Working Paper 511

Linking Census and NHSCR migration data

Peter Boden, John Stillwell and Philip Rees
May, 1988

School of Geography
University of Leeds
Leeds LS2 9JT

CONTENTS

Abstract

The National Health Service Central Register and the Census provide data on internal migration within the United Kingdom. A previous paper (Boden, Stillwell and Rees, 1987a) reported results from a preliminary comparison of the two data sources. This paper extracts results from the preliminary work and compares them with results from a more accurate analysis using NHSCR information accessed from primary unit data (PUD). In this second stage of the comparison a number of more precise alignment and adjustment procedures are adopted which include the allocation of Armed Forces moves to individual FPCAs, the assignment of all age, sex and origin not-stated flows and the matching of the NHSCR age-time plan of observation to that of the Census. The paper illustrates NHSCR-Census differences at a number of spatial scales and levels of age and sex disaggregation for both stages of the comparison.

The general patterns of NHSCR and Census migration are similar although movement from the NHSCR is recorded at a higher level. The NHSCR-Census ratio varies considerably at sub-national levels with the assignment of Armed Forces flows being particularly influential upon non-metropolitan flows in the 15-19 age-range. Explanations for the differences that occur are put forward in terms of the conceptual, operational measurement and population at risk discrepancies that exist between the NHSCR and the Census.

Acknowledgement

The authors wish to acknowledge the assistance of the Office of Population Censuses and Surveys in providing the Census and NHSCR data used in the research reported in this paper.

List of Tables

List of Figures

1. INTRODUCTION

Projections of sub-national population in Great Britain depend increasingly on our ability to predict inter-area migration, yet our knowledge and understanding of current trends in migration levels and patterns continues to be limited by inadequate data. The Census of Population provides the most reliable and comprehensive migration information, but during inter-censal periods it is necessary to rely on migration data from other sources. The importance of NHS patient re-registrations contained in the National Health Service Central Register (NHSCR) is recognised by the Office of Population Censuses and Surveys (OPCS), and the migration assumptions prepared for the official sub-national population forecasting model are derived by using NHSCR data as a means of updating zone-specific gross migraproduction rates (GMRs). Other multi-regional models have been developed to project populations on the basis of sets of either transition or movement accounts (Rees, 1981,1984a). In order to incorporate both Census and NHSCR data into a projection model, careful identification of the characteristics of both types of data and accurate estimation of the statistical relationships between them are essential.

This paper outlines some research which compares migration data from the two alternative sources. Detailed results have already been reported in Boden, Stillwell and Rees (1987a, 1987b). In Section 2 comparative studies by other authors are reviewed, important differences that exist between Census and NHSCR data are systematically described, and the availability of data used for comparison is briefly outlined. Our comparative work has been undertaken in two stages with the first stage utilising computer summaries of NHSCR primary unit data (PUD) and stage two accessing the PUD directly. Section 3 describes the alternative assignment and adjustment techniques adopted in order to 'match' the respective datasets, and Section 4 illustrates a selection of results which highlight the major differences that exist between NHSCR and Census migration information.

2. PREVIOUS RESEARCH AND THE CHARACTERISTICS AND AVAILABILITY OF SOURCES OF MIGRATION DATA

2.1 Previous research

Ogilvy (1980a) has compared one year transition data from the 1971 Census with NHSCR movement data. She compared Census migration, measured from 25/26 April 1970 to 25/26 April 1971, with the available NHSCR data for the closest corresponding time-period which was that associated with moves occurring between April 1st 1971 and March 31st 1972. This NHSCR movement data, on the assumption of an average three-month lag in re-registration, is that assumed to have taken place during the calendar year 1971. Therefore the time-periods associated with the two data sets overlap by only three months (January - March 1971).

In her analysis, which is restricted to gross and net movement between the eight English regions and Wales, Ogilvy points out the obvious difficulties involved in such a comparison of Census migrants versus NHSCR migrations or moves and states that multiple movement is the main reason of the differences between the two measures. NHSCR gross flows were shown to be approximately 20%

higher than those from the Census at this spatial scale although there was a strong correlation (Pearson's correlation coefficient, r=0.997) between the two data sets. The correlation proved to be only slightly weaker when net flows were compared. The transformation of flows into rates again produced a significant positive correlation coefficient. In a subsequent paper, Ogilvy (1980b) summarised the differences by age and sex, highlighting in particular, higher rates of NHSCR movement for children under 5 and teenagers aged 15-19, a differentially higher recording of NHSCR moves by young women as compared to young men and a higher rate of NHSCR movement by persons aged 60 and over.

Thomson (1984) undertook a comparative study based on age and sex-disaggregated 1981 Census data and NHSCR data for flows between metropolitan districts and shire counties in the West Midlands, and flows between these zones and the other standard regions. Thomson showed that there was a generally strong correlation between the two data sets but not for the 15-19 age-group. The possibility of student distortion was emphasised, with the application of a student correction factor producing a more satisfactory relationship for this age-group. The estimated NHSCR : Census ratio for net flows was 1.04 although ratios for individual areas showed wide variation. The gross flow comparison produced ratios which were higher overall but more stable across age-bands. Like Ogilvy, Thomson states that the presence of multiple moves is the major reason for differences between Census and NHSCR figures but argues that whilst significant variation is apparent at the small area level, the NHSCR is a reasonable guide to migration at more aggregate spatial scales.

