or lack of faith, in any estimate can be expressed as a "confidence interval", that is the range of values which may be accepted with a stated degree of confidence, containing the true value. The degree of confidence with which this statement can be made can be calculated and expressed as a percentage. Thus, each estimate made in this analysis (a survival rate and a number of "births" for each list)<sup>8</sup> has a confidence interval attached to it based on a stated level of confidence. It is desirable to have as small a confidence interval for as high a degree of confidence, as possible.

The confidence interval for each estimate can only be calculated from its variance, this calculation being complicated even after the derivation of the basic equation for the variance. For the sake of simplicity, only the estimates have been shown for the present analysis, but it should be realised that these estimates each represent a range and if this range is too large very little confidence can be attached to the estimate.<sup>9</sup> The size of the range depends on a number of factors, but principally on the lack of completeness

---

<sup>8</sup> See below, page 724.

<sup>9</sup> See Appendix 1 for examples of certain confidence intervals and the degree of confidence held in such results.

of the data, and the failure of the data to satisfy certain rigid conditions specified in the statistical model.

The full method, and derivation of the equations used, is set out in Appendix 1, and the tables of figures extracted from the Southampton data in Appendix 2. The estimates themselves can be derived intuitively from certain parameters that can be obtained for each list from the data. The symbol  $\hat{\phantom{x}}$  over any letter indicates that the quantity in question is an estimate. The two quantities required to be estimated are the survival rate ( $\hat{\phi}$ ) *i.e.* the probability<sup>10</sup> that an individual survives from one list to the next, and the "birth" rate ( $\hat{B}$ ) *i.e.* the number of new male heads of household who enter the population between two successive listings.

However, to secure these estimates further estimates are necessary. Firstly, the probability must be estimated of any live individual, *i.e.* male head of household in St. Lawrence parish, being registered ( $\hat{P}$ ), secondly, the number still alive that have already been registered ( $\hat{M}$ ) and thirdly, the total size of the population ( $\hat{N}$ ). All these estimates are required for each list.

The estimates ( $\hat{P}$ ,  $\hat{M}$ ,  $\hat{N}$ ,  $\hat{\phi}$ , and  $\hat{B}$ ) are calculated for each list and each estimate will have its own confidence interval, based on its variance. The estimates are calculated from four sets of numbers ( $n_1$ ,  $m_1$ ,  $z_1$  and  $r_1$ ) each of which must be counted from the data for each list. With a total of 35 lists, a large number of calculations are required for which a computer programme is advantageous.

The estimates of particular interest are the survival rate and the "birth" rate, but these are both dependent on the time interval between successive lists. After the lists had been dated as precisely as possible, the time element was eliminated and the estimates expressed on an annual basis, *i.e.* the probability of surviving for a certain year, and the number of "births" in that year.<sup>11</sup> Where gaps of several years occur between lists, estimates for the intervening years are identical. The results are given in table 4 of the calculation of estimated survival and estimated "births".

### Assumptions

The application of a rigid statistical model to imprecise data undoubtedly causes problems of interpretation. Apart from the importance of the point made above that the results of the analysis are estimates, rather than firm figures, certain other assumptions have to be made and difficulties pointed out.

Firstly, it is assumed that, in a particular list, the estimate for the probability that any individual is registered ( $\hat{P}$ ) is an average figure for those

<sup>10</sup> Probability is measured between 0 and 1, the value 0 representing an impossibility and the value 1 a certainty.

<sup>11</sup> See below, Appendix 2, table 1 and 4.

already registered. Naturally the lower this average figure the less reliable the results will be. Secondly, the estimate ( $\hat{N}$ ) of the population size assumes that the probability of being registered in a particular list is the same for all individuals, whether they have been registered before or not. Thirdly, the estimate ( $\hat{O}$ ) for the survival rate refers only to the people who have been registered. Fourthly, the estimate ( $\hat{B}$ ) for the number of "births" or immigrants assumes that the average survival of all individuals between one list and its successor, is the same whether or not the individuals have been registered. Fifthly, it is assumed that all registered individuals must have the same probability of appearing in some subsequent list.

