Working Paper 112

POPULATION ANALYSIS OF THE CITY REGION

P.H. Rees

Paper prepared for a British Council Course on Growth and Change in the Future City Region, 20th April to 2nd May, 1975.

Department of Geography
University of Leeds
LS2 9JT.

April 1975

CONTENTS

List of Figures

List of Tables

ABSTRACT

A proper analysis of the population system of a city region requires that all processes of change - mortality, fertility and migration - to be taken into account in the forecasting model adopted. The paper outlines one such forecasting model - a variant of the multi-regional cohort survival model and shows how the three processes of change are represented in the model Difficulties are revealed, however, in matching available demographic statistics to the requirements of model rates. In order to clarify the problems involved and to make best use of available data, demographic accounts are described for a city region. From these accounts the forecasting model rates may be derived or an accounts based model may be used as an alternative. The problems of rate forecasting and of converting a population forecast into a household project are then briefly touched upon at the end of the paper.

1. A systems analysis of population

People interested in the city region and in its planning have always been concerned with the number of people who live in it, how that number has changed over time and how it is likely to change in the future. However, simple description of past trends has prevailed and little analytical insight has been brought to bear. It is the purpose of this paper to show how it is possible to go beyond mere description, and to begin to develop a systems analysis of population. The products of this"systems analysis" are conceptual and mathematical tools with which population change can be monitored and forecast.

This first section of the paper outlines the problem of population forecasting and contrasts the approaches of the demographer and the planner. The second section describes the type of data available on population, the sources of this data and the scale at which it is available. Then in section three this information is organized in a forecasting model - a multi-regional version of the cohort survival model. The next section , the fourth, discusses how the demographic rates and flows needed in the model are generated from existing data. The difficulties and confusions of the process lead on to an outline of demographic accounts where demographic statistics are related to the demographic transitions involved in modelling the population. Section six looks at how rates in either cohort survival or accounts based models may be forecast, and section seven poses the problem of how population data may be converted into household data. Section eight restates the case for population analysis in summing up of the paper.

The focus of interest will be demographic. Demographic models alone may not be sufficient for the accurate forecasting of population trends in the city region. Knowledge of the state of the economy, level of employment, state of the housing market and level of house building are also essential. But without demographic analysis very little can be said about the age/sex breakdown of the population or its potential for change.

In the main attention will be concentrated on method rather than result.

Traditional forms of demographic analysis are somewhat inadequate when applied to the city-region. For example, Keyfitz and Flieger's Population: Facts and Methods of Demography contains some valuable descriptions of demographi methods and associated computer programmes. In Chapter 17 on Cities the methods are applied to cities and a population projection using fixed age-specific birth and death rates is produced for a number of cities, among them Greater London. The results of this projection are summarized in Table 1.

Whereas the population of Greater London fell by 613,000 in the 1961-71 decade and is estimated to fall further since, Keyfitz and Flieger projected an increase. The error resulted from two features of the model. Firstly, birth rates (by age of mother) were assumed to continue at their 1967 levels. They have subsequently fallen considerably and are continuing to fall, a trend which had already established itself in 1965-67. Secondly, the cohort survival model used neglects migration. People are not allowed to change spatial location.

Population projections carried out by Thompson (1971) and his fellow Greater London Council workers take into account these two sources of error. The first projection of Thompson's listed in Table 1 assumes that the fertility rates characterizing the Greater London population remain constant at their 1968 levels and that migration flows influence the projection. In-migrant numbers are assumed to fall slightly, and out-migrant numbers also fall as the population of Greater London itself falls. Net out flows of between 80 and 90 persons per thousand per annum are assumed. In the second projection, fertility

Table 1. <u>Some population estimates and forecasts for Greater London ('000s)</u>

Date	Census and R.G.'s Annual Estimate	Keyfitz and Flieger	Thompson 1: Constant fertility, varying migration	Thompson 2: Fertility decline to 1975, varying migration
(1)	(2)	(3)	(4)	(5)
1961 (c.d.)	7,992			
1966 (m.y.)			7,836	7,836
1967 (m.y.)		7,805	Projected	Projected
1971 (c.d.)	7,379	Projected	figures ↓	figures ↓
1971 (m.y.)	7,441	figures ↓	7,607	7,567
1972 (m.y.)	7,345	8,064		
1973 (m.y.)	7,281			
1976 (m.y.)			7,417	7.234
1977 (m.y.)		8,275		
1981 (m.y.)			7,162	6,841
1982 (m.y.)		8,422		
1986 (m.y.)			6,941	6,518

Notes on each column in the table:

(1) Date: c.d. refers to the date of the census in late April.
m.y. refers to mid-year (June 30/July 1).

(2) Sources: Office of Population Censuses and Surveys (1971) <u>Census 1971. England and Wales Preliminary Report</u>. HMSO, London, Table 4. Office of Population Censuses and Surveys (1973). <u>The Registrar General's Revised Estimates of the Population of England and Wales, Regions and Local Authority Areas 1971 and 1972</u>. H.M.S.O., London, Table 7.

(3) Source: Keyfitz and Flieger (1971), pp. 518-519

(4) Source: Thompson (1971), Table 14.

(5) Source: Thompson (1971), Table 15.

rates are assumed to fall to 1975 to a lower limit equivalent to a birth rate of 12 per thousand. Thereafter fertility rates are assumed constant at this lower limit.

Events since the 1966/1967 base dates upon which the Keyfitz and Flieger and the two Thompson projections are based have supported the second Thompson projection most strongly. The population has continued to decline, largely as a result of migration losses although natural decrease may well play a part in the future. In fact, the Thompson 2 projection may even have underestimated the fall in population involved if recent trends continue.

A proper model for forecasting the population of a city region must concern itself with time (changes in model rates in the near past and future) and with space (migration into and out of the city region). We will show how time and space can be incorporated in population analysis a little later in the paper.

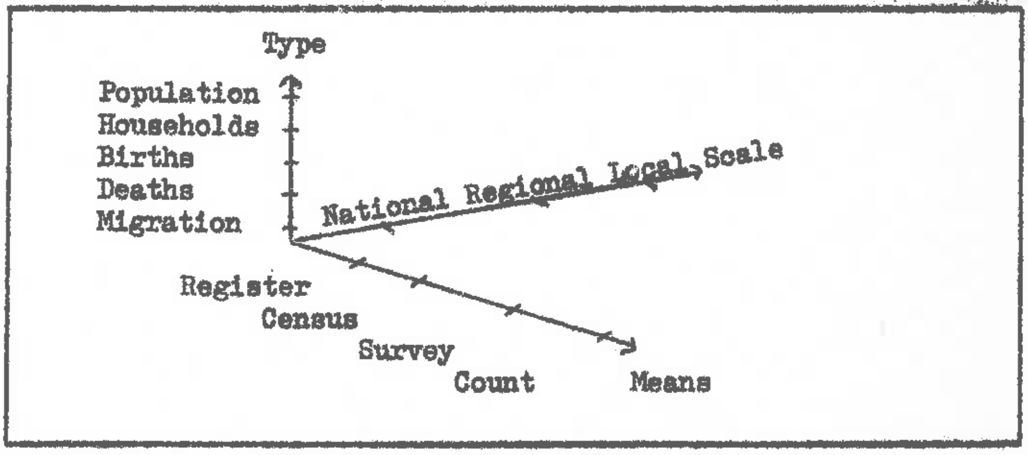
The consequences of incorrect population forecasts can be quite substantial. The investment planned, for example, by the Post Office in extending the telephone service in Britain runs into £100's millions in just the next half decade. The pace of communication network extension depends on detailed forecasts of population for small areas within the city region or rather on forecasts of the number of households and the proportion of households that will have telephones. One might note that a decline in population is not necessarily reflected in a fall in the number of households. In the Greater London case, for example, the number of households fell by only 6,000 despite a population decrease of 613,000 between 1961 and 1971 census dates. The population structure had shifted, new households had formed and the ratio of households to population had increased. In Britain Donnison (1967) has estimated that we have still a substantial potential for new household formation before we reach current Scandinavian or North American headship rates.

Having said something about the focus of the paper, the character of successful population analysis and its importance let us turn to the subject of information about population and examine what kind of data is normally available.

2. Information on population

We must look at the availability of information on population in terms of three dimensions: those of type of data, means of data collection and scale of publication. These three dimensions are pictured in Figure 1.

Figure 1. The dimensions of the demographic information system



To construct a spatial demographic model for a city region one needs the following kinds of information:

(1) data on population stocks or counts of people at a point in time, broken down by age and sex;

(2) data on births or counts of new additions to the population over a period of time, broken down by sex of child and age of mother;

(3) data on deaths or counts of subtractions from the population over a period of time, broken down by sex and age of the deceased;

(4) data on migrants into and out of the regions of interest (and sometimes between them), broken down by age and sex;

(5) data on headship rates or the number of heads of household in a particular sex/age group.