More recently, Devis and Mills (1986) have published a detailed comparison of NHSCR and Census information which analyses some of the differences that exist between the two alternative sources of migration data, and which illustrates the effect of adjusting for these differences. The comparison is based on rates of movement between FPCAs in England and Wales. The respective time-periods of observation for the NHSCR and the Census are more closely matched than those used by Ogilvy (1980a). NHSCR data, when lagged by three months, are associated with the twelve months ending March 31st 1981, whilst the Census data refer to the year prior to April 5th 1981, the date of the 1981 Census.

The total NHSCR in- and out-transfers were shown to exceed the total Census in- and out-flows by approximately 28%. When NHSCR moves by those with unstated age and sex were omitted, together with moves made by persons aged under one, the discrepancy fell to 24%. These crude discrepancies were greater for females than males and greater for moves between regions than for moves between FPCA's within regions. The difference was most significant for women aged 15-19 and boys aged 10-14.

Devis and Mills emphasize that there is no simple reason for these crude differences such as one source including multiple and return moves and the other not doing so, and they outline the main factors which require consideration when attempting to match the alternative data sets. The important types of move are those involving students, Armed Forces personnel and their dependants, non-survivors, prisoners and long-term psychiatric patients and those who move more than once. The quality of Census data is also an important factor. The effect of adjusting for these discrepancies is to reduce the total difference between the data sets to 3%.

The main aim of the analysis reported in Occasional Paper 35 (Devis and Mills, 1986) was to decompose NHSCR re-registrations and Census migrant figures into move or migrant types and to attempt a comparison of the lowest common denominator. Table 1 sets out the decompositions estimated by Devis and Mills although arranged differently from their Tables on pages 1 and 54. Figure 1 illustrates these components in diagram form. Essentially, component A.1 of the NHSCR re-registrations (1,301,306) is compared with component B.1 of the Census migrants plus component B.4, an estimate of missed migrants, which is 1,130,575 plus 172,000 = 1,302,575. The estimated numbers are thus in very close agreement.

The remaining components of both data sets can be divided into three groups:

(i) those that involve conceptual differences between the two sources;
(ii) those that involve operational measurement problems in the two sources; and
(iii) those that involve differences in the populations at risk captured in the two sources.

Subsequent sections examine the importance and effect of these groups upon the comparison of the alternative datasets.

Much of the analysis by Devis and Mills (1986) focusses on whether, at the FPCA scale, the net migration estimates provided by the two sources are in agreement or not. This concern with net migration derives from its use in the final stage of current subnational population projections. The argument is that if the

Table 1. The decomposition of NHSCR re-registrations and Census migrants, 1980-81, estimated by Devis and Mills (1986): migration between FPCAs in England and Wales.

	A. Decomposition of NHSCR re-registrations between FPCAs		Source in Devis & Mills (1986)
A.1	First moves of non-student survivors who are one year of age or more and whose sex and age are stated.	1301306	Residual
A.2	Moves by migrants who die	4662	Table 3.8
A.3	Moves by students	100100	Table 3.2
A.4	Second and further moves (Multiple and return moves)	101672	Table 1.1 & p.16
A.5	Moves by persons under one year of age	17600	Table 2.2
A.6	Moves of persons with sex not-stated	25490	Table 2.2
A.7	Moves of persons with age not-stated	3300	Table 2.2
	Total NHSCR re-registrations	1554130	
	Additional components not measured directly		
A.8	Armed Forces moves between FPCAs	78600	Table 3.5
A.9	Moves between FPCAs by inmates of prisons or psychiatric establishments	7440	p.18
A.10	Sampling Error	+ or - 7330	Appendix C
A.11	Moves between FPCAs not resulting in a re-registration	unknown	
	Possible total NHSCR re-registrations	1647500 to 1632840	
	B. Decomposition of Census migrants between FPCAs		
B.1	Civilian, non-institutional surviving migrants, aged one or more	1130575	Residual
B.2	Armed Force migrants	78600	Table 3.5
B.3	Prisoners and psychiatric patients	7440	p.18
	Total Census migrants	1216615	
	Additional components not measured directly		
B.4	Migrants missed by the Census (origin not-stated, under-enumeration or mis-reporting as estimated by the Post Enumeration Survey)	172000	pp. 18-19
		1388615	

