Population Statistics and Spatial Demographic Accounts
Author(s): A. G. Wilson and P. H. Rees
Source: *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 23, No. 3/4, Statistics and Geography (Sep. - Dec., 1974), pp. 229-257
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2987582
Accessed: 01-10-2020 01:46 UTC

Population Statistics and Spatial Demographic Accounts

A. G. WILSON and P. H. REES, *University of Leeds*

The availability of population statistics at various levels of spatial resolution in the U.K. is reviewed. These data are related to the elements of a population accounting scheme, and an account-based model, which can be used for historical analysis or for projection, is developed in detail.

Introduction

A major interest of geographers has always been the past, present and future spatial distributions of population. Until relatively recently, demographers have concerned themselves largely with single regions, while geographers have analysed spatial distributions without making the best use of demographic tools. Our concern here is to follow in the footsteps of such workers as Rogers (1966, 1968) in a field which may best be described as *spatial demographic analysis*.

We begin by giving a reasonably detailed description of available population statistics at various spatial scales. We then show how these can be related to concepts of demographic accounts, and that many spatial demographic models can be simply and directly expressed in relation to such an accounting format. We present our own model in detail and argue that it makes the best use of available statistics for both historical and projective purposes. It also leads us to make some suggestions in the final section about the desirability of revised methods of data collection and recording.

Current Population Statistics in the U.K.

To describe, model or project population change in a set of regions we need to know the population stocks of the regions, the inter-regional migration flows of people, the deaths of persons in the regions, and the numbers of children born there. If any one of these four sets of statistics is missing, the analysis possible of population change is less than satisfactory.

Information on *population stocks* is published in the periodic census reports of the Office of Population Censuses and Surveys (hereafter referred to as O.P.C.S.), formerly, the General Register Office (hereafter referred to as G.R.O.). Population counts are made at full decennial censuses such as those in 1961 and in 1971 (G.R.O. 1961 and O.P.C.S. 1971) or at sample censuses such as that in 1966 (G.R.O. London and G.R.O. Edinburgh, 1967).† Advanced analyses of full censuses provide the counts in preliminary form (O.P.C.S. London and G.R.O. Edinburgh 1972). Data on population stocks are also published every year in the "Annual Estimates of the Population of England and Wales and of Local Authority Areas", (O.P.C.S. 1972) but these are secondary estimates based on particular population accounting equations rather than actual counts.

Migration data are provided in the published tables of recent censuses (G.R.O. 1966, 1968a, 1968b, 1969; G.R.O. Edinburgh 1968 and 1969; Government of Northern Ireland, G.R.O. 1968) and will be provided in the latest census:

> "Migration statistics analysing persons and households whose address at census date was different from their address one year and five years earlier will be published on a national and regional basis."
> (O.P.C.S. London and G.R.O. Edinburgh 1973, p. xxix)

Information on flows between all regions is given with the exception of emigration outside the U.K. International migration statistics (flows between the U.K. as a whole and other countries) are available on a quarterly basis in the "Registrar General Quarterly Returns" (O.P.C.S. 1973a) and on an annual basis in the "Registrar General's Statistical Review", Part II, Tables, Population (O.P.C.S. 1973c). Use is also made by the Registrar General, when preparing the annual estimates of population, of counts of moves recorded in the National Health Service Register when persons change their doctor and exchange their N.H.S. card. These data are not, however, published.

The population stock of a region is a count of the number of people living there at a point in time (say, midnight on census night). The population can be counted in a number of different ways depending on how visitors, the citizens of other countries, members of the armed forces, students at educational establishments and patients at long stay institutions are treated. Details of the procedures used are given at the beginning of any official census publication, and the researcher needs to pay special attention to these.

The distinction between actual and usual residents, in particular, is a crucial one. The count of all those people actually in a region on census night is called the enumerated population of the region. Most British population data refer to this concept. The count of all those people whose usual residence is a particular region, irrespective of their location on census

† The sources cited here are illustrative of the type of material available. One example of the published tables containing the kind of information referred to is cited.

night, is called the *usually resident* population of the region. This distinction is made clear in a table (Table 1) that classifies people in a region of interest according to their location on census night and their usual or permanent residence.

Table 1

Population stock concepts

		Location at census night		
		Region of interest, region i	*The rest of the world, region* R	*Totals*
Location of usual residence	Region of interest, region i	The "stay at home" population of region i	Visitors to region R from region i	Usually resident population of region i
	The rest of the world, region R	Visitors to region i from region R	The "stay at home" population of region R	Usually resident population of region R
Totals		Enumerated population of region i	Enumerated population of region R	

The most appropriate population concept to use in a population model is that of usual residence. The reason is simple. Statistics on migration, deaths and births are all gathered on a usual residence basis. Otherwise visits would get recorded as migrations, deaths would be concentrated in areas with more than their fair share of hospitals and a large proportion of births would be assigned to a few maternity clinics. The concept used for population stocks should correspond to that used in the "flow" statistics. However, the most detailed breakdown of the population by age and sex refers to the enumerated population, and in practice we must convert this to the usual residence definition by multiplying it by the ratio of the total usually resident population of the region to enumerated population for each sex.

There are three kinds of migration statistics available from census tables or quarterly returns: lifetime migration flows, migration and survival flows, and counts of person moves across borders.

Lifetime migration flows can be extracted from tables that classify the population by place of current residence and by place of birth. These statistics carry no information about when in relation to a period of interest the migration took place, although methods have been developed (Rogers and Von Rabenau 1971) to estimate intercensal migration from the lifetime migration figures of successive censuses. Migration and survival flows are much more useful in population models. They are essentially cross-classifications of the population by place of usual residence at the time of the census and by place of usual residence one year before or five years before. One might note that the migration tables do omit, usually, any mention of the migration behaviour of persons under one, or persons under five respectively, at the time of the census. Cross-classification of persons under one or under five by place of birth would, if carried out yield statistics on birth, migration and survival. We shall see later that these play an important role in population accounts.

Counts of moves are usually made with respect to international borders and the characteristics of the mover recorded at the time of the move. No follow up record is kept of what happens to the person who moves. Care has to be exercised in the use of such statistics: often they are used as if they were migration and survival statistics.

A clear distinction between the different kind of migration statistics can be made through use of a time–space diagram. In Figure 1 are plotted a sample of life lines of persons that migrate between two regions, region i of interest and R, the rest of the world, at some point in their life.

The person whose migration history corresponds to lifeline A in Figure 1 will be recorded as a lifetime migrant† in place-of-residence by place-of-birth census tables, as will the persons with lifelines B and C. Person B will, however, be recorded as a surviving migrant within the period of concern from time t to time $t+T$; person C might be recorded as a surviving infant migrant (born within the period) but is usually not so recorded. Person D is someone who moves into the region of interest but then subsequently moves out, within the same period. The various migration statistics associated with the sample of four persons whose migration histories are displayed in Figure 1 are:

Lifetime migrants	$=3$	
Surviving migrants in period	$=1$	(Surviving infant migration $=1$)
Moves into region i from region R in period	$=3$	

Only persons B and C contribute as migrants to the end of period population of region i. Use of lifetime migration statistics or move statistics will lead to error in assessing the in-migrant component of the end of period population.

† Note that the definition of "lifetime migrant" relates to place of birth and current residence.

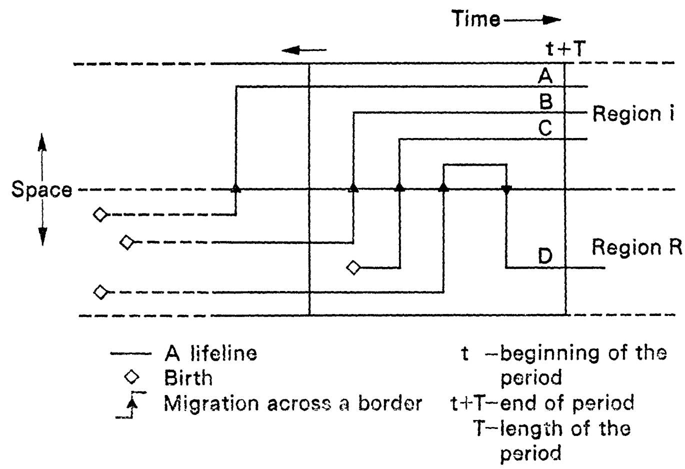

Figure 1

A time space diagram illustrating the different kinds of migration for which statistics are available

From a consideration of the concepts involved, we now turn to a description of the information on population stocks and migrant flows by age and sex. Almost all population statistics are disaggregated by sex. In this and subsequent sections it assumed, unless otherwise stated, that the statistics mentioned are available for males, females and their sum persons.