In addition, models with more sophisticated fertility components may require a marital status disaggregation of the female population and of live births and information on marriages and abortions.

A continuous, comprehensive and compulsory population register will supply all the necessary information, and a few countries are fortunate in having such means of data collection (Netherlands, Sweden, Norway, Hungary and others). Other countries must make do with registers of births,marriages and deaths and with periodic censuses (decennial and quinquennial) for population stock and retrospective migration information. Counts of migrations into and out of a country are also usually made. Severe problems occur because data for different components of a demographic model come from different sources. In particular, the points in time at which censuses are taken often fail to coincide with the beginning or end of periods over which births or deaths are recorded.

However, the most severe problems occur with respect to the third dimension of Figure 1 - the scale of publication. Theoretically, the production of demographic statistics for any spatial unit at any spatial scale should be a relative simple problem since the population contained within that unit is merely an aggregation of individual records, the form in which the data was originally produced. However, in practice the most detailed statistics published at the national or country level, less detail is available at regional or local level. Further detailed statistics can be specially ordered for local areas at a fairly high cost in money and time. In some instances the most detailed statistics for local areas may not have been properly preserved; there is no substitute for formal publication at all levels.

These difficulties may ease with the availability of grid referenced records which the user can aggregate to the desired spatial units but procedures for doing this easily and cheaply are still in the development stage.

There is another kind of difficulty putting together information on population, that of matching the available data to the demographic model being used, apart from ensuring that the age group breakdowns and regional definitions from the various sources are consistent. In many countries death statistics are only classified by age group of the deceased at time of death. Most models demand that deaths also be classified by year of birth so that they can be expressed in terms of age group at the beginning of the historical or projection period. Similarly, deaths are classified by region of usual residence at time of death and not by region of usual residence at the beginning of the period, which is what most models demand. Similar considerations apply with even more force to births classified by age group and region of mother. These are usually given as age group and region of usual residence at time of birth. They are needed for the beginning of the period in most models. If a country's official demographers are aware of the demands of the modellers, the age group difficulty is readily overcome. The regional difficulty demands careful analysis of a population register in order to yield the correct information.

3. Organization of demographic information using a multi-regional cohort survival model.

Best use of available population statistics can be made through use of demographic modelling and accounting principles developed in the recent past by Rogers (1967, 1968, 1971, 1973), by Stone (1966, 1971a, 1971b, 1972, 1973) and by Rees and Wilson (1973, 1975a, 1975b) and Wilson and Rees (1974a, 1974b). But before we outline these principles and show how demographic accounts can be constructed for a city-region, we will outline a typical demographic forecasting model of the cohort survival type and show how is has an implied accounting basis. We will then connect this accounting basis with proper demographic accounts.

Let us assume we are interested in one city region in particular, and other regions (city or rural) only in so far as they interact with the region of interest. The population of the region is broken down by five year age groups and by sex. We adopt the following notation. From these basic definitions the reader should be able to ascertain the meaning of composite variables.

i,j	region labels; region i is the region of interest.
X	the sex label that takes on two values X = M,Male and X = F, Female.
r	age group label; r is a label that runs from 1 to R where the Rth age group is the last. Age group 1 refers to people aged 0-4; age group 2 refers to people aged 5-9, and so on to age group R which is open ended and refers to people aged 75 and over. Age group r = 0 refers to persons born in a period.
w	refers to population.
t	a point in time at the start of a period
t+5	a point five years later at the end of a period
t,t+5	the two points in time that bracket and therefore define a period
s	a survival rate
m	a migration rate
f	a fertility rate
*	refers to summation of the subscript or superscript replaced
σ	a sex proportion (e.g. $\sigma^M = 0.51$, $\sigma^F = 0.49$ which are the proportions of live births that are male and female respectively).
d	a death rate
b	a birth and survival rate
bm	a birth and migration rate
S	survivors
M	Migrants
BM	migrants born in the period

D	deaths
B	births

The pieces of the jigsaw puzzle listed above can be put together in a demographic model as follows

$$w_r^{iX}(t+5) = s_{r-1\,r}^{i*X}(t,t+5)\, w_{r-1}^{iX}(t) - \sum_{j\neq i} m_{r-1\,r}^{ijX}(t,t+5)\, w_{r-1}^{iX}(t) + \sum_{j\neq i} M_{r-1\,r}^{jiX}(t,t+5) \qquad \text{for } 1 < r < R \qquad (1)$$

$$w_1^{iX}(t+5) = \sum_{r=\alpha}^{\beta}\left(s_{01}^{i*X}(t,t+5)\,\sigma^X f_r^i\, w_r^{iF}(t)\right) - \sum_{j\neq i} m_{01}^{ijX}(t,t+5) \sum_{r=\alpha}^{\beta} \sigma^X f_r^i\, w_r^{iF}(t) + \sum_{j\neq i} M_{01}^{jiX}(t,t+5) \qquad \text{for age group } r=1 \qquad (2)$$

$$w_R^{iX}(t+5) = s_{R-1\,R}^{i*X}(t,t+5)\, w_{R-1}^{iX}(t) - \sum_{j\neq i} m_{R-1\,R}^{ijX}(t,t+5)\, w_{R-1}^{iX}(t) + \sum_{j\neq i} M_{R-1\,R}^{jiX} + s_{RR}^{i*X}(t,t+5)\, w_R^{iX}(t) - \sum_{j\neq i} m_{RR}^{ijX}(t,t+5)\, w_{R-1}^{iX}(t) + \sum_{j\neq i} M_{RR}^{jiX} \qquad (3)$$

In verbal terms we can express equation (1) (with some considerable simplification) as

$$\begin{array}{lclcl} \text{population in age group } r & = & \text{survival rate age group } r\text{-1 to } r & \times & \text{population in age group } r\text{-1} \\ & - & \text{out migration rate age group } r\text{-1 to } r & \times & \text{population in age group } r\text{-1} \\ & + & \text{in-migrants age group } r\text{-1 to } r & & \end{array} \qquad (4)$$

Equation (3) has a similar structure except that there are two sets of such terms on the right hand side. Equation (2) has the following simplified meaning:

$$\text{population in age group 1} = \text{sum over the fertile age group} \left[\text{survival rate for infants born in the period} \times \text{sex proportion} \times \text{fertility rate for age group } r \times \text{number of women in age group } r \right] \qquad (5)$$

Leslie (1945, 1948) and Rogers (1967, 1968, 1971) have shown how it is possible to arrange this kind of demographic model in matrix form very conveniently. Here we adopt the slightly different version outlined by Matras (1973):

$$\underline{w}'(t) \quad \underline{H}(t, t+5) = \underline{w}'(t+5) \qquad (6)^{*}$$

---

* This is a transposition of the more usual Rogers version $\underline{w}(t+T) = \underline{G}\ \underline{w}(t)$ where $\underline{w}(t+T)$ and $\underline{w}(t)$ are column vectors of population broken down by age group and $\underline{G}$ is a matrix of "growth" rates with the survival rates in the first diagonal below the principal diagonal. The matrices $\underline{G}$ and $\underline{H}$ are related thus: $\underline{H} = \underline{G}'$.

where $\underline{w}'(t)$ is a row vector of populations broken down by age group at time t ($\underline{w}(t)$ is a column vector which we have transposed), $\underline{w}'(t+5)$ is a row vector of populations at time t+5 broken by age group, and $\underline{H}(t,t+5)$ is a matrix of survival and birth and survival rates for the period t, t+5. Equation (6) is a general population growth equation for a period. There are many ways in which the $\underline{w}$ vectors and $\underline{H}$ matrix can be specified. To express the demographic model outlined in equations (1) to (5) in matrix form we need to modify equation (6) and then define carefully the contents of the matrices and vectors:

$$\underline{w}'(t)\,\underline{S}(t,t+T) - \underline{w}(t)\underline{M}^{out}(t,t+T) + \underline{m}^{in'}(t,t+T) = \underline{w}'(t+5) \quad (7)$$

where $\underline{S}$ is a matrix of survival rates and birth and survival rates, $\underline{M}^{out}$ is a matrix of out-migration rates and birth and out-migration rates, and $\underline{m}^{in'}$ is a row vector of in-migrants. All these are specific to the period t to t+5.

Let us illustrate how this model works by constructing an example for a city-region, West Yorkshire.*

---

*The exact definition of this study area is given in Illingworth, Smith and Rees (1974). It is slightly larger than the current West Yorkshire Metropolitan County. It combines the SMLA s (Standard Metropolitan Labour Areas) of Leeds, Halifax, Huddersfield, Dewsbury and Wakefield defined in Hall *et al* (1973).