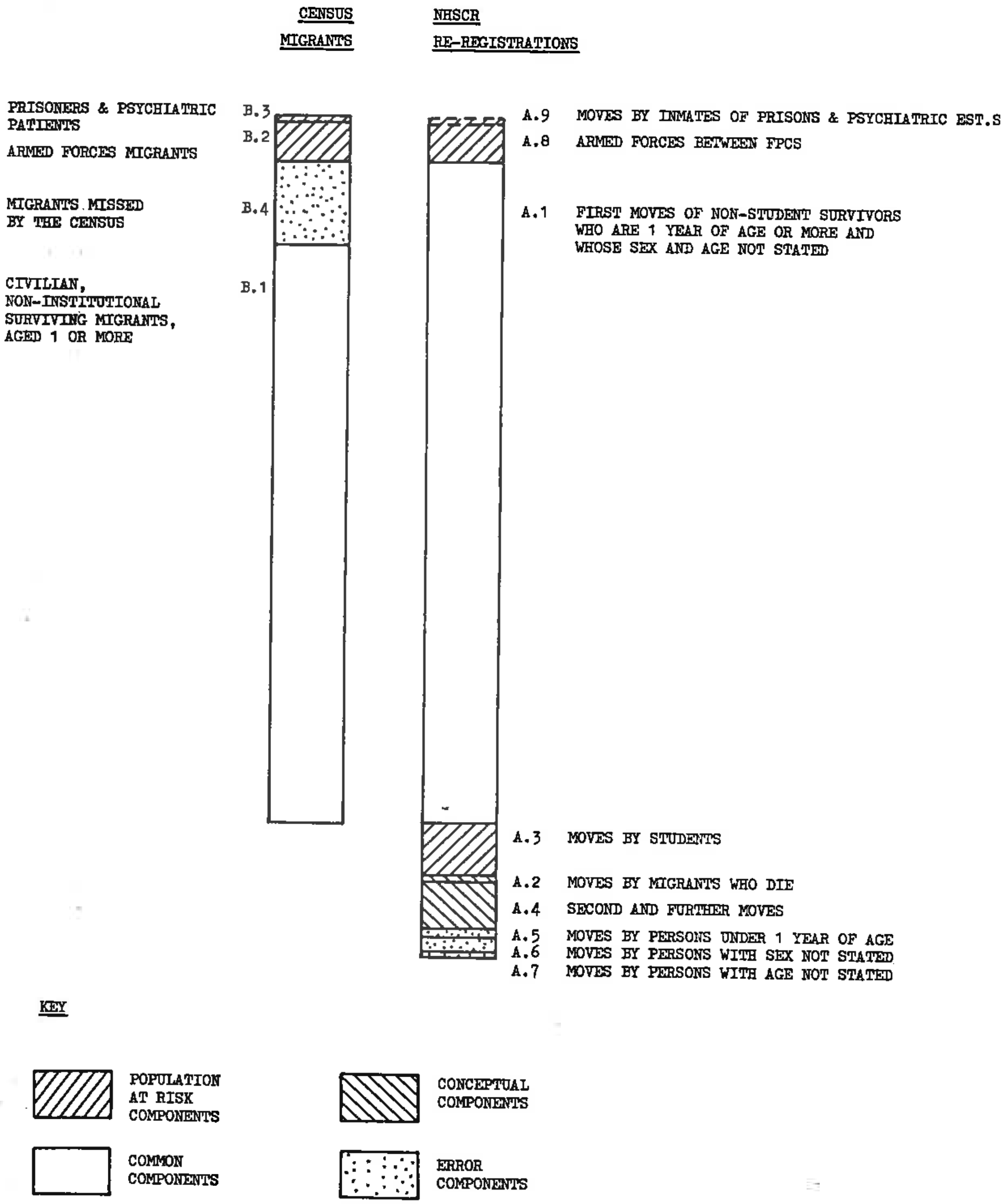


Figure 1: The components of NHSCR re-registrations and Census migrants, 1980-81, estimated by Devis and Mills (1986)

two sources are shown to be in agreement, then we can be confident in using NHSCR re-registration data for inter-censal years to provide data for trending the migration inputs to the subnational projections. However, in its earlier stages, the subnational population projection model utilizes gross migration by age and sex into and out of local authority areas, and gross migration for broad age groups between local authority areas (OPCS, 1984). Alternative projection models which might be explored also use gross migration stream data. Hence it is important to explore the goodness of fit between the NHSCR re-registrations data set and the Census migrant data set by looking at the matrix of inter FPCA flows at various spatial scales.

The comparison reported here takes a slightly different view of the components that make up the two types of migration measure. It aims to provide the justification for the implementation of a multiregional population projection based on "movement" data, and so is concerned to use, not downwardly adjusted NHSCR re-registrations A.1 but upwardly adjusted NHSCR re-registrations - components A.1 to A.7 (total measured re-registrations) plus components not measured A.8, A.9 and A.11, adding to NHSCR re-registrations estimates of Armed Forces and institutional migration wherever possible. A study of the discrepancies between Census migration and NHSCR migration should help in this respect.

To summarize this short review, it is clear that the analysis of Devis and Mills (1986) is crucial to an understanding of the differences between NHSCR re-registration and Census migrant data at FPCA level. Devis and Mills emphasize the variety of reasons why differences occur between the two data sources, and adjust each data set successively in order to reduce the differences between them. However, they do conclude that

> "... although care should be taken in each area with the treatment of various sub-populations, NHSCR data can be an effective tool for the annual measurement of net population changes through migration."
> (p. 28)

The conceptual, operational measurement and population at risk differences between the two sets of data are now examined in more detail.