The age group information available for the population stocks of U.K. regions in published form at full censuses is specified in Table 2. Less detailed information is available in published form in the 1966 Sample Census or in the 1971 Advance Analysis. The available statistics match model requirements well: five year age groups can be used at all scales; one year age groups can be used at all scales. One year age groups can be used directly at the national scale, and the regional, local and small area five year age group populations can be deconsolidated into one year age groups using national proportions quite simply.

The migration statistics available from published tables in full or sample censuses are less comprehensive than the statistics available on population stocks. For example, migration flows to the rest of the world outside the U.K. are not available in the census.† Inter-regional migration flows are available in aggregate form, in published tables,

(1) amongst all countries within the U.K. (England, Wales, Scotland,

† Because to obtain them one would have to interview the migrants abroad at census date, an impossible task. Out-migrants to the rest of the world can be estimated from the counts of international migrants made by O.P.C.S. (1973c.)

Table 2

Age group information available at various scales at full census England and Wales (Census 1961, Census 1971)

National	*Regional*	*Local*	*Small area*	*National*	*Regional*	*Local*	*Small area*	*National*	*Regional*	*Local*	*Small area*
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
0	0	0	0	40				80			
1	1	1	1	41				81			
2	2	2	2	42	40–44	40–44	40–44	82	80–84	80–84	65 and over (cont.)
3	3	3	3	43				83			
4	4	4	4	44				84			
5	5	5	5	45				85			
6	6	6	6	46				86			
7	7	7	7	47	45–49	45–49	45–49	87	85–89	85–89	
8	8	8	8	48				88			
9	9	9	9	49				89			
10	10	10	10	50				90			
11	11	11	11	51				91			
12	12	12	12	52	50–54	50–54	50–54	92	90–94	90–94	
13	13	13	13	53				93			
14	14	14	14	54				94			
15	15	15	15	55				95			
16	16	16	16	56				96			
17	17	17	17	57	55–59	55–59	55–59	97	95 and over	95 and over	
18	18	18	18	58				98			
19	19	19	19	59				99			
20	20	20	20	60				100			
21	21	21	21	61				101			
22				62	60–64	60–64	60–64	102			
23	22–24	22–24	22–24	63				103			
24				64				104			
25				65				105			
26				66				106			
27	25–29	25–29	25–29	67	65–69	65–69	65 and over	107			
28				68				108			
29				69				109 and over			
30				70							
31				71							
32	30–34	30–34	30–34	72	70–74	70–74					
33				73							
34				74							
35				75							
36				76							
37	35–39	35–39	35–39	77	75–79	75–79					
38				78							
39				79							

Notes to Table 2

1. The numbers in the boxes refer to age at last birthday.

2. The solid boxes with numbers in refer to data published in census volumes by O.P.C.S. (or the G.R.O.). The dashed line boxes refer to data available in the form of computer printout or magnetic tape records for purchase from O.P.C.S. for the cost of processing the necessary files.

3. The scales refer to the following:

National —the U.K.
England and Wales, Scotland, Northern Ireland.

Regional —the standard regions of England and Wales, and the conurbations and region remainders, the regions of Scotland.

Local —Counties, local authority areas.

Small area—wards within local authority areas, parishes in rural districts, enumeration district.

4. Examples of the sources for the items in the columns indicated:

(1) G.R.O. 1964.
(2) The regional figures can be derived by aggregating the local data.
(3) G.R.O. (1963–64). O.P.C.S. (1972–73).
(4) O.P.C.S. supply this data on request.

5. One slight difference between 1961 Census and 1971 Census was that single years of age information was published for four years under 25 in 1971 but only under 22 in 1961, for regional and local scale areas.

Northern Ireland in G.R.O. 1968a; G.R.O. Edinburgh 1968 and Government of Northern Ireland, G.R.O. 1968).

(2) amongst standard regions within the U.K. (North, Northwest ... Wales II, Scotland in G.R.O. 1968a and G.R.O. Edinburgh 1968).

(3) amongst all counties in each country (English and Welsh counties, Scottish counties ... in G.R.O. 1968b, and G.R.O. Edinburgh 1968).

(4) amongst only a small selection of local authority areas where figures are greater than a certain threshold and where the local authority is larger than a certain size (G.R.O. 1968b).

Full matrices of inter local authority area flows within regions or countries can be obtained from O.P.C.S. (Gilje and Campbell 1973), or the missing flows can be estimated using entropy maximizing methods (Chilton and Poet 1973).

Much less information is published on migration flows in each age group. Table 3 shows what is available from the 1966 Sample Census, which contained questions designed to ascertain migration and survival over the one year and five year periods prior to census date, 24/25 April 1966. It is

Table 3

Age group information available at various scales in the Sample Census, 1966, England and Wales for total out-migration and total in-migration: five year migration

National scale (totals and inter-nation flows)		*Regional scale (totals and inter-region flows)*		*Local scale (total in- and out-migration)*		*Small area scale (totals)*		*Index No. of ages group*
$s-1$	s	$s-1$	s	$s-1$	s	$s-1$	s	
Birth	0–4							1
0–4	5–9	0–9	5–14	0–9	5–14	0–9	5–14	2
5–9	10–14							3
10–14	15–19	10–14	15–19	10–39	15–44	10–39	15–44	4
15–19	20–24	15–19	20–24					5
20–24	25–29	25–39	25–39					6
25–29	30–34							7
30–34	35–39							8
35–39	40–44							9
40–44	45–49	40–54	45–59	40–54	45–64	40–54	45–64	10
45–49	50–54							11
50–54	55–59							12
55–59	60–64	55–59	60–64					13
60–64	65–69	60 and over	65 and over	60 and over	65 and over	60 and over	65 and over	14
65–69	70–74							15
70–74	75–79							16
75–79	80–84							17
80–84	85–89							18
85–89	90–94							19
90 and over	95 and over							20

Notes to Table 3

1. The symbol s refers to age group at census date, 1966; $s-1$ to age group at census date 1961. Age groups are defined in terms of age ranges at last birthday.

2. The scales (national, regional, local, small area) are as defined in Table 2.

3. The dashed boxes in the national scale columns refer to unpublished data available from O.P.C.S. Data broken down by five year age group are also available at the regional and local scale from O.P.C.S.

4. Sources for the items at the scales indicated:

(1) National scale
(2) Regional scale } G.R.O., 1968a and 1969; G.R.O., Edinburgh, 1968 and 1969; Government of Northern Ireland, G.R.O., 1968.

(3) Local scale: G.R.O. 1968b and G.R.O., Edinburgh, 1968.

(4) Small area scale: obtainable from O.P.C.S.

5. Data for five year age groups is obtainable from O.P.C.S. at each of these scales, at cost, subject to disclosure rules.

important to note that the age groups referred to in the tables are those at the end of the migration period and that these must be "aged back" to obtain the age groups at the beginning of the period, which are the ones needed for calculating rates required in population models.

The published migration statistics suffer from three major deficiencies. The age groups used are too few and irregular, and do not match up with the much more detailed breakdown available for population. They are not available on an annual basis but only a periodic one. This makes for difficulties when studying and projecting trends in migration rates. And thirdly, an important flow, from a U.K. region to the rest of the world, is difficult to estimate, although it has been of considerable importance in the past.

Having examined the population stock and migration statistics available, we must now look at the nature of the vital statistics and the sorts of information available on births and deaths.

It is a legal requirement in the U.K. that births and deaths are registered and registration forms filed at local register offices. The births and deaths statistics generated from the registration forms are published in the Registrar General's Weekly Returns, the Registrar General's Quarterly Returns (O.P.C.S. 1973a) and in the Registrar General's Statistical Review (O.P.C.S. 1973b and 1973g each year) for the country concerned (England and Wales, Scotland, Northern Ireland).