We distinguish two other regions, the rest of England and Wales and the rest of the world. In doing so we are introducing the principle of spatial comprehensiveness: that is, the regional system defined in a demographic model or set of accounts must include all the regions with which those of interest can interact. In practice this means including all the world. In conceptual terms with the relevant mathematical terms inserted, equation (7) applied to this three region system looks like this for females in region 1:

$$\left[\, w_1^{1F}(t)\; w_2^{1F}(t)\; w_3^{1F}(t)\; \ldots\; w_{19}^{1F}(t)\; w_{20}^{1F}(t) \,\right] \qquad \underset{1\times 20}{\underline{w}^{1F}(1961)}$$

$$\times$$

$$\begin{bmatrix}
0 & s_{12}^{1*F} & 0 & 0 & 0 & & \ldots & 0 \\
0 & 0 & s_{23}^{1*F} & 0 & 0 & & \ldots & 0 \\
b_{31}^{1*F} & 0 & 0 & s_{34}^{1*F} & 0 & & \ldots & 0 \\
b_{41}^{1*F} & 0 & 0 & 0 & s_{45}^{1*F} & & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\
b_{101}^{1*F} & 0 & 0 & 0 & 0 \ldots & s_{1011}^{1*F} & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & & 0 & s_{1920}^{1*F} \\
0 & 0 & 0 & 0 & 0 & & 0 & s_{2020}^{1*F}
\end{bmatrix} \qquad \underset{20\times 20}{\underline{S}^{F}(1961\text{-}66)}$$

$$=$$

$$\left[\, w_1^{1F}(t)\; w_2^{1F}(t)\; w_3^{1F}(t)\; \ldots\; w_{19}^{1F}(t)\; w_{20}^{1F}(t) \,\right] \qquad \underset{1\times 20}{\underline{w}^{1F}(1961)}$$

$$\times$$

$$\begin{bmatrix} 0 & (m_{12}^{12F}+m_{12}^{13F}) & 0 & 0 & 0 & \cdots & & 0 \\ 0 & 0 & (m_{23}^{12F}+m_{23}^{13F}) & 0 & 0 & \cdots & & 0 \\ (bm_{31}^{12F}+bm_{31}^{13F}) & 0 & 0 & (m_{34}^{12F}+m_{34}^{13F}) & 0 & \cdots & & 0 \\ (bm_{41}^{12F}+bm_{41}^{13F}) & 0 & 0 & 0 & (m_{45}^{12F}+m_{45}^{13F}) & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ (bm_{101}^{12F}+bm_{101}^{13F}) & 0 & 0 & 0 & 0 & \cdots (m_{1011}^{12F}+m_{1011}^{13F}) & \cdots & 0 \\ \vdots & \vdots & \vdots & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & (m_{1920}^{12F}+m_{1920}^{13F}) \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & (m_{2020}^{12F}+m_{2020}^{13F}) \end{bmatrix} \quad \underset{20\times 20}{\underline{M}\ (1961\text{-}66)}$$

$$+$$

$$\left[(BM_{41}^{21F}+BM_{41}^{31F})\ (M_{12}^{21F}+M_{12}^{31F})\ (M_{23}^{21F}+M_{23}^{31F})\ (M_{34}^{21F}+M_{34}^{31F})\ (M_{45}^{21F}+M_{45}^{21F})\ \cdots\ (M_{1920}^{21F}+M_{1920}^{31F}+M_{2020}^{21F}+M_{2020}^{31F})\right] \quad \underset{1\times 20}{\underline{m}'\ (1961\text{-}66)}$$

$$=$$

$$\left[w_1^{1F}(t+5)\ w_2^{1F}(t+5)\ w_3^{1F}(t+5)\ \cdots\ w_{19}^{1F}(t+5)\ w_{20}^{1F}(t+5)\right] \quad \underset{1\times 20}{\underline{w}'\ (1966)} \qquad (8).$$

On the right hand side of the expanded vectors and matrices we have noted the equation (7) terms to which they correspond and their overall size (dimensions).

For the inter-censal period 1961-66 (April 23/24 to April 24/25) we can fill in equation (8) as follows, showing a selection of the numbers involved:

$$
\begin{bmatrix} 77352 & 69903 & 81200 & 61718 & \ldots & 1324 & 185 \end{bmatrix}
$$

$$
\times
$$

$$
\begin{bmatrix}
0 & .99493 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & .99857 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & .99834 & 0 & \ldots & 0 \\
.03692 & 0 & 0 & 0 & .99769 & \ldots & 0 \\
.30210 & 0 & 0 & 0 & 0 & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
.00035 & 0 & \ldots & & & .97645 & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
0 & 0 & 0 & 0 & 0 & \ldots & .14959 \\
0 & 0 & 0 & 0 & 0 & \ldots & .08270
\end{bmatrix}
$$

$$
=
$$

$$
\begin{bmatrix} 77352 & 69903 & 81200 & \ldots & 1324 & 185 \end{bmatrix}
$$

$$
\times
$$

$$
\begin{bmatrix}
0 & (.04213+.01677) & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & (.04213+.01676) & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & (.03554+.04708) & 0 & \ldots & 0 \\
.00123 & 0 & 0 & 0 & (.08779+.05354) & \ldots & 0 \\
.01005 & 0 & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
.00000 & & \ldots & & & (.02342+.00435) & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
0 & 0 & 0 & 0 & 0 & \ldots & (.01915+.00230) \\
0 & 0 & 0 & 0 & 0 & \ldots & (.01641+.00000)
\end{bmatrix}
$$

$$+ \left[(1985+865)(3618+2019)(3289+1733)(2888+1565)(5715+2286) \ldots (28+2+5+0)\right]$$

$$= \left[88301 \quad 78041 \quad 70708 \quad 78810 \ldots \quad 1872 \quad 217\right] \tag{9}$$

We can follow through the calculations for the fourth age group at the end of the period, for example (the 15-19 year olds in West Yorkshire):

$$\begin{aligned} 78810 &= (81200)(.99834) - (81200)(.03554+.04708) + (2888+1565) \\ &= 81065 - 6709 + 4453 \end{aligned} \tag{10}$$

with a slight rounding error of 1 person (the right hand side comes to 78809).

The model we have outlined here, which traces the change in the city region's population, is a variant of the multi-regional cohort survival model first put forward by Rogers. There are a large number of alternative ways in which such models may be arranged (see Rees and Wilson, 1975c, Chapter 2 and Rees and Wilson, 1975b for a review of this variety) and the one outlined above is only one among many. In fact, it approximates most closely to that used by Thompson (1971) in the projection of the population of Greater London.

The key question we must now answer is where do the rates used in the model come from.* There are

*The population stocks derive from the census, as do the in-migrant numbers.

really two answers. The first describes conventional practice. The second outlines an alternative accounting basis for rate definition. We first discuss conventional practice.

4. Rates for the multi-regional cohort survival model: conventional practice.

Survival rates, the s_{r-1r}^{i*x}'s, are required in the model outlined in section 3. Survival as such is not an event which is directly measured so that survival rates are defined as the complements of death rates

$$s_{r-1r}^{i*x}(t,t+5) = 1 - d_{r-1*}^{i*x}(t,t+5) \tag{11}$$

where $d_{r-1*}^{i*x}(t,t+5)$ is the rate at which persons of sex x in region i at time t in age group r-1 die in the next five years before time t+5. This death rate is defined as

$$d_{r-1*}^{i*x}(t,t+5) = D_{r-1*}^{i*x}(t,t+5) / w_{r-1}^{ix}(t) \tag{12}$$

where $D_{r-1*}^{i*x}(t,t+5)$ are the corresponding deaths and $w_{r-1}^{ix}(t)$ is the initial population of region i in age group r-1.