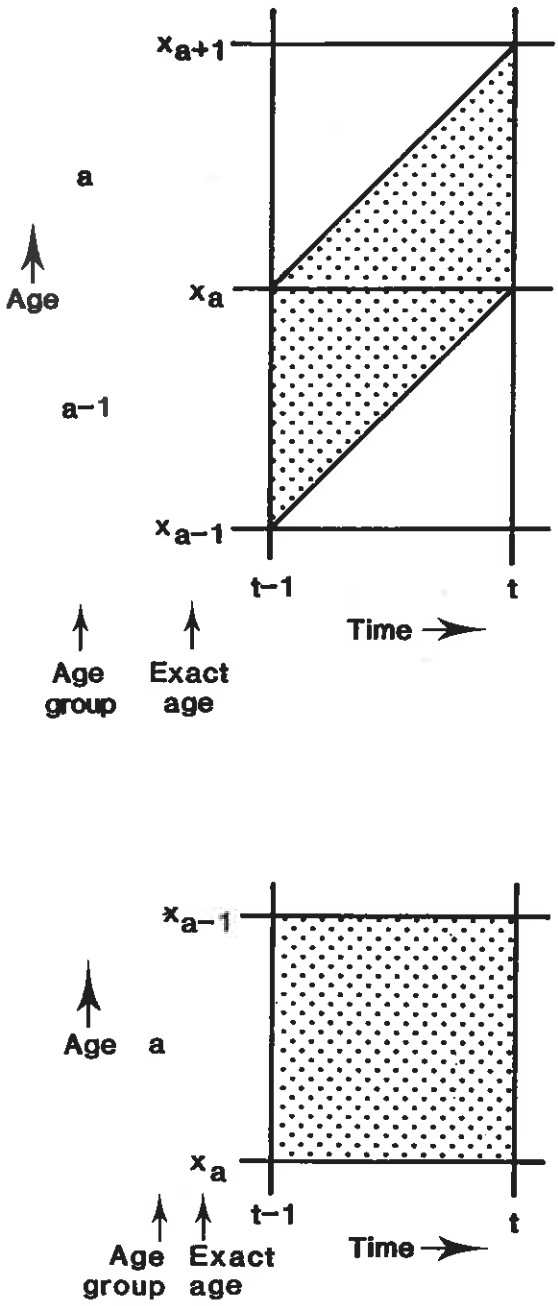
2.2. Conceptual differences between the data sources

Conceptual differences that exist between the alternative types of migration data, although well documented, (Courgeau 1980, Ledent 1980, Rees 1984b) have been somewhat neglected within previous comparative analyses. Within this category fall components A.2, moves by migrants who die and component A.4, second and further moves for NHSCR re-registrations.

Rogers and Castro (1986) note that,

> 'Most information regarding migration is obtained from population censuses or registers that report migration data for a given time interval, in terms of counts of migrants or of moves respectively.'
> (p157)

A person undergoing a change of residence will be classified in the NHSCR by age at the time of move. If the period of observation of events is limited to one year then all moves by persons of a given age will be counted in the age-time space illustrated in Figure 2a - the period age-time plan of observation. A move recorded in this way



b. The age-time plan of observation for transition data

a. The age-time plan of observation for movement data

Figure 2: Age-time plans of observation for movement and transition data

is defined as,

> 'an event in which only the immediately anterior state and the immediately posterior state are known, not the states of the mover at the beginning or end of the time interval.'
>
> (Rees,1986 p101)

The NHSCR will record all such moves taking place within a given time-period and provides, therefore, a count of migration events rather than a count of the number of persons migrating.

Alternatively, the population Census provides a retrospective measure of migration. The age of a migrant is measured at the end of the period of observation and migrant transitions are counted for cohorts of the population who are aged between X_{a-1} to X_a at time t-1 and aged between X_a and X_{a+1} at time t, assuming a one year period is observed. The transitions are measured as a period cohort (Figure 2b) and can be defined as,

> 'classifications of the populations by initial and final states in a time interval; the intermediate states through which a person may have passed are unknown.'
>
> (Rees,1986 p101)

The 1981 Census records migrants as persons whose usual residence on census date was different from that one year or five years previously.

The Census and the NHSCR, therefore, employ contrasting methods of migration classification which will invariably produce contrasting levels of mobility if compared over a similar time-period. For example, an individual undertaking several changes of address within a given period of observation will have each move recorded in a population register, provided that a re-registration is made at each new destination, but will only be classed as a migrant in the Census if the persons address at the beginning of the period is different to that at the time of enumeration. If an individual returns to his initial address before the end of the period of observation, after previous changes of address, he/she will be recorded in the Census as a non-migrant as the address at the beginning and end of the period are the same. In this instance a population register would record a return movement. Furthermore, a person who changes residence during the period in question yet dies before the end of the period will constitute one move in a population register but will be ignored by the retrospective census. The result of such discrepancies is that,

> 'since at least some migrants, by census definition, will have been involved, by registration deinition, in more than one migratory event, counts from registers should normally exceed those from censuses.'
> (Rogers and Castro,1986 p158)

2.3. Operational measurement difficulties

Within this second category of differences one can identify component A.6, moves of persons with age not-stated, component A.7, moves of persons with sex not-stated, component A.10, sampling error and component A.11, moves not resulting in a re-registration for the NHSCR data (see Figure 1). For Census migration data the corresponding component is B.4, migrants with origin not-stated, migrants with mis-reported migrations and migrants missed through under-enumeration.

The accuracy of information collected by a national survey such as the Census has been measured through a Post Enumeration Survey (PES)

(Britton and Birch,1981) which was carried out immediately after the Census. Devis and Mills (1986) estimate from the PES that the Census failed to record approximately 172,000 migrants because of either non-statement of origin or incorrect completion of forms and mis-representation of those usually resident on Census night. Migrants in the 20-29 age range were estimated as being those most affected by the deficiency of recording suggested by the PES. The extent of this unrecorded Census migration provides further discrepancy between data sources, although it is possible to assign the total origin not-stated migrants on the basis of known flows.