Births and deaths totals are counts of events that take place within a specified period, in specified regions. The children born are classified by sex. Their mothers and fathers are classified by age *at time of birth*. Persons who die are classified by sex and age *at time of death*.

Little information is provided in the U.K. vital statistics about the prior location, age or other states of the mother or deceased. We shall see later how this information is vital in the proper accounting or modelling of population change. Figure 2 illustrates this problem with respect to prior location of "mother-to-be" and "those about to die". Mothers who gave birth in a region i of interest can be in-migrants from the rest of the world. Persons who die in a region can have started the period of interest in that region or in another. We must include many more people in the populations at risk of giving birth or of dying than demographic convention normally encompasses.

There are good reasons, of course, why such prior locational information is not recorded. It would be difficult to obtain from respondents and would need to be geared to a commonly agreed period of population accounting or projection. Prior age information is, however, easier to obtain simply by

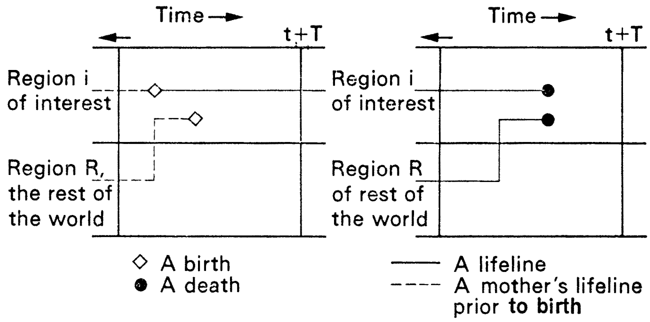

Figure 2

A time–space diagram illustrating how "mothers-to-be" and "those about to die" can begin the period in many regions

requiring the birth date of the mother or of the deceased, as well as the age at time of the appropriate event. For example, if we know that a mother is aged 25 in 1973 at the time of birth of her child she can either be a member of the 1947 cohort or the 1948. Again this should modify our calculation of the population at risk of giving birth or of dying in a particular age group, or we should move to a cohort view of births and deaths (cf. Pressat 1972). Mother's date of birth and the date of birth of the deceased are both collected and birth and death statistics at present just classified by age could easily be cross classified by cohort.

The age group information available in published and unpublished form for births and deaths at various scales is detailed in Tables 4 and 5. At the national scale the data for both births and deaths is in a single year form, from aged 11 to 55 and over for mothers and from aged 0 to 110 and over for persons dying. Published regional information is on a coarser mesh (five year age groups for mothers, ten year age groups for deaths) and there is no age disaggregation of local scale deaths. However, more detailed breakdowns can be obtained from O.P.C.S. At the small area scale no statistics are published although interested researchers have been granted access to local files, in order to compile statistical summaries for small areas.

It can be appreciated from the descriptions of the U.K. demographic statistics so far given that it is difficult to assemble them on a common time period, region and age group basis.

Population stock statistics are available in published form for April census dates every five years as actual counts and as estimates for 30 June every year. Migration and survival statistics are available for five year intercensal periods or for one year periods prior to a census. Birth statistics and deaths statistics are available on a calendar year basis at regional and local scales, and on a quarterly and weekly basis at the national scale.

Whether a single calendar year, a single mid point to mid point year, a single census date to census date year or a five year multiple of one of these is chosen, considerable effort has to be mounted to convert the non-conforming statistics into the form needed. This effort may involve expenditure on special tabulations from O.P.C.S. or it may consist of time spent in devising methods of estimating the data required in the population model from the data published. Migration statistics are the most difficult to convert from the published basis to any other because there is an as yet poorly understood non-linear relationship between migration and survival and the length of time period of measurement.

Difficulties with the definition of regions are rarely serious unless one wishes to adopt regions for study that do not coincide with local authority area boundaries. However, care must be taken in adjusting for regional boundary changes over the period of study. This will lead to particular difficulties in 1974 at the time of local government reorganization.

The matching of age groups in the various statistical series poses problems. The age groups used in published data for population stocks, migration flows, births and deaths do not match as they stand. Further information must be requested from O.P.C.S. in particular, on migration flows at all scales, on births and deaths at regional, local and small area scales if five year age groups are to be adopted. Further information and estimation methods must be applied if one year age group information is required. Note that the age group interval (one, five years) must match the length of the period chosen if survival calculations are not to become complicated.

There are other kinds of difficulties involved with the use of British population statistics, however. These concern the proper definitions of the demographic rates that enter any population model. All population models require the calculation of migration rates, birth rates and death rates. Death rates, for example, conventionally involve one of two alternatives

$$\begin{array}{c}\text{death rate in an age group}\\ \text{in a region in a period}\end{array} = \frac{\begin{array}{c}\text{total deaths in an age group in a}\\ \text{region in a period}\end{array}}{\begin{array}{c}\text{the start of period population in the age}\\ \text{group in the region}\end{array}}$$

or

$$\begin{array}{c}\text{death rate in an age group}\\ \text{in a region in a period}\end{array} = \frac{\begin{array}{c}\text{total deaths in an age group in a}\\ \text{region in a period}\end{array}}{\begin{array}{c}\text{the mid-point or mid-period population}\\ \text{in the age group in the region}\end{array}}$$

Table 4

Information available on age group of mother at birth at various scales, England and Wales

National	*Regional*	*Local*	*National*	*Regional*	*Local*
(1)	(2)	(3)	(1)	(2)	(3)
11			35		
12			36		
13			37	35–39	35–39
14			38		
15	Under 20	Under 20	39		
16			40		
17			41		
18			42	40–44	40–44
19			43		
20			44		
21			45		
22	20–24	20–24	46		
23			47		
24			48	45 and over	45 and over
25			49		
26			50		
27	25–29	25–29	51		
28			52		
29			53		
30			54		
31			55 and over		
32	30–34	30–34			
33					
34					

Notes to Table 4

1. The solid lines show what breakdown of births by age of mother is available in published sources. The broken lines indicate the age groups for which unpublished information is available.

2. Sources for items in the columns indicated (the sources are published at regular intervals).

(1) *National*. O.P.C.S. (1973c), Table AA(a), Maternities and births, by age of mother at birth, legitimacy and, for births, sex and whether live or still.

(2) *Regional*. O.P.C.S. (1973c), Table BB, Births by age of mother at birth, legitimacy and whether live or still.

(3) *Local*. O.P.C.S. (1973c), Table E, Home population, live births, still births, deaths, infant mortality and perinatal mortality. In the table births are disaggregated by sex but not by age. However, "The following unpublished items are available for reference at the Office of Population Censuses and Surveys, or copies can be supplied on payment:

(i) Live births and still births separately by sex, legitimacy and month of occurrence for each region, conurbation and local authority area in England and Wales.

(ii) Live births and still births separately by place of confinement, age and parity of mother for each county borough and administrative county in England and Wales" (O.P.C.S., 1973c, p. ix).

The first definition is that used in cohort survival models, simple, single region or multi-region models (Rees and Wilson 1973b) to define survival rates. The rate denominator has to be the same population as the base population used in the model. However, as we have already seen in Figure 2, some of the persons recorded as dying in the numerator (total deaths) need not necessarily have originated in the denominator population. Conversely, some of the denominator population may have died in another region or in another age group.