It is, therefore, assumed throughout that the behaviour of the sub-population of registered individuals is the same as that of the whole population of male heads of household in St. Lawrence parish. There is no evidence to suggest that the above assumption should not be justified regarding the population listings used. In this particular example, with so many possible causes of error, it would be advisable to calculate the variances of the estimates in order to examine the plausibility of the estimates obtained. Even without the variances, however, trends appear which suggest certain interpretations.<sup>12</sup> The arguments from which the variance formulae are derived are outlined elsewhere.<sup>13</sup>

Thus, when applying capture-recapture analysis to similar data, the assumptions outlined above must be satisfied in full. The computer programme has been designed for general use and is available if required.<sup>14</sup>

### 3. *The results and their interpretation*

Any conclusions drawn are, as stated above, by no means concrete, and should be interpreted with due regard to the limitations inherent in the data. On the other hand the survival of historical data of all kinds is haphazard: it is not possible to know to what extent vital information is missing in any historical study.

In considering the survival rate figures, it appears that there are very low results for the years 1546-47, 1575 and 1592. Evidence of outbreaks of plague in these particular years is lacking for Southampton, although some kind of minor epidemic may have occurred in 1592-93 as the parish register of St. Michael, Southampton shows a slight excess of burials over baptisms. Such excesses are however, not reflected in the results of this analysis in other years in the late XVIth century when they are known to have occurred.

---

<sup>12</sup> See below, this page.

<sup>13</sup> G. A. SEBER, *Estimation of Animal Abundance*, (London 1973), chapter 5.

<sup>14</sup> The program is filed at the Computing Laboratory, University of St. Andrews.

The low figures for survival in 1546-47, 1575 and 1592 therefore remain something of a mystery, perhaps explicable only in terms of the inadequacy of the data. It is noteworthy that the circumstances of the high turnover rate in the years mentioned above varied in that the population of heads of household remained fairly constant in 1546-47, whereas it dropped appreciably to the figures recorded for 1575 and 1592.<sup>15</sup>

By chance, few listings were taken in St. Lawrence around the period of the influenza epidemic of 1558 and the plague epidemic of 1563, or in the early 1580s, there being an outbreak of plague in 1583. Such a lack of lists shows up in the results as an extended period of low survival rather than a sudden sharp fall as would undoubtedly have appeared had listings survived closer to the period of the epidemics. The worst plague in the period under consideration was that of 1604. More regular lists exist for the beginning of the XVIIth century, but the predictable low figure in the results appears in the period 1602-03 rather than in 1604 for which there is a list. This could be for a variety of reasons. Firstly, an error may have occurred in the dating of the lists. Secondly, it may have been that the plague which was so disastrous in St. Michael parish and in the Walloon community may have struck less fiercely in St. Lawrence, which was small and wealthy, than in the poorer areas of the town, particularly St. Michael. Thirdly, the plague may have affected a younger age group than the heads of household: the Walloon register in particular shows a large number of young people being carried off in the 1604 plague.<sup>16</sup>

It can be seen that the figures for "births" into the population, shown in Table 4, are highly volatile with large oscillations. If the period from 1501 to 1604 is taken as a whole the overall trend tends towards an increase in the number of "births" per year. A moving average curve of the annual number of births, while not entirely reliable, would display this general increase more clearly.

A further noteworthy point arose from the comparison of the survival ( $\hat{O}$ ) and "birth" ( $\hat{B}$ ) figures. From 1501 to 1551 the two sets of figures move in entirely opposite directions, that is to say an increase in "births" accompanies a reduction in survival rate and *vice versa*. From 1551 to 1577 the exact opposite occurs: an increase in "births" accompanies an increase in survival rate and *vice versa*.<sup>17</sup> From 1577 to 1590, without exception, the behaviour recorded in the period 1501 to 1551 recurs, while in the fourth period, from 1591 to 1604 the oscillation appears at first sight to be random. Closer examination reveals that the fluctuations bear a marked

---

<sup>15</sup> See below, Appendix 2, tables 1 and 2.

<sup>16</sup> SRO, PR 7/1/1, register of St. Michael parish, Southampton; *Registre de l'église Wallonne de Southampton*, ed. H. M. Godfray (Publications of the Huguenot Society of London, iv, 1890), 107-110.