The quality of the NHSCR can be questioned on several counts. First, the NHSCR does not record migrants who do not register with a doctor. Certain groups, especially young male adults, may not register with a new FPC until they require treatment and may even neglect registration totally. Devis (1984) emphasises the fact that household surveys have shown that over a year, 28% of the population never consult a family doctor. Young adults are the most mobile members of the population so their failure to re-register upon moving to a new FPCA may falsify the true level of mobility. Other sub-groups, however, such as the elderly or mothers with young children, are likely to re-register as soon as a move is made - these groups being the most likely to require frequent medical treatment. The speed of registration varies, therefore, between age-groups and although it is impossible to measure accurately, the general assumption has been made that the average time-lag between a move and its accompanying re-registration is three months. The implication is that moves recorded in a certain period refer to moves made, on average, three months earlier. Devis and Mills (1986) explored the use of alternative lags in their comparison of NHSCR and Census data but found little variation in the results that were obtained. The three month lag has, therefore, been universally accepted.

Finally it is important to recognise that movement data from the NHSCR has been obtained by OPCS on a 10% sample basis, whereas the 1981 Census count of migration was 100%. The possibility of sampling error must be recognised when attempting to explain variations in the ratio of NHSCR to Census flows.

2.4. Differences between populations at risk of capture

This third category comprises NHSCR component A.3, moves by students, component A.5, moves by persons under one year of age, component A.8, moves between FPCAs by Armed Forces personnel and their dependants, A.9, moves between FPCAs by inmates of prisons and psychiatric establishments and the corresponding components B.2, Armed Forces migrants, and B.3, prisoner and psychiatric migrants for Census data.

Certain sub-groups of the population are handled differently in each data source. The movement of students, for example, tends to be ommitted from the 1981 Census because persons completing Census forms were instructed in 1981 to include,

> 'any persons who usually live with your household but who are absent on census night. For example, on holiday, in hospital, at school or college.'
> (OPCS, 1981)

Students tend, therefore, to be recorded by the Census as living at the parental home and all moves to places of education during the previous year are excluded from subsequent tabulations. In contrast the NHSCR will record student moves to places of education if re-registration takes place. The extent and timing of re-registration in a new FPCA varies between educational establishments with some institutions having compulsory or block registrations with a GP upon arrival and others leaving the timing of registration to the individual.

The recording of moves made by Armed Forces personnel and their dependents provides another form of discrepancy between the NHSCR and the Census. Such moves are included in the Census yet excluded from the NHSCR. Table 2 taken from Devis and Mills (1986) illustrates how each source treats Armed Forces movement. The Census records recruitments and internal postings but discharges are recorded as civilian migrants provided that address at enumeration was different from one year previously. The NHSCR, however, records only recruitment and discharge moves by origin and destination FPCA respectively. No moves within the Armed Forces are captured by the NHSCR as personnel are recorded, on enlistment, as being members of the Forces and from then until discharge are excluded from any movement tabulations. The destination FPCA on recruitment and the origin FPCA at time of discharge are not distinguishable. These FPCAs are tabulated as 'Armed Forces'. Recruitment and discharge totals therefore include all moves whether by personnel or dependent. Published tabulations of NHSCR data (1976-83) exclude movement by Armed Forces personnel altogether.

Moves by prisoners and long-term psychiatric patients are also included in the Census but excluded from the NHSCR. The 1981 Census specified the usual residence of an inmate as the institution

Table 2. Classification of type of moves between FPCAs by Armed Forces personnel and their dependants

Type of move	Identified in Census	Identified in NHSCR	Recorded in Census	Recorded in NHSCR	Identification Census orig	Identification Census dest	Identification NHSCR orig	Identification NHSCR dest
Recruitment	Yes	Yes	AF mign.	Move from FPCA to AF	FPCA	FPCA	FPCA	AF
Posting	Yes	No	AF mign.	No record	FPCA	FPCA	-	-
Discharge	No	Yes	Civil. mign.	Move from AF to FPCA	FPCA	FPCA	AF	FPCA

Source: Devis and Mills (1986 Table 3.6 p16) concerned if the person had been in the institution for more than six months, or as the home address if less than six months. Again a person is classed as a migrant if address on Census night and one year previously differed. The NHSCR excludes all moves to such institutions and all moves between them but will include patient/prisoner discharges on the condition that the person re-registers in a different FPCA to that of the institution.

Another important difference between the NHSCR and the Census involves the recording of transfers made by infants - those aged less than one at the end of the period. The NHSCR, being a continuous register and observing events in a period age-time plan (Figure 2a), records all registered moves regardless of age. The Census, on the other hand, recording information for cohorts defined by end-of-period ages, will omit all migrants aged less than one at the time of enumeration. Alternative procedures to adjust for infant flows are outlined in a subsequent section. This section of the paper finishes with a short summary of the Census and NHSCR data files that have been assembled.

2.5 Data availability

The Census transition data used in the comparative analysis has been obtained from the 1981 Census district by district migration matrix supplied by OPCS on magnetic tapes. The coded records of information on each tape were processed into the required migration data at the FPCA level disaggregated by age and sex. The three arrays produced for comparison were an inter-FPCA array of flows and gross out and in-flows disaggregated by five-year age-groups and sex.

Movement data was made available by OPCS in two stages, thus necessitating a two-stage comparison with data from the Census. Firstly, NHSCR migration information for the period April 1 1980 to March 31 1981 was provided in the form of computer summaries of Primary Unit Data in files of a similar format to those of the Census data. No Armed Forces moves or origin not-stated moves were recorded within this dataset. Secondly, PUD itself was provided and used used to construct the necessary datasets for comparison, thus allowing a more accurate conversion and estimation routines to be adopted. The following sections introduce the alternative assignment and adjustment techniques utilised in the two stages of the comparative analysis with results from both illustrated in Section 4.