The second death rate definition is employed in the calculation of historical rates. Its use in population projection models requires that these projections be iterative in form. On first iteration an approximate value for the mid-point or mid-period population can be employed and an initial estimate made of the end of period population. This can then be used to define first estimate of the value for the mid period population to use in the projection model to calculate the number of deaths in the projected period. The process is repeated until successive values of the mid-period population are very nearly equal. Normally, however, iteration is not used. Rather the death rate is corrected so that it approximates to the first definition

$$\begin{array}{c}\text{death rate in an age group}\\ \text{in a region in a period}\end{array} = \frac{\begin{array}{c}\text{total deaths in an age group in a}\\ \text{region in a period}\end{array}}{\begin{array}{c}\text{mid-point or mid-period population plus half}\\ \text{of the total deaths in an age group in a}\\ \text{region in a period}\end{array}}$$

Table 5

Information available on age group at death at various scales, England and Wales

National	*Regional*	*Local*	*National*	*Regional*	*Local*	*National*	*Regional*	*Local*
(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
0	0		40	35–44		80		
1			41			81		
2	1–4		42			82		
3			43			83		
4			44			84		
5			45			85		
6			46			86		
7			47			87		
8			48			88		
9			49			89		
10	5–14		50	45–59		90		
11			51			91		
12			52			92		
13			53			93		
14			54			94		
15			55			95		
16			56			96		
17			57			97		
18			58			98		
19			59			99		
20	15–24		60	55–64		100		
21			61			101		
22			62			102		
23			63			103		
24			64			104		
25			65			105		
26			66			106		
27			67			107		
28			68			108		
29			69			109		
30	25–34		70	65–74		110 and over		
31			71					
32			72					
33			73					
34			74					
35			75					
36			76	75 and over				
37			77					
38			78					
39			79					

Notes to Table 5

1. The solid lines show what breakdown of deaths by age is available in published sources. The broken lines indicate the age groups for which unpublished information is easily available.

2. Sources for items in the columns indicated (the sources are published at regular intervals).

(1) *National.* O.P.C.S. (1973b), Table 14, Deaths by age, sex and marital condition, England and Wales.

(2) *Regional.* The same reference as under (1) National, Table 19, Deaths by cause, sex and age-group and standardized Mortality Rates, 1971. England and Wales, standard regions, Wales, conurbations, urban and rural aggregates, hospital region.

(3) *Local.* The same reference as under (1) National, Table 13, Home population, live births, still births, deaths, infant mortality and perinatal mortality. England and Wales, standard region, subdivisions of standard regions, Wales, conurbations, urban and rural aggregates, hospital regions, countries, Greater London, City of London, London and County boroughs, urban and rural districts. This table gives deaths disaggregated by sex only but "Deaths by sex and five year age groups for each county borough, London borough municipal borough, urban district and rural district in England and Wales", . . . "are available for reference at the Office of Population Censuses and Surveys, or copies can be supplied on payments" (O.P.C.S., 1973b, p. vi).

3. There is also very detailed information about deaths under 1 year of age in Table 17 of the Medical Tables.

This correction leads us back to the difficulties associated with the first definition of the death rate. Use of the second death rate and an iterative procedure involves fewer conceptual difficulties, as the mid-point or mid-period population does allow for a contribution from migrants and from persons not in the age group at the beginning of the period. However, if the in-migration stream, is very different in size from the out-migration stream, the mid-point or mid-period population is a less adequate surrogate for the time population at risk of dying in an age group in a region.

Exactly the same difficulties occur when birth rates are defined as

$$\begin{array}{l}\text{birth rate in a region to}\\ \text{mothers in an age group}\end{array} = \frac{\text{total births in a region to mothers in an age group}}{\text{females in an age group in a region at the start of a period}}$$

or

$$\begin{array}{l}\text{birth rate in a region to}\\ \text{mothers in an age group}\end{array} = \frac{\text{total births in a region to mothers in an age group}}{\text{the mid-point or mid-period female population in an age group in a region}}$$

Migration and survival rates, when used in population models, are normally defined as

$$\begin{array}{c}\text{migration rate between two} \\ \text{regions from one age group} \\ \text{to another}\end{array} = \frac{\begin{array}{c}\text{surviving migrants in an age group in the} \\ \text{destination region}\end{array}}{\begin{array}{c}\text{population in the age group in the origin} \\ \text{region at the beginning of the period}\end{array}}$$

This definition is satisfactory as long as the age groups are matched correctly. For example, the migration rate of persons aged 0–4 at the beginning of the period and 5–9 at the end from Greater London to the West Riding of Yorkshire over a five year period should involve persons recorded in age group 5–9 years at the end of the period in the West Riding who were in the 0–4 year old population of Greater London in 1961. Very often this correct definition of migration rate is not used and the number of surviving migrants will get divided, incorrectly, by the end of period population of the origin region or of the destination region.

A considerable body of valuable statistics on population and the components of population change exists for U.K. regions. Great care has to be exercised, however, in preparing the inputs to population accounting or projection models in a consistent fashion. There are also a number of conceptual difficulties connected with the way populations move between regions and from one age group to another over time that make conventional rates and models less than satisfactory.

In order to solve some of these difficulties and to use the available population statistics in the best way possible, we need to develop principles of population accounting and associated projection models that match the sets of population statistics consistently.

Principles of Accounting

We begin with the concept of a "state", and assume that each member of the population is in one and only one of a set of states at any time. For a population of given sex, we might typically take age and location to characterize states for demographic analysis. Accounting consists of being able to trace the history of each member of the population through state transitions in time. We will be concerned with a particular form of accounting which identifies the number of people in each state at a time t, together with their state at some subsequent time $t+T$. This information can be recorded as a matrix $\boldsymbol{K}=\{K^{ij}(t, t+T)\}$, whose (i, j)th element is the number of people who were in state i at time t and state j at time $t+T$. This form of accounting is particularly appropriate when major data sources are available which identify the numbers of people in each state at cross-sections in time, as does the Census of Population of course.

(In other cases, if data was available on the exact time at which each state transition took place, a different algebraic format for recording the information would be appropriate.)

Let $\boldsymbol{w}(t)=\{w^i(t)\}$ be a vector which gives the population totals in each state at time t:

$$w^i(t)=\sum_j K^{ij}(t, t+T) \tag{1}$$

$$w^i(t+T)=\sum_j K^{ji}(t, t+T) \tag{2}$$

Then, the usual form of rate-based analysis based on an accounting matrix and these "old" and "new" populations proceeds as follows. Define a matrix of rates, $\boldsymbol{G}$, whose (i, j)th element is

$$G^{ji}(t, t+T)=K^{ij}(t, t+T)/w^i(t) \tag{3}$$

That is, $\boldsymbol{G}'$ (where a prime denotes transposition) is a matrix of transition rates formed by dividing each element of the accounting matrix, $\boldsymbol{K}$, by the corresponding row sum. The following identity can then be seen to hold:

$$\boldsymbol{w}(t+T)=\boldsymbol{G}\boldsymbol{w}(t) \tag{4}$$

$\boldsymbol{G}$ is sometimes referred to as a "growth matrix". It is an operator which converts the old population into the new. However, an identity is not a model. Equation (4) becomes a model equation if the rates in the matrix $\boldsymbol{G}$ are supplied independently—that is, not using equation (3). Much demographic analysis is concerned with the estimation of such rates and their use in model equations such as (4) above.

The methods which have been used vary considerably and are not always presented in the matrix format of equation (4), though they can be converted into this form. They include the cohort survival model for a closed system (as described by Usher 1973), the single region cohort survival model/component model (Rogers 1968 and 1971; Gilje and Campbell 1973), the multi-region cohort survival model (Rogers 1968 and 1971), the projection model based on the multi-region life table (Rogers 1973). Stone (1972) has used an accounting method in a rather aggregative form.

We will show below that an alternative approach to the development of an account-based demographic model can be more fruitful, and enables some common mistakes in rate definition to be avoided (Rees and Wilson 1973a; Wilson and Rees 1974). First, however, we discuss typical state definitions for demographic analysis.

Levels of Resolution and System Description

The definition of a state and hence the form of system description depends on the level of resolution to be adopted. The possible levels of resolution are heavily constrained by data availability. Indeed, it is a central feature of our work on the account-based model to be presented in the next section that such constraints are met. There are three elements of this model design decision to be dealt with concerned with time, age and location. We consider finite time periods t to $t+T$ which, typically, will be the times of the Census of Population, though they may occasionally refer to other times. We also consider age to be related to finite age groups which are numbered consecutively from 1, 2, 3, . . . up to R, say. The last age group is usually considered to be open ended. Thus, for five year age groups, 1, 2, 3, . . . R would be 0–4, 5–9, 10–14, . . . 95+, and R would be 20. Location will also be identified by regions of finite size. Demographic data is available, as we have shown, for quite small units such as Census Enumeration Districts. However, for such small units, demographic flows would have to be considered as stochastic variables rather than as "stable averages" which can be treated in a more deterministic way. Thus, this is likely to be the constraint on size of region, and the smallest regions to be considered in practice would probably be local authorities. They may be as large as is useful—say countries, or Economic Planning Regions within the U.K. We must also remember that a demographic system must always be closed, and so one region must always be created which is the "Rest of the world" in relation to the rest of the system. Typically, there will be problems in finding data for migration flows between the "rest of the world" region and the system of interest. Clearly, an important population characteristic is sex; for simplicity, however, we shall assume we are dealing with a single sex population, say of females. It is always straightforward to extend the models to both sexes, usually by adding an additional sex index. The *only* problem is that care has to be taken to recognize that births occur only to females.