<sup>17</sup> There is an exception to the pattern in 1572.

resemblance to the period from 1551 to 1577, moving parallel, apart from an occasional lag of one year in their movement. Thus four periods of alternating population trend can be discerned. Given a stable population size, as was the case in St. Lawrence parish over the period from 1501 to 1604,<sup>18</sup> the behaviour of the first and third periods is predictable, a high "birth" rate replacing the positions vacated by the low survival rate. It is possible that during the second and fourth periods the population size was less stable, although the estimates for  $N$  do not show this clearly. The possibility cannot be ruled out. An investigation of the correlation between the survival and "birth" rates could be carried out, and also investigation of the effect of survival and "birth" rates on the population size but it was felt that in this study the variability was too great for these to be informative.

The main conclusion of the whole exercise is that capture-recapture analysis, which is devised for incomplete data sets, is a method of wide application in the field of historical demography for the estimation of the size of populations and sub-populations. Strong emphasis has been laid on the limitations and assumptions contained in the method which the data must satisfy. It is encouraging that the data for St. Lawrence parish, Southampton between 1454 and 1610, with its limitations, can produce results. The results shed some light on the formerly intractable problem of the measurement of population turnover, or survival, while the method strongly recommends itself for application to comparable data sets, for comparative purposes. At later periods than those under discussion, when dating is more certain and listings taken more regularly, the data would satisfy the rigid statistical conditions more fully: a more reliable picture of "births" and immigrants in the population could be obtained.

Even as it stands the method of capture-recapture analysis makes a significant contribution to the field of the quantification of demographic data, especially as it is capable of application to data from the pre-industrial era in which field little has been achieved so far.

---

<sup>18</sup> See above, pp. 726-727.

APPENDIX 1

THE ANALYSIS OF THE DATA

The principles behind capture-recapture and the equations which bring forth the required estimates are stated by CORMACK in *The Logic of Capture-Recapture Estimates*, « *Biometrics* », June 1972. It would perhaps be helpful to go through the notation, equations and assumptions, as applied to this problem.

The following notation was used:

- $P_i$  the probability that any live individual, *i.e.* male head of household in St. Lawrence parish, Southampton, is registered in the  $i^{\text{th}}$  list.
- $m_i$  the number of individuals registered in the  $i^{\text{th}}$  list that have already been registered.
- $M_i$  the number at the time of the  $i^{\text{th}}$  list that have already been registered.
- $N_i$  the total size of the population at the time of the  $i^{\text{th}}$  list.
- $\phi_i$  the survival rate between the  $i^{\text{th}}$  and the  $(i+1)^{\text{th}}$  lists.
- $n_i$  the total number of individuals in the  $i^{\text{th}}$  lists, all of whom are "returned" to the population after the  $i^{\text{th}}$  list was taken.
- $B_i$  the number of immigrants (*i.e.* new male heads of households) between the  $i^{\text{th}}$  and  $(i+1)^{\text{th}}$  lists.
- $z_i$  the number of individuals that are registered before and after the  $i^{\text{th}}$  list but not in it.
- $r_i$  the number of individuals that are registered in the  $i^{\text{th}}$  list and are subsequently registered in any other list.

The principal estimate is that for  $M_i$ , which as Jolly points out,<sup>1</sup> has an intuitive derivation, in the same way as the others. To calculate this estimate the number of individuals present and registered in the population, but not in the particular list under consideration must be examined, *i.e.*  $M$  minus the number of individuals in a list that have appeared in any previous list (the set of numbers  $m_i$ , to be counted from the data).<sup>2</sup> It can be seen that this quantity is equal to the number of individuals not in a particular list, but registered previously and subsequently to it (the parameter  $z_i$  to be counted from the data) divided by the probability that an individual in the list is registered again after it. This probability is equal to the number in the list who appear again subsequently (the parameter  $r_i$  to be counted from the data) divided by the number in the list (the parameter  $n_i$  to be counted from the data).

Thus we have:

$$(\hat{M}_i - m_i) = \frac{z_i}{(r_i/n_i)} \quad \text{i.e.} \quad \hat{M}_i = m_i + \frac{n_i z_i}{r_i}$$

The estimate  $\hat{P}_i$  is merely the number of individuals in the list that have already been registered ( $m_i$ ) divided by the number in the population at the time of the list that have been registered ( $\hat{M}_i$ ).

$$\hat{P}_i = \frac{m_i}{\hat{M}_i}$$

<sup>1</sup> G. M. JOLLY, *Explicit estimates from capture-recapture data with both death and immigration-stochastic model*, « *Biometrika* », lli (1963), 225-247.