3. ALIGNMENT, ADJUSTMENT AND ASSIGNMENT PROCEDURES ADOPTED

3.1 Age-time plan adjustment of NHSCR data

In both stages of the comparative analyses two important adjustments were made to the NHSCR data. Firstly the exclusion of a proportion of infant moves from the inter-FPCA array, and secondly the conversion of the age-time plan of observation for NHSCR data to be consistent with that of Census data. Both adjustments are outlined in this section since they are closely linked in the second stage of the comparison.

An important discrepancy between the two data sets is the inclusion of moves made by infants in the NHSCR data. For the initial comparison (Stage 1) utilising computer summary data the NHSCR inter-zonal flows are reduced in volume by a constant which corresponded to the proportion of infant moves within the system as a whole. Figures from Devis and Mills (1986) suggest a constant, c, defined as,

$$c = \frac{\text{Total moves by persons aged under one}}{\text{Total moves by persons of all ages}} = 0.011 \qquad (1)$$

Section 2 described some of the conceptual differences that exist between Census and NHSCR data. To make the age disaggregated data sets more comparable in Stage 1 it is necessary to convert NHSCR movement data for five-year age groups to Census five year period-cohort data, (or vice versa). The diagrams in Figure 3 illustrate how estimation techniques are adopted to convert NHSCR data to a cohort basis consistent with Census flows, for the first, intermediate and final age-groups.

(a) The first age group

The NHSCR records moves made by all persons in age groups 0 to 4 years during the one year period (Figure 3a). The Census records transitions made by persons in the period-cohort defined by the 1-4 age-group at the end of the period. The census therefore does not include migrants aged less than 1 year. The NHSCR inmove data for the first age group can be estimated, for both sexes, as:

$$M^{1}_{*j}(c) = 0.8\, M^{1}_{*j}(n) \qquad (2)$$

where,

M = NHSCR re-registrations or moves

$M^{1}_{*j}(c)$ = total number of NHSCR moves into zone j recorded in the cohort defined by the end-of-period age-group 1

$M^{1}_{*j}(n)$ = total number of NHSCR moves into zone j recorded in period age group 1

and 0.8 is the proportion of the NHSCR moves in the first age-group to be included.

(b) The final age-group

For the purposes of comparison, the 16th and final age-group defined as those persons aged 75 or over. The penultimate Census period cohort contains all transfers made by those in the 70-74 age-group at the end of the year (Figure 3c). So, to match the movement data in this way, a proportion of the penultimate NHSCR age-group must be combined with the final NHSCR age- group, using the following equation for total in-movement to destination j:

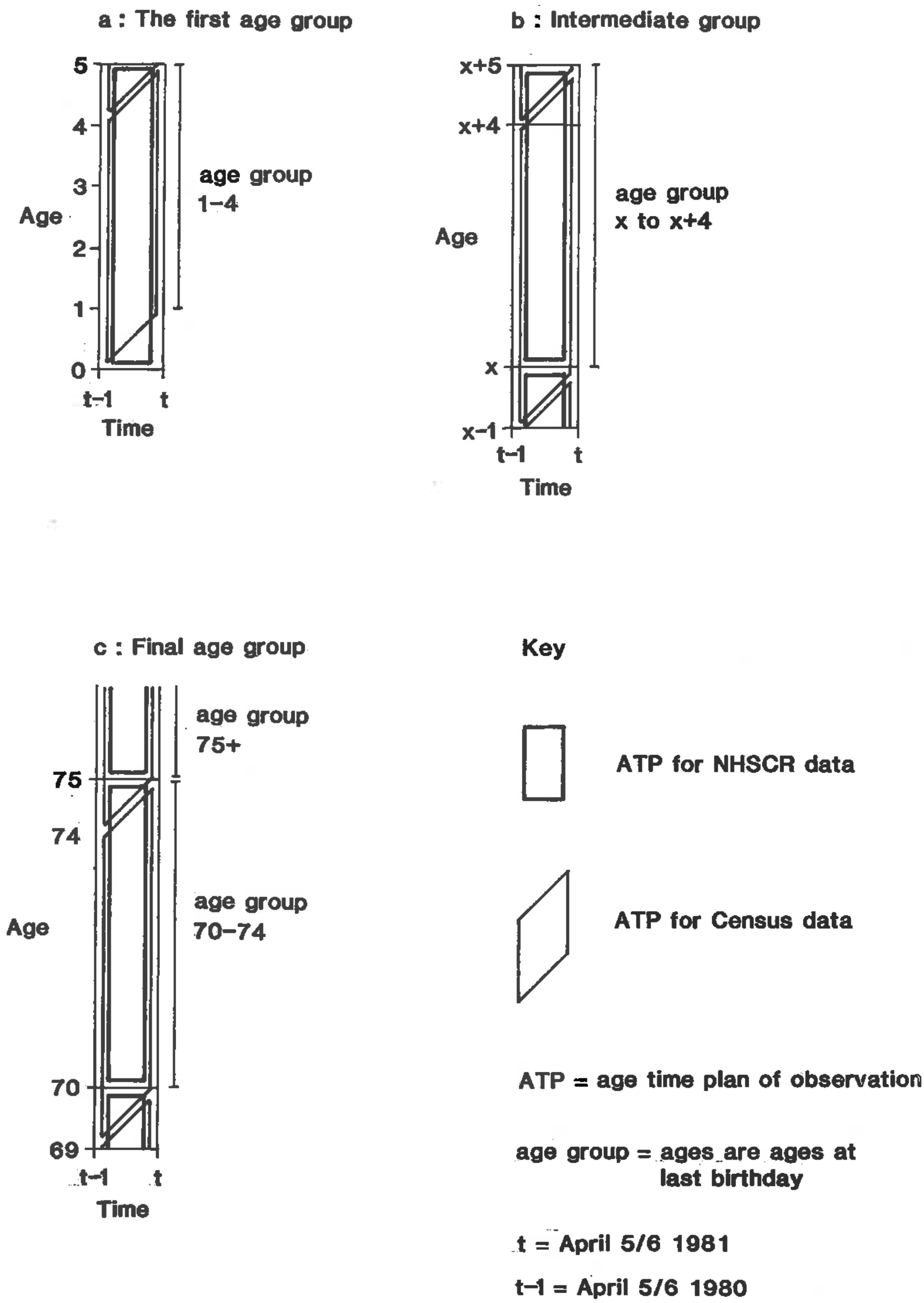


Figure 3: Age-time adjustments required to convert movement to transition data in first, intermediate and final age groups

$$M^{16}_{*j}(c) = 0.1\, M^{15}_{*j}(n) + M^{16}_{*j}(n) \qquad (3)$$

where 0.1 is the proportion of the penultimate age group total to be included. The 15th age group is 70-74 years of age and the 16th refers to 75 or more years of age.

(c) <u>The intermediate age-groups</u>

The inmove totals for intermediate age groups (Figure 3b) can be adjusted as follows:

$$M^{a}_{*j}(c) = 0.1\, M^{a-1}_{*j}(n) + 0.9\, M^{a}_{*j}(n) \quad \text{for } 1<a<16 \qquad (4)$$

where a-1 and a are consecutive age groups used in the NHSCR re-registration dataset.

While this ATP adjustment method proved adequate for the preliminary (Stage 1) comparison of gross out and in-migration totals, the availability of PUD allows a more accurate conversion routine to be utilised and, at the same time, enables the estimation of the relevant proportion of infant moves to be excluded from the aggregate NHSCR origin-destination array.

Figure 4 illustrates the age-time plans of observation for Census-period cohorts in relation to annual cohorts. Census data relates to cohorts defined by end-of-period age-groups. The aim is to transform the NHSCR movement data so as to be consistent with the Census transition form of age-time plan. The coded PUD gives information for each move on the basis of calendar year of birth,