For people who exist at time t and at time $t+T$, the above discussion shows that we can characterize their initial state by two indices, say i and r, denoting region and age group respectively at time t, and their final state by (j, s) giving location and age group at time $t+T$. We have to add to these definitions the possibilities of births and deaths. Someone who is born during the period t to $t+T$ in region i has an initial state characterized by the symbol $\beta(i)$. Someone who dies during the period t to $t+T$ in region j has a final state characterized by the symbol $\delta(j)$ and age group s, where s is the age group *at time of death* (and again this decision is made in relation to the way in which data is collected). There is no need to associate an age group of course with an initial birth state, but we take advantage of this and add "age group of mother" as part of such an initial state description. We can now construct, using these definitions, variables which will form a demographic accounting matrix as follows:

$K_{rs}^{ij}(t, t+T)$ is the number of people in age group r in region i at time t

who are in age group s in region j at time $t+T$. Note that if $i \neq j$, this represents surviving migrants.

$K_{rs}^{\beta(i)j}$ is the number of people born in region i to mothers aged r at time t who survive into age group s in region j at time $t+T$. If $i \neq j$, these represent surviving infant migrants.

$K_{rs}^{i\delta(j)}$ is the number of people in age group r in region i at time t who die in region j in age group s at time of death. If $i \neq j$ there was a migration before death, so such terms represent non-surviving migrants.

$K_{rs}^{\beta(i)\delta(j)}$ is the number of people born in region i to mothers in age group r at time t who die in region j while in age group s. If $i \neq j$, there has been a migration between birth and death, and such flows represent non-surviving infant migrants.

These elements can be analysed in a demographic accounting matrix, for N regions, as follows:

$$\boldsymbol{K}=\begin{bmatrix} \boldsymbol{K}^{11} & \boldsymbol{K}^{12} & \boldsymbol{K}^{1N} & \boldsymbol{K}^{1\delta(1)} & \boldsymbol{K}^{1\delta(2)} & \boldsymbol{K}^{1\delta(N)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{K}^{N1} & \boldsymbol{K}^{N2} & \boldsymbol{K}^{NN} & \boldsymbol{K}^{N\delta(1)} & \boldsymbol{K}^{N\delta(2)} & \boldsymbol{K}^{N\delta(N)} \\ & & & & & \\ \boldsymbol{K}^{\beta(1)1} & \boldsymbol{K}^{\beta(1)2} & \boldsymbol{K}^{\beta(1)N} & \boldsymbol{K}^{\beta(1)\delta(i)} & \boldsymbol{K}^{\beta(1)\delta(2)} & \boldsymbol{K}^{\beta(1)\delta(N)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{K}^{\beta(N)1} & \boldsymbol{K}^{\beta(N)2} & \boldsymbol{K}^{\beta(N)N} & \boldsymbol{K}^{\beta(N)\delta(1)} & \boldsymbol{K}^{\beta(N)\delta(2)} & \boldsymbol{K}^{\beta(N)\delta(N)} \end{bmatrix} \quad (5)$$

where $\boldsymbol{K}^{ij}$, $\boldsymbol{K}^{i\delta(j)}$, $\boldsymbol{K}^{\beta(i)j}$ and $\boldsymbol{K}^{\beta(i)\delta(j)}$ are sub matrices whose (r, s)th elements are K_{rs}^{ij}, $K_{rs}^{i\delta(j)}$, $K_{rs}^{\beta(i)j}$, and $K_{rs}^{\beta(i)\delta(j)}$. $\boldsymbol{K}$ is a $2NR \times 2NR$ matrix.

So far, our definitions have been quite general in the sense that the projection period T can take any value relative to the widths of each age group (which we can define to be $\Delta_1, \Delta_2, \ldots \Delta_R$, though Δ_R, we recall, is considered to be open ended. These relationships can be displayed on a Lexis (1875) diagram, as in Figure 3.

There is a rectangle for each age group, r, as shown, of width T and height Δ_r. A lifeline is represented as a downward sloping line at 45° to the horizontal. It intersects the left hand vertical at the exact starting age. Thus, using the Lexis diagram in an obvious way, we can see how the structure of the ageing process in period T is represented in relation to a particular set of age group definitions. We shall refer to Figure 3 as the "general" case. There is one particularly "simple" case which is often useful, defined by the property

$$\Delta_r = T \text{ for all } r \neq R \quad (6)$$

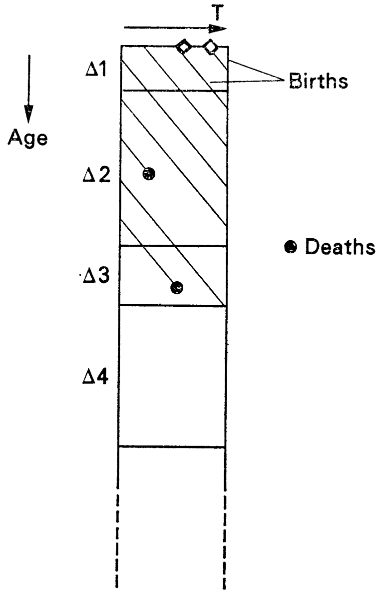

Figure 3

It can easily be seen that in such a case

(i) all people born in the period and surviving are recorded as being in age group 1 at the end of the period. Infants who die during the period must be in age group 1 at time of death;

(ii) people in age group r at time t who survive to time $t+T$ *must* be in age group $r+1$, the next highest, at time $t+T$ (unless they are in age group R at time t);

(iii) people who are in age group r at time t and who die during the period *must* be either in age group r or age group $r+1$ at time of death.

If this simple representation is used, it means that many of the elements of the submatrices of the accounting matrix $\boldsymbol{K}$ in equation (5) will be zero. For example, $K_{rs}{}^{ij}$ can only be *non-zero* if $s=r+1$ or $r=s=R$. We will usually use this simple representation to explain the concepts of our account-based model, though quite often, the data does not come in such a tidy form. Then, either the more general formulation has to be used or manipulation of the data carried out to get it into a suitable form.

The Account-based Model

Our own model has a resemblance to the usual transition rate model of equation (4). Indeed, it can be stated in this form. It is also worth noting that our work with this model originated in attempts to build a Roger's (1966) model for a British region (Rees 1971; Wilson 1972). Initially, our work was a response to problems of data availability and although we would claim that we have been led to features of the model which are in themselves more important, we have continued to observe the principle that the recommended model should match the data available. It would be possible, however, to use the principles on which the account-based model is based to amend it for use with other data sets.

We had three main kinds of worry with the traditional model of equation (4) which arose as follows. First, if an accounting matrix, $\boldsymbol{K}$, as defined earlier, is used as a basis for defining the matrix of transition rates, $\boldsymbol{G}$, it will be observed that there are many kinds of flows, for example the "migration and death" flows, which are not directly measured. Thus, the associated transition rates cannot be measured. Typically, such flows and rates are omitted from transition rate models, or at best dealt with indirectly in an imperfect manner. Secondly, even if all the transition rates appear in the model, many of them, in practice, are incorrectly measured using available data as we showed earlier. Thirdly, such rates may not be the best for predictive purposes. Our own solution to these problems involved working with the flows in the accounting matrix directly, and defining rates where, unlike those described below, the numerator *did* match the denominator. Further, we only define such rates where absolutely necessary.