<sup>2</sup> The suffix "i" denotes the number of the actual list under review, *i.e.*  $i = 1, 2, \dots, 35$

*Measurement of the Change in Populations Through Time*

The estimate  $\hat{N}$  for the total size of the population is the size of the list ( $n_i$ ) divided by the probability that an individual appears in that list: ( $\hat{P}_i$ ).

$$\hat{N}_i = \frac{n_i}{\hat{P}_i} = \frac{n_i \hat{M}_i}{m_i}$$

The estimate for the survival rate  $\hat{O}$  for a certain list (*i.e.* the probability of an individual surviving to the next list), is given by the estimate at the time of the next list of those that have been registered, divided by the estimate, immediately after the present list, of those that have been registered *i.e.* the proportion of those already registered immediately after the present list who are alive at the time of the next list.

$$\hat{O}_i = \frac{\hat{M}_{i+1}}{\hat{M}_i + n_i - m_i}$$

Finally the estimate  $\hat{B}$  the number of newcomers, or "births", between the present list and the next list is given by the estimate of the total population times the survival rate.

$$\hat{B}_i = \hat{N}_{i+1} - \hat{O}_i \hat{N}_i$$

Thus throughout, except for  $\hat{M}_i$ , estimates are used to derive further estimates, and thus the concept of estimation and the errors involved must be constantly borne in mind.

The data under review was arranged in a  $35 \times 396$  array, 35 being the number of lists and 396 the number of individual names. A "1" was entered where a name appeared in a certain list, otherwise the spaces were left blank. A computer programme was written to calculate the estimates.<sup>3</sup>

The results of the parameter counting are given in appendix 2 (table 2), below, and the calculated estimates are given in table 3.

The programme is also capable of calculating all the variances and covariances or degree of interaction and interdependence between estimates, but as stated above, the variances are not included in this paper.

To illustrate how the estimates are calculated, the case can be examined where  $i = 30$ , that is to say an examination of the thirtieth list in the data, dated September 1601.<sup>4</sup>

For this list the values are as follows:

$$\begin{array}{ll} n_i = 41 & m_i = 25 \\ r_i = 34 & z_i = 3 \end{array}$$

Thus there are 41 names in the list in question, of which 25 have appeared in previous lists, and 34 appear in some subsequent list. Also there are 3 names not appearing in this list that do appear both prior and subsequently to it. To calculate the estimates applicable to this list it is also necessary to know the parameter values for the  $(i + 1)^{\text{th}}$  list, that is to say list 31.<sup>5</sup>

These are:

$$\begin{array}{ll} n_{i+1} = 40 & m_{i+1} = 30 \\ r_{i+1} = 33 & z_{i+1} = 7 \end{array}$$

<sup>3</sup> Grateful acknowledgement is due to the Computing Laboratory, University of St. Andrews, for writing and implementing the program.

<sup>4</sup> SRO SC 5/17/12.

<sup>5</sup> See Appendix 2, table 2.

By substituting the values for  $i = 30$  into the equation

$$\hat{M}_i = m_i + \frac{n_i z_i}{r_i}$$

we have  $\hat{M}_{30} = 25 + \frac{41 \times 3}{34} = 28.62$  (cf. App. 2, table 3, list 30)

Also for  $i = 31$ ,  $\hat{M}_{31} = 30 + \frac{40 \times 7}{33} = 38.48$  (cf. App. 2, table 3, list 31)

From  $\hat{N}_i = \frac{n_i \hat{M}_i}{m_i}$ , by substitution we have:

$$\hat{N}_{30} = \frac{41}{25} \times 28.62 = 46.93$$
 (cf. App. 2, table 3, list 30)

and  $\hat{N}_{31} = \frac{40}{30} \times 38.48 = 51.31$  (cf. App. 2, table 3, list 31)

From the other three estimates are obtained by further substitutions into the relevant equations:

(a)  $\hat{P}_i = \frac{m_i}{\hat{M}_i}$  giving  $\hat{P}_{30} = \frac{25}{28.62} = 0.8736$