Unfortunately, D_{r-1*}^{i*x} statistics are not usually available. Death statistics come in the form

D_{*r-1}^{*ix} = the deaths for persons of sex x in region i and age group r-1 <u>at death</u>

There is no information as to location or age group at the beginning of the period. What is sometimes done is to assume that

$$D_{r-1*}^{i*x} = D_{*r-1}^{*ix}(t,t+5) \tag{13}$$

and to calculate death rates and thus survival rates on that basis. This assumption may involve considerable error at the youngest and oldest ages. A better assumption is to assume that

$$D^{i*X}_{r-1\,*}(t,t+5) = D^{*iX}_{r-1\,*}(t,t+5) \qquad (14)$$

and to try and estimate $D^{*i}_{r-1\,*}(t,t+5)$.
This will involve some attempt to break down the deaths in each age group into the two components belonging to separate cohorts:

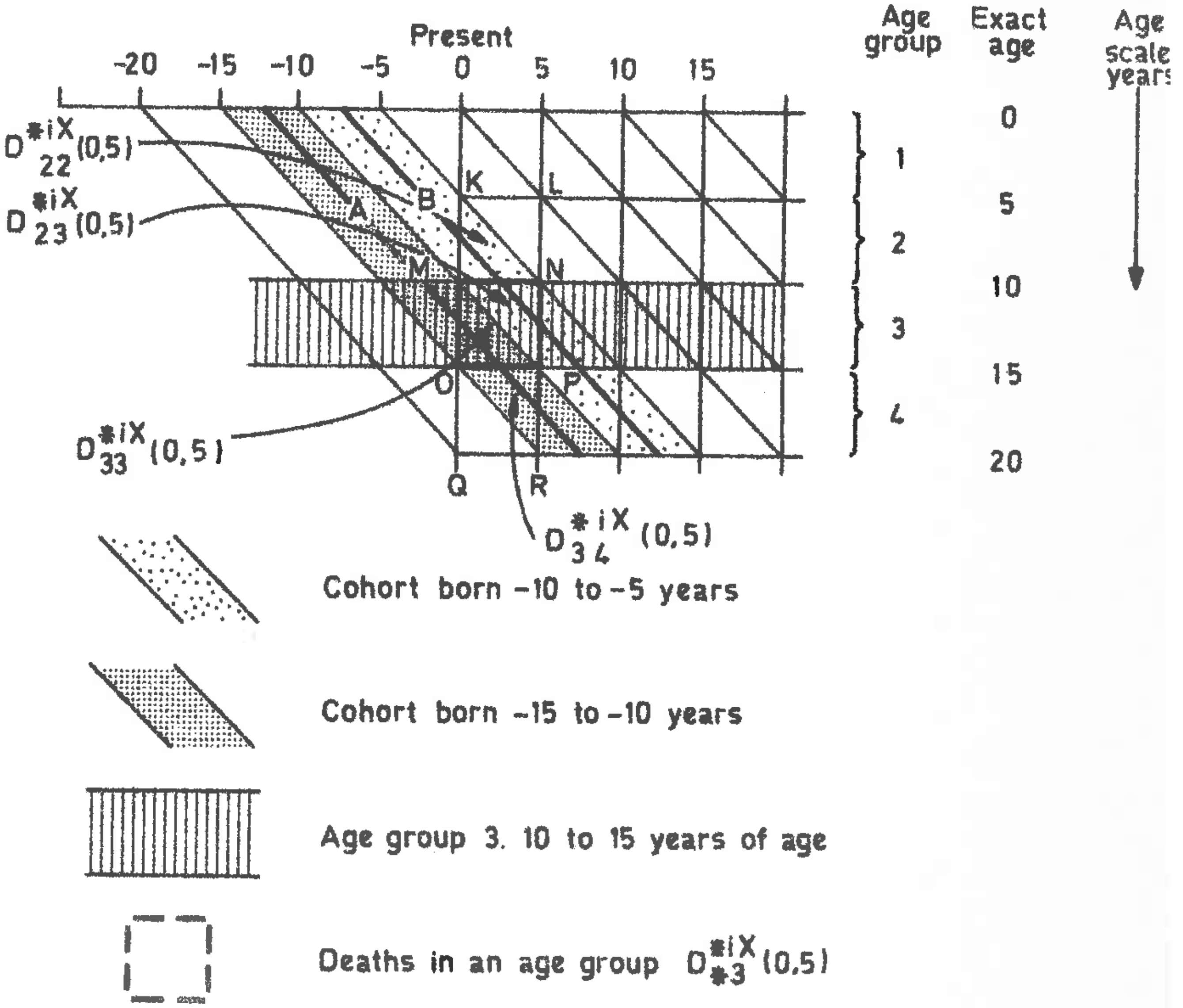
$$D^{*iX}_{*r}(t,t+5) = D^{*iX}_{r-1\,r}(t,t+5) + D^{*iX}_{rr}(t,t+5) \qquad (15)$$

The variable $D^{*iX}_{r-1\,r}$ refers to deaths to persons of sex in age group r-1 at time t who die in age group r in region i in period t to t+5. Given this breakdown the components may be reassembled thus

$$D^{*iX}_{r-1\,*}(t,t+5) = D^{*iX}_{r-1\,r-1}(t,t+5) + D^{*iX}_{r-1\,r}(t,t+5) \qquad (16)$$

to yield all the deaths to persons in age group r-1 at time t that occur in the period t to t+5 in region i.
Figure 2 illustrates the concepts involved.

Figure 2. A Lexis diagram illustrating the way in which deaths are conventionally measured and the way in which they enter a demographic model



This illustration is a Lexis diagram (Lexis, 1875) in which age is plotted on the vertical axis and time on the horizontal. Lifelines are located at 45° in this space and slope down from the left. People born in the same period of time (a cohort) move through the Lexis diagram in 45° corridors. Person A, for example, was born in the period 15 to 10 years before the present, and person B was born in the period 10 to 5 years before the present. Events, such as death, occur on lifelines in this space and can be counted within particular portions of it.

Deaths in an age group are counted in a square: the deaths in age group 3 are picked out on the diagram in □ MNPO. They are the mathematical variable $D_{*3}^{*ix}(0,5)$. This square is made of triangles △MNP and △MPO in which deaths $D_{23}^{*ix}(0,5)$ and $D_{33}^{*ix}(0,5)$ occur. If we are interested in the deaths occurring to persons in age group 3 at time 0 over the subsequent 5 years we need to count the deaths in parallelogram ▱ MPRO made up of the two components $D_{33}^{*ix}(0,5)$ and $D_{34}^{*ix}(0,5)$ in triangles △ MPO and △ OPQ respectively.

This breakdown of deaths by age group at death is simply achieved if deaths are also classified by cohort of birth. Otherwise estimation procedures must be adopted.(see Smith and Rees (1974) for a detailed description of these). Table 2 shows our estimates of deaths, doubly classified by age group at census date 1961 and at death in 1961-66, for West Yorkshire.

Table 2 Deaths in West Yorkshire classified by age group at census date 1961 and at death in 1961-66

Part 1.

		Age group at death										
Cohort (Years of birth)	Age group at c.d 1961	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45- 49	Totals
61-66	birth	1846										1846
56-61	0-4	323	73									396
51-56	5-9		53	48								101
46-51	10-14			53	80							133
41-46	15-19				72	79						151
36-41	20-24					78	104					182
31-36	25-29						106	137				243
26-31	30-34							168	215			383
21-26	35-39								266	396		662
16-21	40-44									438	560	998
11-16	45-49										738	1747
Totals		2167	126	101	154	158	210	305	481	835	1299	

Part 2

Age group at death

Cohort (Years of birth)	Age group at c.d. 1961	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-95	95 and over	Total
11-16	45-49	1009										174
06-11	50-54	1160	1545									270
01-06	55-59		1748	2334								408
96-01	60-64			2698	3578							627
91-96	65-69				3816	4865						868
86-91	70-74					5211	5692					1090
81-86	75-79						5744	5866				1161
76-81	80-84							5456	4296			975
71-76	85-89								2946	1582		4528
66-71	90-94									845	288	113
61-66	95 and over										173	173
Totals		2168	3292	5034	7395	10075	11436	11322	7243	2426	462	66689

Note: Rounding error of ±1 or 2 occurs in the columns. The grand total is a sum of the column totals and is 4 greater than the sum of row totals.

A commonly adopted alternative is to ignore the problem, and to rely instead on survival rates derived from national life tables multiplied by differentials for the city region being studied (as in Thompson (1971), for example). The survival rates that are used are not those used directly in the construction of the life table but the ratio of two "life-years lived" quantities. The mathematics involved in the multi-regional case has been outlined in Rogers(1973), Rogers and Ledent (1974) and Rees and Wilson (1975c, Chapter 16). Again many assumptions are involved and error may creep in. The alternative described in section 5 is to measure death rates and survival rates directly from demographic accounts for the city region concerned.

Migration rates, the m_{r-1r}^{ijx}'s, are normally derived by dividing migrant numbers given in the census migration tables by the origin region's initial population:

$$m_{r-1r}^{ijx}(t,t+5) = M_{r-1r}^{ijx}(t,t+5) / w_{r-1}^{ix}(t) \qquad (17)$$

The migrant data is in the correct form for the model.

Birth and survival rates, the b_{r1}^{i*x}'s, record the rate at which mothers in age group in region i and in age group r gave birth to infants of sex x who survive the period and are alive in age group 1 at time t+5. No statistics of this exact nature exist. Births are recorded by region in which the mother is resident at time of birth. This region may be quite different from that of the mother at the start of the period and of the child at the end of the period.

An approximate method is adopted to estimate these rates:

$$b_{r1}^{i*x} = s_{01}^{i*} \sigma^{x} f_{r}^{i} \qquad (18)$$

or the birth rate is estimated by the survival rate of infants multiplied by the proportion of births that are of sex X multiplied by the fertility rate of women in age group r in region .