The account-based model is based on two main sets of concepts and one additional assumption. First, we have the concept of *accounting equations*: each row and column of the accounting matrix in equation (5) sums to a total which is either known, or is a quantity which we want to estimate. These equations form the backbone of the model. Secondly, in relation to recorded births and deaths in a region we introduce the concept of "population-at-risk"—of giving birth and of dying respectively. We can then use these quantities as denominators to calculate appropriate birth and death rates. Thirdly, the major assumption is needed as follows: we distinguish between unknown major flows and unknown minor flows in the accounting matrix in a way which is consistent with common sense and is such that there are exactly as many usable accounting equations as unknown major flows. We then make an assumption about the rates at which "minor flow" events occur as a basis for estimating such flows. These concepts, and the assumption, then allow us to state the account-based model. The resulting set of equations must be solved iteratively.

The model can be described conveniently in seven steps, two of them preliminary, as follows:

Step 1: assemble known data;

Step 2: manipulate this where necessary into a form suitable for the model;

Step 3: obtain initial values of the unknown major flows by setting the unknown minor flows to zero in the usable accounting equations;

Step 4: calculate "at risk" populations;

Step 5: calculate birth and death rates;

Step 6: calculate unknown minor flows;

Step 7: solve the accounting equations for the unknown major flows and the new population totals.

Steps 4 to 7 must be repeated until convergence is achieved. We describe each of these steps below, initially for a historical period, and for simplicity using the "simple case" definition of age groups and time period of equation (6) above.

Step 1. The known data is assumed to consist of: K_{r*}^{i*}, the initial population; $K_{**}^{\beta(i)*}(K_{*u}^{*m(i)})$, which is the total births in region i to mothers aged u at time of birth (and where $K_{*u}^{*m(i)}$ is used as a label to indicate the characteristics of the mother); $K_{*s}^{*\delta(i)}$, total deaths in region i of persons aged s at time of death; K_{*s}^{ij}, $i \neq j$, the number of surviving migrants from i to j who are in age group s at the end of the period; and $K_{rs'}^{\beta(i)j}$, $i \neq j$, a similar quantity for surviving migrating infants, r in this case being age of mother. This data can be assembled, though not without difficulty, from the data sources described in section II.†

Step 2. The model requires the birth data to relate to age of mother at the beginning of the period rather than age at birth of the child. Mothers aged r at the beginning of the period could be in age groups r or $r+1$ at the time of

† There will always be adjustment problems, as outlined in section II. The main data sources—the Census and the Registrar General's Returns relate to different time periods and so must be "matched up", usually by adopting an inter-censal period as standard. There may be similar matching up in relation to regional definitions, especially if a fine level of resolution is adopted. There will almost always be problems with migration from the "rest of the world zone", for which diverse data sources will have to be drawn together, differing definitions of migration dealt with, and estimates made. The data may have to be adjusted for (*a*) under enumeration, (*b*) regional boundary changes and (*c*) the usual residence definition. Finally, if the "simple case" conventions are adopted some such method as the use of "deconsolidation operators" will have to be employed to transform "data age groups" to "model age groups". There is a fuller discussion of these issues in Rees and Wilson (1974).

birth. We let c_{rr} and c_{rr+1} be exogenously specified coefficients which are proportions such that

$$K_{r*}{}^{\beta(i)*}=\sum_{u=r}^{r+1} c_{ru}K_{**}{}^{\beta(i)*}(K_{*u}{}^{*m(i)}) \tag{7}$$

We also require migration data to relate to age at the beginning of the period. In the simple case, we have

$$K_{r-1*}{}^{ij}=K_{*r}{}^{ij}, \qquad r<R \tag{8}$$

For the last age group, we could define coefficients q_{R-1R} and q_{RR} such that

$$K_{R-1*}{}^{ij}=q_{R-1R}K_{*R}{}^{ij} \tag{9}$$

and

$$K_{R*}{}^{ij}=q_{RR}K_{*R}{}^{ij} \tag{10}$$

These quantities can henceforth be treated as known "data".

Step 3. The accounting equations for the simple case can be written as follows. Each of the first NR rows sum to the initial population:

$$K_{rr+1}{}^{ii}+\sum_{j\neq i} K_{rr+1}{}^{ij}+K_{rr}{}^{i\delta(i)}+K_{rr+1}{}^{i\delta(i)} + \sum_{j\neq i}(K_{rr}{}^{i\delta(j)}+K_{rr+1}{}^{i\delta(j)})=K_{r*}{}^{i*}, \qquad r<R \tag{11}$$

For the last age group, R, we have

$$K_{RR}{}^{ii}+\sum_{j\neq i} K_{RR}{}^{ij}+K_{RR}{}^{i\delta(i)}+\sum_{j\neq i} K_{RR}{}^{i\delta(j)}=K_{R*}{}^{i*} \tag{12}$$

Note that in these and subsequent equations, we distinguish the $i-i$ terms from the $i-j$, $i\neq j$, terms, as this will help us to differentiate major and minor flows.

The next NR rows sum to total births. There are entries only for the child bearing age groups, the upper and lower limits of which we define to be λ and μ respectively:

$$K_{r1}{}^{\beta(i)i}+\sum_{j\neq i} K_{r1}{}^{\beta(i)j}+K_{r1}{}^{\beta(i)\delta(i)}+\sum_{j\neq i} K_{r1}{}^{\beta(i)\delta(j)}=K_{r1}{}^{\beta(i)*}, \qquad \lambda\leqslant r\leqslant\mu \tag{13}$$

The second NR columns sum to total recorded deaths. Those in age group 1 at time of death may be in age group 1 at the beginning of the period, or born during the period, and that equation is

$$K_{11}{}^{i\delta(i)}+\sum_{j\neq i} K_{11}{}^{j\delta(i)}+\sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)\delta(i)}+\sum_{r=\lambda}^{\mu}\sum_{j\neq i} K_{r1}{}^{\beta(j)\delta(i)}=K_{*1}{}^{*\delta(i)} \tag{14}$$

For age groups $s>1$, we have

$$K_{s-1s}{}^{i\delta(i)}+K_{ss}{}^{i\delta(i)}+\sum_{j\neq i}(K_{s-1s}{}^{j\delta(i)}+K_{ss}{}^{j\delta(1)})=K_{*s}{}^{*\delta(i)} \tag{15}$$

The above equations represent the "useable" accounting equations at this stage of the calculation. The first NR columns sum to the "new population" totals, and we reserve these for tests of the model. However, for completeness here and for later use, they are presented as follows.

$$K_{s-1s}{}^{ii} + \sum_{j \neq i} K_{s-1s}{}^{ji} = K_{*s}{}^{*i}, \qquad 1 < s < R \tag{16}$$

$$K_{R-1R}{}^{ii} + K_{RR}{}^{ii} + \sum_{j \neq i} (K_{R-1R}{}^{ji} + K_{RR}{}^{ji}) = K_{*R}{}^{*i} \tag{17}$$

$$\sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)i} + \sum_{r=\lambda}^{\mu} \sum_{j \neq i} K_{r1}{}^{\beta(j)i} = K_{*1}{}^{*i} \tag{18}$$

The unknown flows we designate as major flows are: $K_{rr+1}{}^{ii}$, $K_{rr}{}^{i\delta(i)}$, $K_{rr+1}{}^{i\delta(i)}$, $K_{r1}{}^{\beta(i)i}$, $K_{r1}{}^{\beta(i)\delta(i)}$. The unknown minor flows are: $K_{rr}{}^{i\delta(j)}$, $K_{rr+1}{}^{i\delta(j)}$, $K_{r1}{}^{\beta(i)\delta(j)}$. All other flows are "known" from data. Thus, for this step, we set the unknown minor flows to zero and solve for the unknown major flows as follows. We begin with the "death equations" (14) and (15). In this case, we still have more unknowns than equations, and so we solve for a sum of two or more terms, and then use exogenously supplied coefficients to disaggregate further. From equation (14) we get

$$K_{11}{}^{1\delta(i)} + \sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)\delta(i)} = K_{*1}{}^{*\delta(i)} - \sum_{j \neq i} K_{11}{}^{j\delta(i)} - \sum_{r=\lambda}^{\mu} \sum_{j \neq i} K_{r1}{}^{\beta(j)\delta(i)} \tag{19}$$