(b)  $\hat{O}_i = \frac{\hat{M}_{i+1}}{\hat{M}_{i+1} + n_i - m_i}$  giving  $\hat{O}_{30} = \frac{38.48}{28.62 + 41 - 25} = 0.8625$

(c)  $\hat{B}_i = \hat{N}_{i+1} - \hat{O}_i \hat{N}_i$  giving  $\hat{B}_{30} = 51.31 - (0.8625 \times 46.93) = 10.83$

Consequently, an extract of table 3 can be reconstructed for list 30:

List	$\hat{M}_i$	$\hat{P}_i$	$\hat{N}_i$	$\hat{O}_i$	$\hat{B}_i$
30	28.62	0.8736	46.93	0.8625	10.83

That is to say, at the time of the thirtieth list (Sept. 1601), the number of individuals that had already been registered is estimated at 28.62, the probability that any live individual is listed is estimated at 0.8736, the total size of the population is estimated at 46.93, the probability that any live individual will survive till the time of the next list is estimated at 0.8625, and the number of new individuals entering the population before the next list is estimated at 10.83. It is again emphasized that individuals in the population are only the male heads of household. All other estimates are calculated in a similar manner.

To show how the time element was eliminated lists 29 and 30 can be examined. The two lists were taken in July 1599 and September 1601, *i.e.* they are 26 months apart. The estimate of  $B_i$  for list 29 is 16.65, that is to say there were 16.65 new heads of household appearing in St. Lawrence between July 1599 and September 1601. Thus, on average, the figures for the twelve months of 1600 would be:

$$\frac{16.65}{26} \times 12 = 7.68$$
 (cf. table 4)

*Measurement of the Change in Populations Through Time*

For the survival rate the process is more complex, as probabilities of surviving successive periods are multiplicative and not additive.<sup>6</sup> The reading for  $\hat{O}_i$  in list 29 is 0.8517, *i.e.* the probability of a male head of household surviving between July 1599 and September 1601 is estimated at 0.8517. This figure can be seen as the product of 26 equal monthly survival rates with a value  $p$ .

Therefore we have:  $p^{26} = 0.8517$   
*i.e.*  $p = \sqrt[26]{0.8517}$

Thus for the 12 months of 1600, the survival rate is:

$$(\sqrt[26]{0.8517})^{26} = 0.928 \quad (\text{cf. table 4})$$

Consequently an extract of table 4 can be reconstructed for 1600:

Date	Survival rate	Number of "births"
1600	0.928	7.68

That is to say, in the year 1600, it is estimated that there was a probability of 0.928 that a male head of household would survive (*i.e.* not die, or leave the parish) for the year and that 7.68 male heads of household would "arrive" in the population (be it by immigration, inheritance, or whatever).

It must be remembered that all these figures are estimates only.

Each estimate represents, in fact, a range of values known as a confidence interval.<sup>7</sup> To show this, we shall examine the estimates for the thirtieth list, and construct confidence intervals within which it can be said with a 95% degree of confidence, that the true values lie. These ranges of values are centred on each estimate, the outer limits being determined by taking the square roots of the variances attached to each estimate and multiplying these by the factor 1.96 (a number which is specified in all 95% confidence intervals, but whose derivation is complicated). The relevant variances of parameters calculated for the thirtieth list are:

Variance $\hat{O}_{30}$ 0.004056	Variance $\hat{N}_{30}$ 26.50	Variance $\hat{B}_{30}$ 8.741
-------------------------------------	----------------------------------	----------------------------------

Thus the 95% confidence intervals are:

(i) for  $\hat{O}_{30}$   
 $0.8265 \pm 1.96 \times 0.004056$ , *i.e.* 0.738 to 0.988

that is to say, it is 95% certain that the probability of an individual surviving from the time of thirtieth list to the thirty-first is between 0.738 and 0.988

(ii) for  $\hat{N}_{30}$   
 $46.93 \pm 1.96 \times 26.50$ , *i.e.* 36.84 to 57.02

that is to say, it is 95% certain that the actual population at the time of the thirtieth list numbered between 36.84 and 57.02

(iii) for  $\hat{B}_{30}$   
 $10.83 \pm 1.96 \times 8.741$ , *i.e.* 5.03 to 16.63

that is to say it is 95% certain that between the thirtieth and thirty-first lists, there were between 5.03 and 16.63 new members of the population.