For example,

$$b_{41}^{1*F} = s_{01}^{1*F}\,\sigma^{F} f_{4}^{1}$$
$$= (.97957)(.485)(.63589)$$
$$= .30210 \qquad (19)$$

The figure of .30210 was entered earlier in the S matrix in equation (9). This is the product of a 98 per cent survival rate multiplied by a female proportion of 48.5 per cent multiplied by a fertility rate of 635 per 1000, equivalent to an annual rate of 127 per 1000 (all divided by 10^7 to produce a proportion).

Birth and migration rates, the b_{r1}^{ijX}'s , are similarly estimated using the following rate multiplication

$$b_{r1}^{ijX} = m_{01}^{ijX}\,\sigma^{X} f_{r}^{i} \qquad (20)$$

where the component rates on the right hand side are estimation as

$$m_{01}^{ijX} = M_{01}^{ijX} / B^{i*X} \qquad (21)$$

$$\sigma^{X} = B^{i*X} / B^{i**} \qquad (22)$$

$$f_{r}^{i} = B_{r*}^{i*X} / w_{r}^{iF}(t) \qquad (23).$$

There are difficulties in classifying births by age group of mother at time t and by mother's location at time t similar to those already described for death. Use of the correct cohort cum age group figures is even more essential in this context.

In-migrant data, the $M_{r\iota r}^{jix}(t,t+5)$, of the multi-regional cohort survival model, are obtained from census migration tables. Infant in-migrant data, that is information on children born in a region in the period who subsequently migrate to another region before the end of the period, is not usually given, but could easily be produced by the census authorities by the matching of birthplaces and residences on census night of the under-fives.

5. Organization of demographic information into demographic accounts

By now it will have become clear that there are many difficulties in relating available demographic statistics to the requirements of a multi-regional cohort survival model. However, there is a way to ensure that the available data are used most efficiently and consistently: that is, through the construction of demographic accounts. Demographic accounts are devices for specifying all the state-to-state transitions that occur in a period in multi-regional populations. Originally developed on a national basis by Stone (1966, 1971a, 1971b, 1972), demographic accounts have been generalized to a multi-regional basis by Rees and Wilson (1973, 1975a) and by Wilson and Rees (1974a).

Accounts are based on a simple and comprehensive enumeration of the transitions that can occur to people in a period in a comprehensive regional system. Such an enumeration is shown in Figure 3 for a two region system consisting of a region i of interest and a region R signifying the rest of the world. Region i might be a city region like West Yorkshire, for example. People begin the period in one of two possible "life-states" -- existence at the start of the period, or birth at some point in the period. Each of these life-states can occur in one of the two regions giving four origin states for the population. People end the period either alive at the end of the period or deceased sometime before it in the life-states survival or death respectively. Again, survival and death can occur in one of two regions giving four destination states. When we combine the four origin states with the four destination states, and trace the possibilities in a logical tree (Figure 3) we obtain some 16 possible sequences or transitions. An alternative representation of these transitions is in a set of time-space diagrams. Figure 4 shows the life-lines of individuals who experience each of the 16 transitions, following the numbering of Figure 3.

These transitions can be arranged as a matrix in tabular form by using the origin states as rows and the destination states as columns. This is done in Figure 5. This is a demographic accounts matrix for a two region system. In order to explore and exploit

Figure 3. <u>The logical tree of possibilities for demographic transitions in a single period in a comprehensive two region system.</u>

Origin states: Life-state	Origin states: Region	Destination states: Life-state	Destination states: Region	Transition number
Existence	Region i	Survival	Region i	1
			Region R	2
		Death	Region i	3
			Region R	4
	Region R	Survival	Region i	5
			Region R	6
		Death	Region i	7
			Region R	8
Birth	Region i	Survival	Region i	9
			Region R	10
		Death	Region i	11
			Region R	12
	Region R	Survival	Region i	13
			Region R	14
		Death	Region i	15
			Region R	16

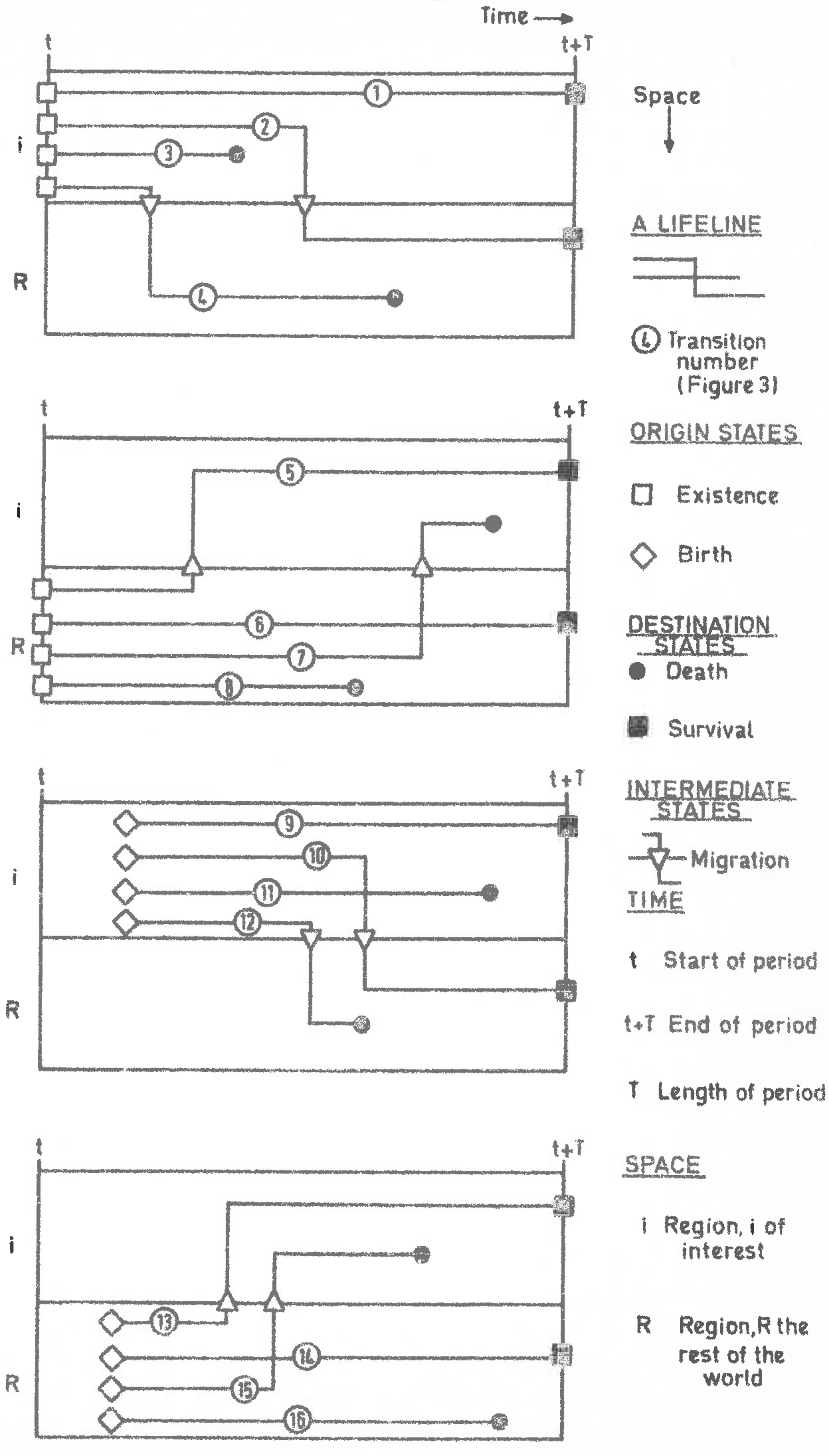


Figure 4. A time-space diagram illustrating the demographic transitions possible in a single period in a comprehensive two region system

Figure 5. A demographic accounts matrix for the 16 population transitions.