(with the minor flows set to zero though they are still shown, here and in subsequent equations, for later convenience), with the subscript 0 denoting the birth state,

$$\sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)\delta(i)} = c_{01} K_{*1}{}^{*\delta(i)} \tag{20}$$

$$K_{11}{}^{i\delta(i)} = c_{11} K_{*1}{}^{*\delta(i)} \tag{21}$$

We also assume the existence of a set of coefficients e_r giving the proportion of dying infants whose mothers were in age group r at the beginning of the period. Then

$$K_{r1}{}^{\beta(i)\delta(i)} = e_r \sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)\delta(i)} \tag{22}$$

Similarly, from equation (15):

$$K_{s-1s}{}^{i\delta(1)} + K_{ss}{}^{i\delta(i)} = K_{*s}{}^{i\delta(i)} - \sum_{j \neq i} (K_{s-1s}{}^{j\delta(i)} + K_{ss}{}^{j\delta(i)}) \tag{23}$$

and

$$K_{s-1s}{}^{i\delta(i)} = c_{s-1s} K_{*s}{}^{i\delta(i)} \tag{24}$$

$$K_{ss}{}^{i\delta(i)} = c_{ss} K_{*s}{}^{i\delta(i)} \tag{25}$$

We can now use the "births" equation (13) to give

$$K_{r1}{}^{\beta(i)i} = K_{r1}{}^{\beta(i)*} - \sum_{j\neq i} K_{r1}{}^{\beta(i)j} - K_{r1}{}^{\beta(i)\delta(i)} - \sum_{j\neq i} K_{r1}{}^{\beta(i)\delta(j)}, \qquad \lambda \leqslant r \leqslant \mu \tag{26}$$

and equations (11) and (12) to give

$$K_{rr+1}{}^{ii} = K_{r*}{}^{i*} - \sum_{j\neq j} K_{rr+1}{}^{ij} - K_{rr}{}^{i\delta(i)} - K_{rr+1}{}^{i\delta(i)} - \sum_{j\neq i} (K_{rr}{}^{i\delta(j)} + K_{rr+1}{}^{i\delta(j)}), \qquad r < R \tag{27}$$

$$K_{RR}{}^{ii} = K_{R*}{}^{i*} - \sum_{j\neq i} K_{RR}{}^{ij} - K_{RR}{}^{i\delta(i)} - \sum_{j\neq i} K_{RR}{}^{i\delta(j)} \tag{28}$$

and this completes our estimation of the initial values of the unknown major flows.

Step 4. We now show how to calculate at risk populations in relation to births and deaths. Strictly, only the "deaths" populations are needed for the model used for an historical analysis, but we present the "births" case also for later convenience. We define the population in age group r at the beginning of the period at risk of giving birth in region i during the period as $\hat{K}_{r*}{}^{B*i}$, and it is made up of a weighted sum of population flows $K_{r*}{}^{jk}$, where j and k range over all states (except birth states for j, since we assume that a person born during the period cannot generate a birth in the same period). The weights are defined to be ${}^{i}\theta^{Bjk}$—the mean proportion of the period which a j, k flow is "at risk" for giving birth in i. Thus, formally,

$$\hat{K}_{r*}{}^{B*i} = \sum_{jk} {}^{i}\theta^{Bjk} K_{r*}{}^{jk} \tag{29}$$

We show elsewhere (Wilson and Rees 1974; Rees and Wilson 1974) how these θs can be calculated. Similarly, we define $\hat{K}_{rs}{}^{D*i}$ ($s = r$ or $r+1$) as the population in age group r at the beginning of the period, at risk of dying in region i, age s. We can also define

$$\hat{K}_{*s}{}^{D*i} = \sum_{r} \hat{K}_{rs}{}^{D*i} = \hat{K}_{s-1s}{}^{D*i} + \hat{K}_{ss}{}^{D*i} \tag{30}$$

We let $r = 0$ denote the birth state when used as the first subscript of $\hat{K}_{rs}{}^{D*i}$. Then defining a new set of coefficients, we have

$$\hat{K}_{0s}{}^{D*i} = \sum_{j\varepsilon B, k, u} {}_{s}{}^{i}\theta_{0u}{}^{Djk} K_{*u}{}^{jk} \tag{31}$$

$$\hat{K}_{rs}{}^{D*i} = \sum_{j\varepsilon NB, k, u} {}_{s}{}^{i}\theta_{ru}{}^{Djk} K_{ru}{}^{jk}, \qquad r > 0 \tag{32}$$

$\hat{K}_{r*}{}^{B*i}$, and $\hat{K}_{*}{}^{D*i}$ are the populations at risk which correspond to total regional births $K_{r*}{}^{\beta(i)*}$ and total regional deaths $K_{*s}{}^{*\delta(i)}$ and will lead us to appropriate rate definitions in the next step.

Step 5. We define the necessary birth and death rates as follows.†

$$b_{r*}{}^{*i} = K_{r*}{}^{\beta(i)*}/\hat{K}_{r*}{}^{B*i} \tag{33}$$

$$d_{rs}{}^{*i} = K_{rs}{}^{*\delta(i)}/\hat{K}_{rs}{}^{D*i} \tag{34}$$

$$d_{*s}{}^{*i} = K_{*s}{}^{*\delta(i)}/\hat{K}_{*s}{}^{D*i} \tag{35}$$

It can now be seen that the pattern of subscripts and asterisks used on the rate and at risk population variables reflect those on the main flow variable in the numerator of the corresponding rate definition. This notation emphasizes the fact that numerators and denominators do match.

Step 6. We can now calculate unknown minor flows by a further development of the concept of at risk population, and using the rates defined in the previous step. The flows to be calculated are $K_{rr}{}^{i\delta(j)}$, $K_{rr+1}{}^{i\delta(j)}$, $i \neq j$, and $K_{r1}{}^{\beta(i)\delta(j)}$, $i \neq j$. We can define, using a now obvious notation, $\hat{K}_{rr}{}^{Di\delta(j)}$, $\hat{K}_{rr+1}{}^{Di\delta(j)}$ and $\hat{K}_{01}{}^{D\beta(i)\delta(j)}$ as the at risk populations for these deaths. Each of these is, of course, a subset of the at risk populations defined in equations (31) and (32), and so they can be found in a straightforward way. We then assume that the death rates found in Step 5 apply to these subjects of deaths. This gives the following equations:

$$K_{rr}{}^{i\delta(j)} = d_{rr}{}^{*j}\hat{K}_{rr}{}^{Di\delta(j)} \tag{36}$$

$$K_{rr+1}{}^{i\delta(j)} = d_{rr+1}{}^{*j}\hat{K}_{rr+1}{}^{Di\delta(j)} \tag{37}$$

$$K_{r1}{}^{\beta(i)\delta(j)} = e_r d_{01}{}^{*j}\hat{K}_{01}{}^{D\beta(i)\delta(j)} \tag{38}$$

where, in the last equation, we use the coefficients e_r, introduced earlier, to split the deaths by age of mother at the beginning of the period. These equations can be solved for the unknown minor flows, though not completely straightforwardly as the left hand side unknowns also appear in the at risk population terms on the right hand side. The details are given elsewhere (Rees and Wilson 1974).

Step 7. Finally, the accounting equations can be solved for the major flows. Rather than repeat the equations, we can note that the solution procedure follows that of Step 3, though now the unknown minor flows have non-zero values. The procedure can be summarized as follows:

(i) solve equation (19) for $K_{11}{}^{i\delta(i)} + \sum_{r=\lambda}^{\mu} K_{r1}{}^{\beta(i)\delta(i)}$, and use c_{01}, c_{11} and e_r coefficients in equations (20)–(22) to obtain $K_{11}{}^{i\delta(i)}$, $K_{r1}{}^{\beta(i)\delta(j)}$.

(ii) solve equation (23) for $K_{s-1s}{}^{i\delta(i)} + K_{ss}{}^{i\delta(i)}$ and use c_{s-1s} and c_{ss} coefficients in equations (24) and (25) to obtain $K_{s-1s}{}^{i\delta(i)}$ and $K_{ss}{}^{i\delta(i)}$.

---

† $K_{*s}{}^{*\delta(i)}$ is the "deaths" quantity available from data. But we assume $K_{s-1s}{}^{*\delta(i)}$ and $K_{ss}{}^{*\delta(i)}$ can be obtained by applying the coefficients c_{s-1s} and c_{ss}.