<sup>6</sup> For example when tossing an unbiased coin the probability of obtaining a "head" is  $\frac{1}{2}$  (a one-in-two chance), but the probability of obtaining two "heads" consecutively is  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ .  
<sup>7</sup> See above p. 723.

Because comparisons are likely to be made between parameter estimates in successive lists, it is perhaps relevant to illustrate the confidence intervals of differences between such parameters. In this connection lists 29 and 30 were compared.

The method of constructing these intervals is more complicated than that outlined above, so these intervals shall simply be quoted to demonstrate the point.

(i) The difference between  $\hat{O}_{29}$  and  $\hat{O}_{30}$  is 0.0108; however, to this value is attached a 95% confidence interval of “-0.214 to 0.236”; that is to say it is 95% certain that the actual difference between these two parameters lies between -0.214 and 0.236.

(ii) The difference between  $\hat{N}_{29}$  and  $\hat{N}_{30}$  is 11.38; however, to this value is attached a 95% confidence interval of “1.94 to 20.82”; that is to say, it is 95% certain that the actual difference between these two parameters lies between 1.94 and 20.82.

(iii) The difference between  $\hat{B}_{29}$  and  $\hat{B}_{30}$  is 5.82; however, to this value is attached a 95% confidence interval of “-3.93 to 15.57”; that is to say, it is 95% certain that the actual difference between these two parameters lies between -3.93 and 15.57.

As can be seen from all six confidence intervals calculated above, large intervals seem to be prevalent, indicating the inadequacy of this data. The above calculations are included merely to illustrate the concept of the confidence interval in this context.

APPENDIX 2

TABLE 1

DATES FOR EACH LIST, TAKEN AT THE BEGINNING OF THE GIVEN MONTH

List Date

1	1454	SRO SC 13/1 <sup>a</sup>
2	1500	SRO SC 14/2/3
3	1524	PRO E 179/173/175
4	1543	PRO E 179/173/228
5	1545	PRO E 179/173/239
6	1548	SRO SC 15/17 <sup>a</sup>
7	1552	PRO E 179/239/161
8	1552	SRO SC 10/1/1
9	1556	SRO SC 13/2/2
10	1563	SRO SC 14/5
11	1567	SRO SC 13/2/3
12	1570	SRO SC 13/2/7
13	1571	PRO E 179/174/387
14	1573	SRO SC 13/2/4
15	1575	SRO SC 5/17/1
16	1576	SRO SC 5/17/1
17	1578	SRO SC 14/2/6
18	1579	SRO SC 13/2/5
19	1584	SRO SC 13/2/7
20	1585	SRO SC 5/4/76
21	1589	SRO SC 13/2/8
22	1591	SRO SC 14/2/7
23	1592	SRO SC 5/17/4
24	1593	SRO SC 10/1/2
25	1593	PRO E 179/174/415
26	1593	SRO SC 14/2/13 <sup>a</sup>
27	1595	SRO SC 10/1/6
28	1598	PRO E 179/174/432
29	1599	PRO E 179/174/446
30	1601	SRO SC 5/17/12
31	1602	SRO SC 2/1/6
32	1603	SRO SC 5/17/14
33	1604	SRO SC 5/17/15
34	1605	SRO SC 5/17/16
35	1610	SRO SC 5/17/19

Abbreviations: SRO = Southampton Civic Record Office, Civic Centre, Southampton.  
PRO = Public Record Office, Chancery Lane, London.

TABLE 2

## RESULTS OF BASIC COUNTING PROCESSES WITHIN COMPUTER PROGRAM

List	$n_i$	$r_i$	$m_i$	$z_i$
1	41	0	0	0
2	39	2	0	0
3	32	9	2	0
4	30	24	9	0
5	41	24	23	1
6	37	29	22	3
7	16	15	11	21
8	33	24	30	6
9	39	17	28	2
10	28	21	15	4
11	43	33	22	3
12	38	35	33	3
13	31	28	26	12
14	52	39	37	3
15	41	33	34	8
16	31	26	30	11
17	34	33	28	9
18	44	24	41	1
19	48	42	21	4
20	48	31	40	6
21	52	36	32	5
22	55	35	36	5
23	37	34	33	7
24	34	32	28	13
25	34	30	34	11
26	22	17	22	19
27	34	23	29	7
28	21	16	17	13
29	18	15	16	13
30	41	34	25	3
31	40	33	30	7
32	40	34	39	1
33	35	34	34	1
34	36	23	34	1
35	36	0	24	0