Origin state \ Destination state		Survival (at time t+T)		Death (in period t to t+T)	
		Region i	Region R	Region i	Region R
Existence (at time t)	Region i	1	2	3	4
	Region R	5	6	7	8
Birth (in period t to t+T)	Region i	9	10	11	12
	Region R	13	14	15	16

Figure 6. A demographic accounts matrix for the 16 population transitions in K notation

Origin states \ Destination states		Survival (at time t+T) σ		Death (in period t to t+T) δ		Totals
		Region i	Region R	Region i	Region R	
Existence (at time t) ϵ	Region i	$K^{\epsilon(i)\sigma(i)}$	$K^{\epsilon(i)\sigma(R)}$	$K^{\epsilon(i)\delta(i)}$	$K^{\epsilon(i)\delta(R)}$	$K^{\epsilon(i)*(*)}$
	Region R	$K^{\epsilon(R)\sigma(i)}$	$K^{\epsilon(R)\sigma(R)}$	$K^{\epsilon(R)\delta(i)}$	$K^{\epsilon(R)\delta(R)}$	$K^{\epsilon(R)*(*)}$
Birth (in period t to t+T) β	Region i	$K^{\beta(i)\sigma(i)}$	$K^{\beta(i)\sigma(R)}$	$K^{\beta(i)\delta(i)}$	$K^{\beta(i)\delta(R)}$	$K^{\beta(i)*(*)}$
	Region R	$K^{\beta(R)\sigma(i)}$	$K^{\beta(R)\sigma(R)}$	$K^{\beta(R)\delta(i)}$	$K^{\beta(R)\delta(R)}$	$K^{\beta(R)*(*)}$
Totals		$K^{*(*)\sigma(i)}$	$K^{*(*)\sigma(R)}$	$K^{*(*)\delta(i)}$	$K^{*(*)\delta(R)}$	$K^{*(*)*(*)}$

how it is possible to replace the 16 numbers of Figure 5 with 16 variables. Although only two regions are recognised in these accounts, they are quite general in the sense that region i can be replaced by any number of regions of interest, and region R by any number of regions in the rest of the world.

We have included row, column and grand totals in Figure 6 replacing the superscripts which have been summed over by asterisks. The totals $K^{\epsilon(i)*(*)}$ and $K^{\epsilon(R)*(*)}$ refer to the populations of regions i and R at time t, the start of the period. The totals $K^{\beta(i)*(*)}$ and $K^{\beta(R)*(*)}$ are the total births in the period in regions i and R, and $K^{*(*)\delta(i)}$ and $K^{*(*)\delta(R)}$ are the deaths. The $K^{*(*)\sigma(i)}$ and $K^{*(*)\sigma(R)}$ totals are the end of period populations of the regions. Note that this interpretation of the totals, which is very useful for estimating and modelling purposes, is true only if the regions listed constitute a closed system. In practice, they must cover all the world.

To illustrate what accounts for a multi-regional system looks like Figure 7 shows aggregate accounts for the West Yorkshire Metropolitan county* and the rest of England and Wales as regions of interest, and with a rest of the world region.

---

*These are an aggregation of more detailed accounts given in Illingworth (1975) which disaggregate the West Yorkshire Metropolitan County into its five constituent districts (Wakefield, Kirklees, Calderdale, Leeds and Bradford) and the rest of England and Wales into the rest of the West Riding and the rest of England and Wales outside the West Riding.

Figure 7 A demographic accounts table for the aggregate population of the West Yorkshire Metropolitan County, 1961-66

Origin States \ Destination States		Survival at c.d. 1966 WYMC	Survival at c.d. 1966 REW	Survival at c.d. 1966 RTW	Death in i.c.p. 1961-66 WYMC	Death in i.c.p. 1961-66 REW	Death in i.c.p. 1961-66 RTW	Totals
Existence at c.d. 1961	WYMC	1,778,783	97,622	28,418	121,130	2,893	929	2,010,465
	REW	79,006	40,705,357	800,929	2,544	2,482,448	23,801	42,617,977
	RTW	38,189	1,033,059	0	1,267	30,703	0	1,103,218
Birth in i.c.p. 1961-66	WYMC	171,623	4,343	2,103	5,704	65	34	183,872
	REW	4,068	3,897,560	55,471	54	116,703	818	4,074,685
	RTW	1,725	45,808	0	28	676	0	48,237
	Totals	2,054,082	45,783,747	886.920	130,741	2,633,487	25,582	51,514,559

c.d. - census date

i.c.p. - inter-censal period. Slight rounding error occurs in the table.

WYMC - West Yorkshire Metropolitan County

REW - Rest of England and Wales

RTW - Rest of the World

130,741 Items available from British sources

The value of such an accounts table is that it informs us directly about quantities like "the persons who survive the period and stay in West Yorkshire", a population flow which had to be teased out with some difficulty from the version of the multi-regional cohort survival presented earlier. Some 1,778,783 out of an initial population of 2,010, 465 stayed and survived in West Yorkshire.

It also clearly demonstrates the importance of migration in population change in the West Yorkshire city-region. Table 3 shows the accounts of Figure 7 re-arranged so as to reveal the importance of natural increase and net migration as components of change.

Natural increase makes up 122 per cent of population change and net migration -22 per cent of population change. In the 9 years since 1966 the natural increase has declined to near balance increasing the importance of net migration as a component of population change. The total figure of ±2,321,228 at the bottom of the inputs and outputs columns indicates the number of persons "flowing through" West Yorkshire over the five year period. Only 76.6 per cent of these were people who stayed and survived in the city region; some 23.4 per cent were involved in a change of state. A further large proportion of the surviving stayers were, in fact, involved in migration between metropolitan districts, between the old local authorities or within them. Even over the short period of five years a substantial proportion of the population of a city region is involved in life-state and location change.

One practical point to note about Figure 7 is that population transitions with RTW-RTW locational labels are set to zero (effectively ignored, that is). The relevant row totals are totals of in-migrants from the rest of the world; and the column totals are for out migrants to the rest of the world from the system of interest.

Clearly aggregate population accounts are not a great deal of practical use. We must break the population down by age group and by sex. Persons in existence-survival transitions are classified by age group at time t and at time t+T; persons in existence-death transitions are classified by age group at time t and at time of death; persons in birth-survival transitions are classified by age group of mother at time t and by age group of child at time t+T; and persons in birth-death transitions are classified by age group of mother at time t and by age group of child at time of death. A set of age/sex disaggregated accounts for West Yorkshire are displayed in Figure 8.* These show how the female populations in each five year age group of West Yorkshire and of the rest of England and Wales changed over the 1961-66 intercensal period. There are similarities in the structure of the submatrices on the left hand or survival portion of the accounts and the structure of the multi-regional cohort survival model (see equation (8)). Surviving stayers and surviving migrants are located in the diagonal one above the principal diagonal just as were the survival and migration rates of equation (8). The infants born in the period are located in the first column in rows referring to the fertility age groups of females (10-14 to 45-49) as were the birth rates and birth and migration rates in the cohort survival model. The main difference between model and accounts

*These refer to a West Yorkshire Study Area slightly larger than the current Metropolitan County.

is that it is most convenient in the accounts to separate out those born and those existing at time t into separate sets of submatrices.

If each element of the accounts matrix is divided by its appropriate row total, or in the case of births by the total of females in the relevant age group then the rates needed in a multi-regional cohort survival model are generated. Figure 9 shows the transition rates associated with the West Yorkshire accounts, both for persons in existence at time t and for persons born in the period. From this table and other associated tables can be generated the rates needed in almost any version of the multi-regional cohort survival model.

We can now return to the proposition that demographic accounts are a way of using demographic statistics most efficiently and consistently, and that they can be used to supply the correct rates to multi-regional cohort survival models. A careful comparison of the numbers in Figure 8 and 9 and those used earlier should convince the reader of the second part of the proposition.

The survival rates used in the multi-regional cohort survival model can be obtained by summing the transition rates involving existence-survival transitions. If we adopt the letter h to denote transition rate and retain the superscripts of the

K accounts notation and the age subscripts of the model notation; then survival anywhere rates for persons in region i and age group r at time t are

$$h^{\epsilon(i)\sigma(*)X}_{r\ r+1} = \sum_j h^{\epsilon(i)\sigma(j)X}_{r\,r+1}(t,t+5)$$

$$= s^{i*X}_{r*}(t,t+5) \qquad (24).$$

For example, the survival anywhere rate for girls in the 10-14 age group in the West Yorkshire Study Area (WYSA) at census date 1961 is

$$h^{\epsilon(WYSA)\sigma(*)F}_{3\ 4}(1961,66) = h^{\epsilon(WYSA)\sigma(WYSA)F}_{34}(1961,66)$$
$$+ h^{\epsilon(WYSA)\sigma(REW)F}_{3\,4}(1961,66)$$
$$+ h^{\epsilon(WYSA)\sigma(RTW)F}_{34}(1961,66)$$

$$= .91572$$
$$+ .03554$$
$$+ .04708$$
$$= .99834 \qquad (25).$$

Which is the s^{WYSA*F}_{34} figure included in equation (9).

The outmigration rates of the transition rates table and the earlier model are directly equivalent

$$h^{\epsilon(i)\sigma(j)X}_{r\,r+1}(t,t+5) = m^{ij\,X}_{r\,r+1}(t,t+5) \qquad (26)$$

and the total outmigration rate is

$$\sum_{j\neq i} h^{\epsilon(i)\sigma(j)X}_{r\,r+1}(t,t+5) \;,$$

For the 10-14 year olds in West Yorkshire the rates of outmigration are

$$h_{34}^{\epsilon(WYSA)\sigma(REW)F}(1961,66) = .03554 \qquad (27)$$

and

$$h_{34}^{\epsilon(WYSA)\sigma(RTW)F}(1961,66) = .04708 \qquad (28)$$

which have been included as a (.03554 + .04708) term in equation (9). Similar connections can be established for the birth associated rates and the in-migrant numbers.