(iii) solve equation (26) for $K_{r1}{}^{\beta(i)i}$;

(iv) solve equations (27) and (28) for $K_{rr*1}{}^{ii}$ and $K_{RR}{}^{ii}$;

(v) solve equations (16)–(18) for the "new populations", $K_{*s}{}^{*i}$.

Concluding Comments

We have shown that, for a historical period, we can start with a set of available data and incomplete population accounts and use our account-based model to estimate the remaining elements. If future birth and death rates, and migration flows, from exogenous assumptions or submodels, are supplied instead of total births and deaths, we can use the model to make population projections. We mentioned earlier, and we have shown in detail elsewhere (Rees and Wilson 1973b) that the account-based model can be compared with other models by writing it in the form of equation (4). Such comparison shows that, in general, we have identified and "accounted for" some minor flows which are not adequately dealt with in other models, but more importantly, we have based our model on a theoretically sound set of demographic rates for both historical and projection purposes. We argue that, because of the interdependence of the elements of the account specified through the accounting equations—the conservation laws of the population system—this framework enables the best use to be made of available statistics. Tests of the model, to be reported elsewhere (Smith and Rees 1974) are very encouraging also.

It also brings to the fore the need for changes in the methods of data collection and recording and for some associated further theoretical work. For all data, there is a need to achieve greater consistency in relation to time periods, region and age-groups definitions. Or, at least, the Registrar General could publish "estimated" data drawn together from various sources which were consistent. In the case of births and deaths, we emphasized below the need for a "cohort view" to be taken. If this was done, we would have no need for c_{rs} coefficients in the model presented later. In the case of migration flows, we recognize that data will be collected which relates to different time period lengths, and that, especially in relation to the "rest of the world" zone, data will relate to "crossing counts" rather than a "change of address between times" definition. More theoretical work is needed to relate these different kinds of migration data more effectively.

One final comment could be made of a different kind: although our main interest in this paper has been concerned with spatial demographic analysis, the model building methods described should be applicable in a variety of other fields—from subsets of human populations—immigrant communities, students in a University, manpower in a factory, to other kinds of populations, such as machine tools.

REFERENCES

Sources of population statistics for U.K. regions have been marked with an asterisk. Reference is made to a particular volume, but a note is attached in parentheses at the end of the citation to indicate its periodicity where appropriate. The abbreviation O.P.C.S. refers to the Office of Population Censuses and Surveys and the abbreviation G.R.O. to General Register Office.

CHILTON, R. and POET, R. R. W. (1973). An entropy maximizing approach to the recovery of detailed migration patterns from aggregate census data. *Environment and Planning*, **5**, 135–46.

*G.R.O. (1961). *Census 1961. England and Wales Preliminary Report.* London: H.M.S.O.

*G.R.O. (1963–64). *Census 1961. England and Wales, County Reports.* London: H.M.S.O.

*G.R.O. (1964). *Census 1961. England and Wales. Age, Marital Condition and General Tables.* London: H.M.S.O.

*G.R.O. (1966). *Census 1961. England and Wales. Migration Tables.* London: H.M.S.O.

*G.R.O. (1968a). *Sample Census 1966. England and Wales. Migration Summary Tables, Part I.* London: H.M.S.O.

*G.R.O. (1968b). *Sample Census 1966. England and Wales. Migration Regional Report. Yorkshire and Humberside.* London: H.M.S.O.

*G.R.O. (1969). *Sample Census 1966. England and Wales. Migration Tables, Part II.* London: H.M.S.O.

*G.R.O., Edinburgh (1968). *Sample Census 1966. Scotland. Migration Tables, Part I.* Edinburgh: H.M.S.O.

*G.R.O., Edinburgh (1969). *Sample Census 1966. Scotland. Migration Tables, Part II.* Edinburgh: H.M.S.O.

*G.R.O., London and G.R.O., Edinburgh (1967). *Sample Census 1966. Great Britain. Summary Tables.* London: H.M.S.O.

GILJE, E. J. and CAMPBELL, R. A. (1973). A new model for projecting the population of Greater London boroughs. Department of Planning and Transportation, Greater London Council. Paper presented at the Population Projections Seminar, Planning Transportation Research Co., New Year Meeting, University of Sussex, January 1973.

*GOVERNMENT OF NORTHERN IRELAND, G.R.O. (1968). *Census of Population 1966, General Report.* Belfast: H.M.S.O.

LEXIS, W. (1875). *Einleitung in die Theorie de Bevölkerungs—Statistik.* Trubner. Strasbourg. Quoted in Keyfitz, N. (1968). *Introduction to the Mathematics of Population*, 9–11. Reading, Mass.: Addison-Wesley.

*O.P.C.S. (1971). *Census 1971. England and Wales. Preliminary Report.* London: H.M.S.O.

*O.P.C.S. (1972). *The Registrar General's Annual Estimates of the Population of England and Wales and of Local Authority Areas, 1972.* London: H.M.S.O.

*O.P.C.S. (1972–73). *Census 1971. England and Wales. County Reports.* London: H.M.S.O. In three parts: Parts I, II and III.

*O.P.C.S. (1973a). *The Registrar General's Quarterly Return for England and Wales. Quarter Ended 30 June 1973.* No. 498. 2nd Quarter, 1973. London: H.M.S.O.

*O.P.C.S. (1973b). *The Registrar General's Statistical Review of England and Wales for the Year 1971. Part I, Tables, Medical.* London: H.M.S.O.

*O.P.C.S. (1973c). *The Registrar General's Statistical Review of England and Wales for the Year 1971. Part II, Tables, Population.* London: H.M.S.O.

*O.P.C.S., London and G.R.O., Edinburgh (1972). *Census 1971. Great Britain. Advance Analysis.* London: H.M.S.O.

*O.P.C.S., London and G.R.O., Edinburgh (1973). *Census* 1971. *Great Britain. Summary Tables (1 per cent Sample).* London: H.M.S.O.

PRESSAT, R. (1972). *Demographic Analysis.* London: Edward Arnold.

REES, P. H. (1971). How to make the Rogers version of the demographic model operational for an English or Welsh county in about one hundred easy steps. *Working Paper 8,* Department of Geography, University of Leeds.

REES, P. H. and WILSON, A. G. (1973a). Accounts and models for spatial demographic analysis. I: Aggregate population. *Environment and Planning,* **5,** 61–90.

REES, P. H. and WILSON, A. G. (1973b). A comparison of available models of population growth. *Working Paper 54,* Department of Geography, University of Leeds.

REES, P. H. and WILSON, A. G. (1974). *Spatial Demographic Analysis.* Monograph in preparation.

ROGERS, A. (1966). Matrix methods of population analysis. *Journal of the American Institute of Planners,* **32,** 40–4.

ROGERS, A. (1968). *Matrix Analysis of Interregional Population Growth and Distribution.* University of California Press, Berkeley and Los Angeles.

ROGERS, A. (1971). *Matrix Methods in Urban and Regional Analysis.* San Francisco: Holden Day.

ROGERS, A. (1973). The mathematics of multiregional demographic growth. *Environment and Planning,* **5,** 3–29.

ROGERS, A. and VON RABENAU, B. (1971). Estimation of interregional migration streams from place of birth by residence data. *Demography,* **8,** 185–94.

SMITH, A. P. and REES, P. H. (1974). Methods of constructing spatial demographic accounts. *Working Paper 59,* Department of Geography, University of Leeds.

STONE, R. (1972). A Markovian education model and other examples linking social behaviour to the economy. *Journal of the Royal Statistical Society,* A, **135,** 511–43.

USHER, M. B. (1972). Developments in the Leslie matrix model. In *Mathematical Models in Ecology,* J. N. R. Jeffers (ed.), 29–60. Oxford: Blackwell's Scientific Publications.

WILSON, A. G. (1972). Multiregional models of population structure and some implications for a dynamic residential location model. *Patterns and Processes in Urban and Regional Systems,* pp. 217–42. London: Pion.

WILSON, A. G. and REES, P. H. (1974). Accounts and models for spatial demographic analysis. II: Age–sex disaggregated populations. *Environment and Planning A,* **6,** 101–16.