Measurement of the Change in Populations Through Time

TABLE 3

CALCULATED ESTIMATES

List number	$\hat{M}_i$	$\hat{P}_i$	$\hat{N}_i$	$\hat{\phi}_i$	$\hat{B}_i$
1				0	
2	0	0	39.00	0.0513	30.00
3	2.000	1.000	32.00	0.2813	21.00
4	9.000	1.000	30.00	0.8236	19.34
5	24.71	0.9309	44.05	0.6047	16.80
6	25.83	0.8513	43.44	0.8181	13.05
7	33.40	0.3293	48.58	0.9961	— 6.32
8	38.25	0.7843	42.07	0.7900	12.15
9	32.59	0.8592	45.39	0.4665	16.78
10	20.33	0.7377	37.96	0.7773	21.14
11	25.91	0.8491	50.64	0.7729	2.61
12	36.26	0.9102	41.75	0.9522	7.09
13	39.29	0.6618	46.84	0.9258	14.26
14	41.00	0.9024	57.62	0.7846	7.77
15	43.94	0.7738	52.99	0.8464	— 0.30
16	43.12	0.6958	44.55	0.8449	7.62
17	37.27	0.7512	45.26	0.9898	1.18
18	42.83	0.9572	45.97	0.5579	32.80
19	25.57	0.8212	58.45	0.9376	4.35
20	49.29	0.8115	59.15	0.6846	23.24
21	39.22	0.8159	63.74	0.7406	19.80
22	43.86	0.8208	67.00	0.6462	2.24
23	40.62	0.8125	45.54	0.9371	8.10
24	41.81	0.6697	50.77	0.9719	— 2.88
25	46.47	0.7317	46.47	1.003	0.00
26	46.59	0.4722	46.59	0.8446	6.78
27	39.35	0.7370	46.13	0.7681	6.64
28	34.06	0.4991	42.08	0.8302	0.62
29	31.60	0.5063	35.55	0.8517	16.65
30	28.62	0.8736	46.93	0.8625	10.83
31	38.48	0.7795	51.31	0.8286	— 1.31
32	40.18	0.9707	41.21	0.8507	1.01
33	35.03	0.9706	36.06	0.9871	2.06
34	35.57	0.9560	37.66		
35					

Where  $0 \leq \hat{\phi}_i \leq 1$  and  $\hat{B}_i \geq 0$  are contradicted, it should be noted that, were the relevant confidence intervals calculated, they would contain "possible" values, within which the true values would lie, subject to the given degrees of confidence.

TABLE 4

ESTIMATES FOR  $\phi_1$  AND  $B_1$  ON AN ANNUAL BASIS

Date	Survival rate	Number of "births"
1501-23	0.883	1.26
1524	0.922	1.14
1525-42	0.936	1.11
1543	0.929	5.90
1544	0.925	7.48
1545	0.902	7.53
1546-47	0.793	7.75
1548-51	0.952	2.15
1552	0.961	— 4.43
1553-55	0.939	3.24
1556	0.907	2.59
1557-62	0.898	2.37
1563	0.928	4.11
1564-66	0.944	4.97
1567	0.928	3.31
1568-69	0.908	0.98
1570	0.948	3.99
1571	0.972	5.72
1572	0.962	7.12
1573	0.938	6.00
1574	0.890	3.73
1575	0.776	2.50
1576-77	0.936	3.05
1578	0.961	2.10
1579	0.931	4.22
1580-83	0.878	7.29
1584	0.957	2.90
1585	0.932	4.42
1586-88	0.908	5.94
1589	0.884	8.01
1590	0.866	9.51
1591	0.782	5.65
1592	0.752	4.39
1593	0.914	3.19
1594	0.903	4.07
1595	0.913	2.89
1596-97	0.923	2.04
1598	0.878	1.61
1599	0.861	4.21
1600	0.928	7.68
1601	0.898	9.45
1602	0.845	5.94
1603	0.851	— 0.41
1604	0.893	1.36