There is not space here to explain how demographic accounts are generated from available statistics except in most general terms. Details are given in Smith and Rees (1974), Rees and Wilson (1975c), and Rees, Smith and King (1975). There an accounts based model is defined that uses the available population, migration, births and deaths data to fill in the accounts matrix using an iterative procedure. In Figure 7, the items of data normally available from British sources have been distinguished by a box symbol. The rest are then estimated in the accounts based model. In other countries or in other situations more or less data may be available.

One of the features of the accounts based model may be particularly useful in a city-region context, especially in the modelling the population of its constituent zones . It is possible to trace in the model the numbers of children born to mothers who migrate both before and after migration in a period. This makes possible the forecasting of rapid buildups of children in newly growing suburbs which might be missed in more conventional analyses.

6. Forecasting of rates and exogeneous variables.

At the beginning of the paper we said that a proper population forecasting model had to concern itself with time and space. We have shown how space can be incorporated in various ways in the forecasting model. We now consider changes in demographic rates over time and how rates may be forecast.

In national population forecasting attention can be devoted to the study of time series of mortality, fertility and migration rates for the large population of just one region. When city regions are being studied there are many rate series to be analyzed. The populations on which they are based are much smaller, and this may make the rate series for individual regions less reliable. A hierarchical solution is often adopted for mortality rates and fertility rates. National trends are adopted but adjusted to allow for the difference between national andregional rates. Attention can then be concentrated on the forecasting of migration flows and rates which must be based on local analysis.

We can illustrate this technique for death rates which may be used in defining survival rates in the multi-regional-cohort survival model (equations (11) and (12)) or directly in a forecasting version of the accounts based model (Rees and Wilson, 1975c) briefly described at the end of the last section.

If one plots death rates in England and Wales for age groups against time they follow straight line plots on semi-logarithmic paper and can be described by a negative exponential function of the following form (Rees and Wilson, 1975c)

$$d_r^{EW}(\theta) = a_r^{EW} \; e^{-b_r^{EW} t(\theta)} \qquad (29)$$

where $d_r^{EW}(\theta)$ is the death rate in England and Wales in age group r (either at the beginning of a year or at time of death) for period θ , $t(\theta)$ is the mid-point in time of period , e is the base of natural logarithms and a_r^{EW} and b_r^{EW} are equation parameters particular to England and Wales and age group r. The first, a_r^{EW} , gives the intercept or starting point of the death rate and the second, b_r^{EW} , measures how fast it falls. The fact that the model is a negative exponential one implies that the absolute improvement in death rates each year gets smaller and smaller as time progresses. Figure 10 shows the improvement in female death probabilities

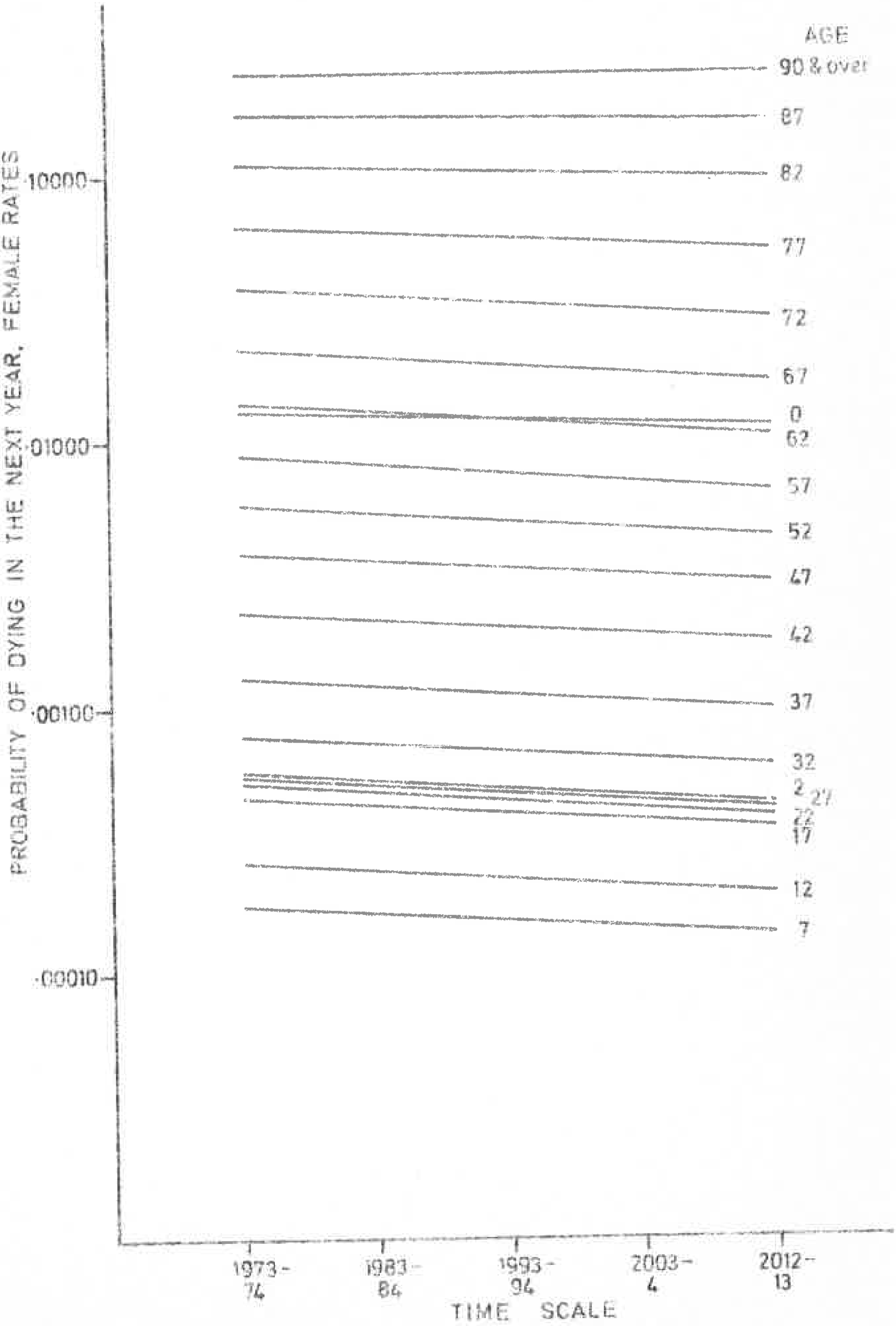


Figure 10 Forecast death probabilities in the next year for females 1973-4 to 2012-13 in England and Wales

expected by the Government Actuary over the next 40 years (O.P.C.S., 1974, Appendix Table V, p.89). The rates at which death rates are projected to fall have been revised downwards in the light of recent trends. There is not a great deal of difference in the slopes of the different age functions. There is, however, a tendency for the slope to diminish with age.

The equation for age 2 is

$$d_2^{EWF}(\theta) = .00730 \; e^{-(.00708)t(\theta)} \qquad (30)$$

where time is measured with 1973-74 as the first year and the $t(\theta)$ zero point at midnight on December 31st 1973/January 1st 1974. The equation for age 32 is

$$d_{32}^{EWF}(\theta) = .00130 \; e^{-(.00706)t(\theta)} \qquad (31)$$

and for age 90 and over it is

$$d_{90+}^{EWF}(\theta) = .27600 \; e^{-(0)\,t(\theta)} \qquad (32)$$

which means that a constant death rate is assumed.

A forecast of the corresponding regional rates can be accomplished by replacing a_r^{EW} in equation (29) by a_r^{1} or the starting level of death rates in the region so that the forecasting equation is

$$d_r^{1}(\theta) = a_r^{1}(\theta) e^{-b_r^{EW}\; t(\theta)} \qquad (33).$$

This is equivalent to shifting the forecasting function up or down depending on the relationship between regional and national rates. This is illustrated in Figure 11 for age 4, interpolated from the series plotted in Figure 10. This is the rate that corresponds to the annual equivalent of the death rate of the 0-4 age group over five years. The assumption is made that the regional differential remains constant over the forecasting period. We would use the death rate forecast to generate forecasts of "survival anywhere rates" using equation (11).

Similar techniques could be applied to fertility rate forecasting for the city region, though here the Government Actuary's assumptions are more controversial. The key to making good forecasts in this area is guessing the likely completed family size of the cohorts of women looking at past behaviour, family intentions and at likely influences. The argument is put forward that some of the recent fall in births represents postponement by career oriented families. The births will occur later. On this basis, the Government Actuary (O.P.C.S., 1974, p21) argues that current low fertility levels will recover in the late 1970's and 1980's to give a total period fertility rate of 2.2 by 1984 compared with a 1973 figure of 2.015 and a 1974 estimate of 1.90. Other demographers have challenged this argument saying instead that the population has shifted to lower family size goals. To what extent regional populations or those of city regions follow national trends in fertility is not entirely clear. Northern Ireland, for example, has maintained a

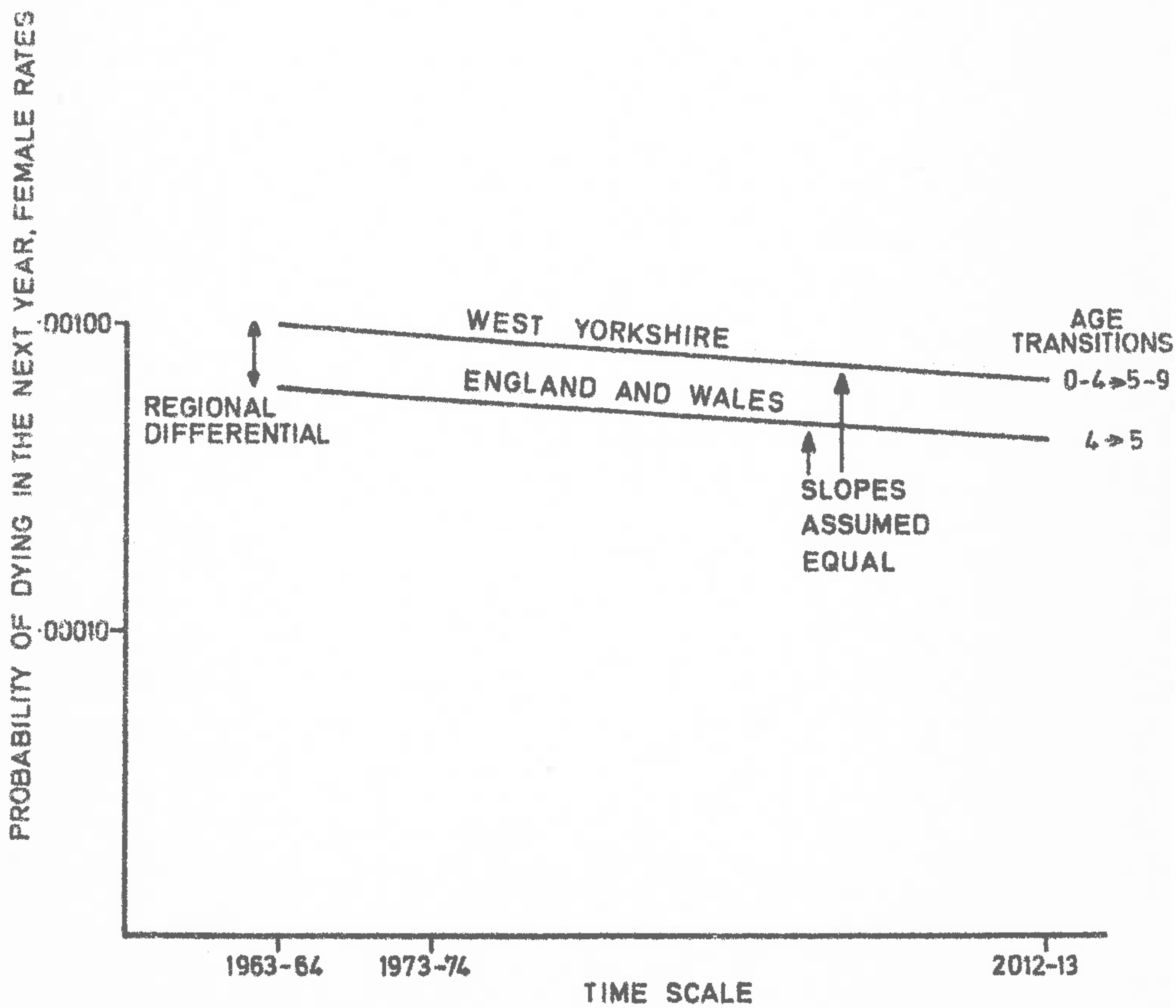


Figure 11. Regional forcasts and national forecasts of deathrates

consistently higher fertility rate than the rest of the U.K. and the variations in this have been less marked (O.P.C.S , 1974, p23). It may, in fact, be unwise in this instance to adopt the kind of model represented by equation (33), and it may be wise to conduct a separate analysis of fertility trends in the city region.

The most difficult rates to forecast are those for migration, primarily because of a lack of good time series, and because migration is heavily influenced by exogenous factors. We now have in Britain data on migration in three one year time periods (1960-61, 1965-66, 1970-71) and in two five year periods (1961-66, 1966-71)*, but little analysis of the trends involved. The Registrar General produces what is, in effect, a net migration series when he makes his annual estimates, and there appear to be clear links (Cordey Hayes and Gleave, 1974) between net migration rates and in- and out-migration rates which may make feasible use of a net migration series as a leading indicator.

We know quite a lot about how migration is distributed in space given knowledge about the distribution of opportunities (Willis, 1974). Thus,

---

*The Census 1971 migration data are not yet available in published form although prepublication figures have been supplied at special request.

if the planner knows the short term plans of his housing department and has been given some idea of the locations of new employment in his city region, and he can link the houses built and jobs created to numbers of in-migrants broken by households, age and sex, he can work out the likely origin of those migrants using a spatial interaction model of migration. These forecasts can be fed into the population forecasting model.

7. Putting population into households.

The population forecast yields numbers of people classified by age and sex, and perhaps by marital status. It is equally useful and in many cases essential to know what these numbers mean in terms of households. Households are estimated by multiplying the numbers in each age/sex group by appropriate headship rates (the proportions of that group who are heads). This yields numbers of households in the city region classified by the characteristics of the head. The method suffers from a lack of good data on headship rates. In addition, it would be useful to produce a more complex "mapping" of population into household types. There is potential for further research in this area.

8. Conclusions.

This paper has outlined one of the key problems of population analysis in the city region context, that of forecasting its population. For some parts of the task sophisticated methods are available. Considerable progress has been made in recent years in the conceptual and mathematical formulation of cohort survival and accounting methods. Similar progress has yet to be made in other components of the task, and considerable intellectual and financial investment will have to be made.to push the development of population forecasting methods forward. However, such investment will be cheap compared with the cost of errors in public and private investment strategies that might stem from cruder guesses as to the future population of our city regions.

References

Cordey Hayes, M. and Gleave, M.B. (1974), Migration movements and the differential growth of city regions in England and Wales. Papers, Regional Science Association, 33, 99-123.

Donnison, D. (1967), The government of housing. Penguin, Harmondsworth, London.

Hall, P. et al (1973), The containment of urban England. Volume one; Allen and Unwin, London.

Illingworth, D.R. (1975), Demographic accounts and their relationship to the life histories of individuals. Working Paper 103, Department of Geography, University of Leeds.

Illingworth, D.R.; Smith, A.P.; and Rees, P.H. (1974), Notes on the definition of the West Yorkshire Study Area and the new West Yorkshire Metropolitan County. Working Paper 90, Department of Geography, University of Leeds.

Keyfitz, N. and Flieger, W. (1971), Population: facts and methods of demography. Freeman, San Francisco.

Leslie, P.H. (1945), On the use of matrices in certain population mathematics. Biometrika, 23, 183-212.

Leslie, P.H. (1948), Some further notes on the use of matrices in population mathematics. Biometrika, 35, 213-245.

Lexis, W. (1975), Einleitung in die Theorie der Bevolkerungs Statistik. Trubner, Strasbourg.

Matras, J. (1973), Populations and societies. Prentice Hall, Englewood Cliffs, N.J

Office of Population Censuses and Surveys (1971), Census 1971, England and Wales. Preliminary Report. H.M.S.O., London.

Office of Population Censuses and Surveys (1973), The Registrar General's Revised Estimates of the population of England and Wales, Regions and Local Authority Areas 1971 and 1972. H.M.S.O., London.

Office of Population Censuses and Surveys (1974), Population projections No.4, 1973 - 2013. H.M.S.O., London.

Rees, P.H.; Smith, A.P.; and King, J.R. (1975), Demographic models. Chapter 3 of a Final Report to the Social Science Research Council on Empirical and theoretical developments in models for cities and regions.

Rees, P.H. and Wilson, A.G. (1973), Accounts and models for spatial demographic analysis 1: aggregate population. Environment and Planning, 5, 61-90.

Rees, P.H. and Wilson, A.G. (1975a), Accounts and models for spatial demographic analysis 3: rates and life tables. Environment and Planning A, 7, 199-231.

Rees, P.H. and Wilson, A.G. (1975b), A comparison of available models of population change. Regional Studies, 9, 39-61.

Rees, P.H. and Wilson, A.G. (1975c), Spatial demographic analysis, Volume 1: Theory. Edward Arnold, London, forthcoming.