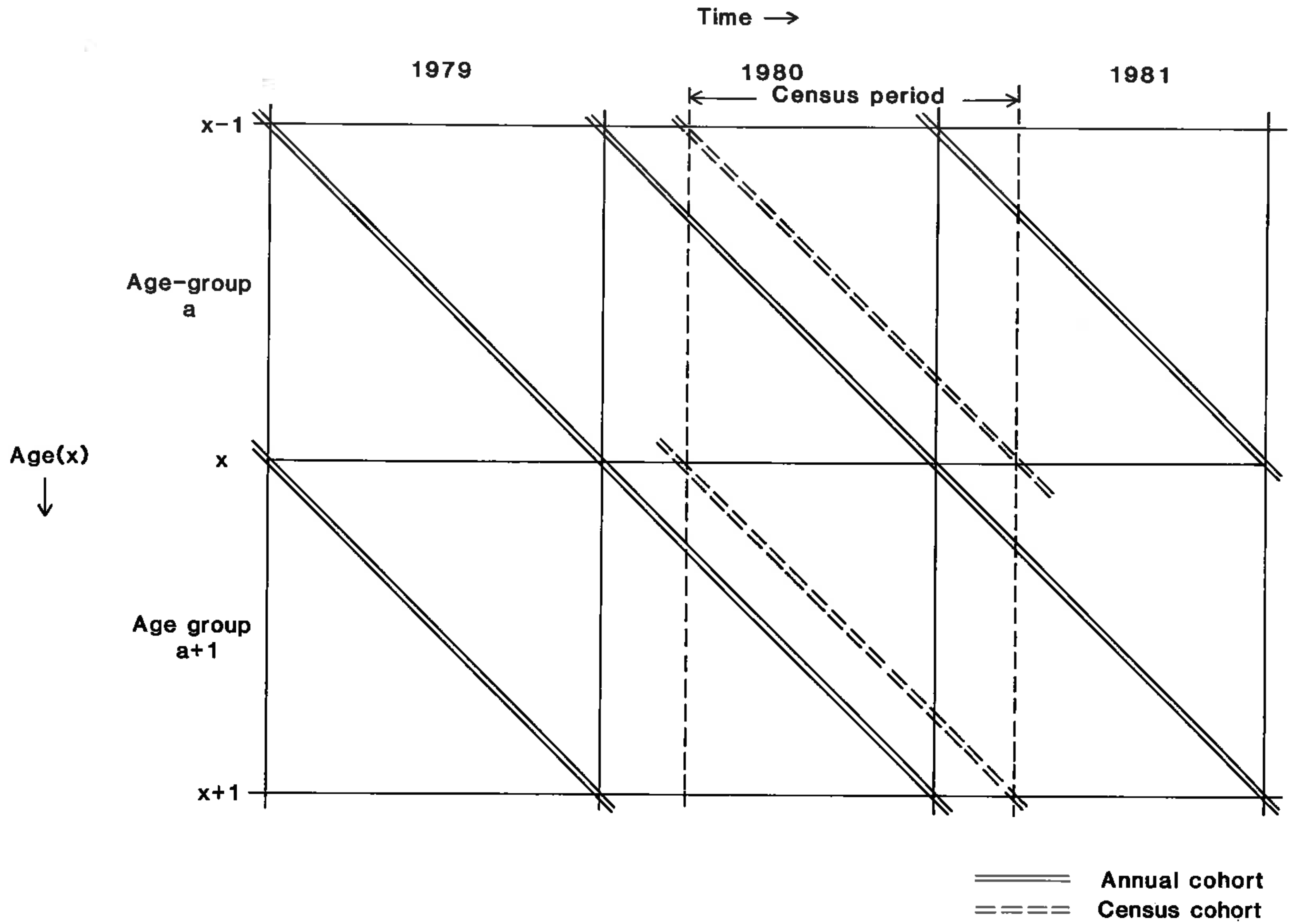
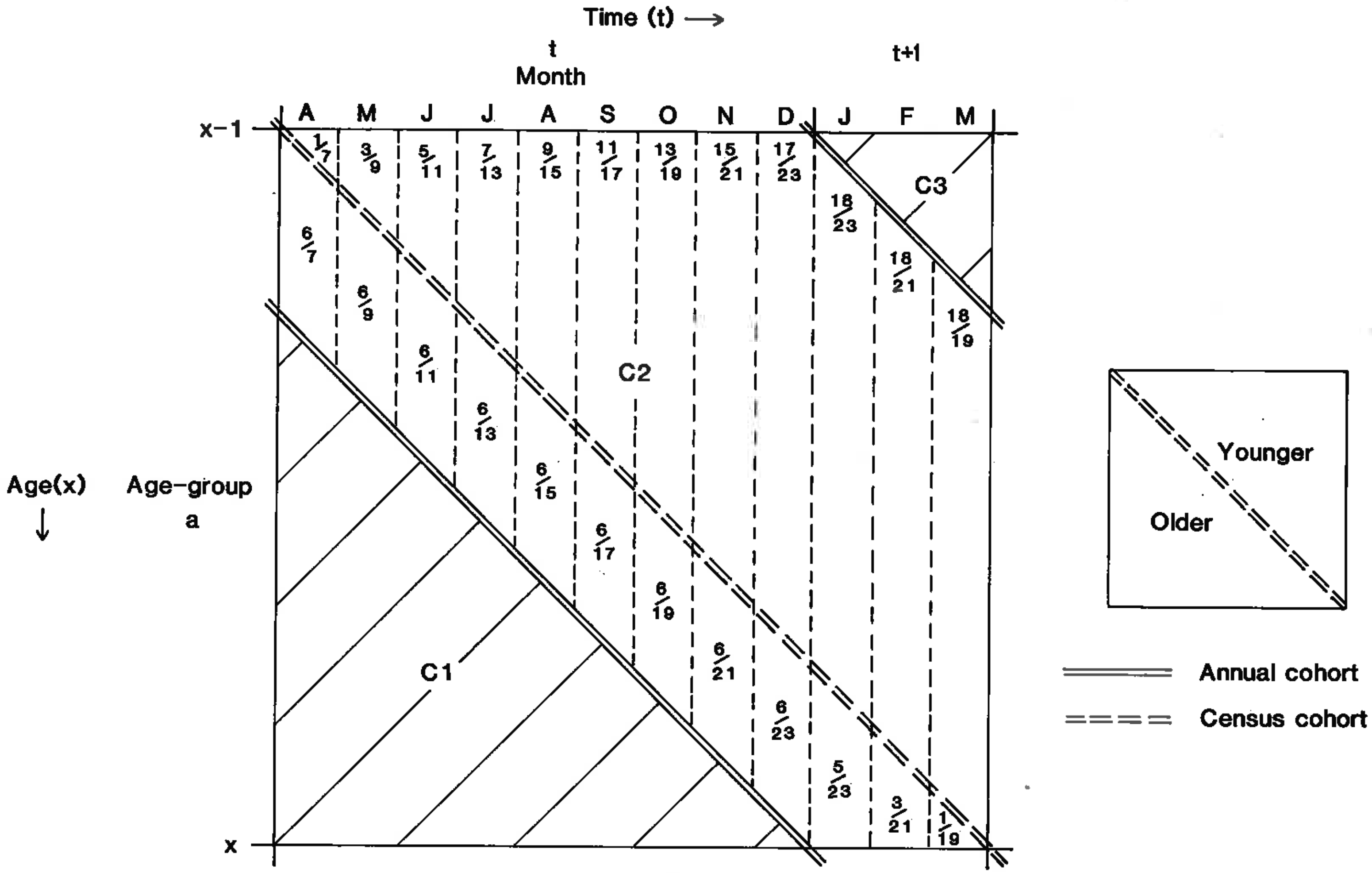


Figure 4: Age-time plan of observation for Census cohorts in relation to annual cohorts

thus allowing the NHSCR data to be recorded in annual cohorts. The PUD relates to the period April 1st 1980 to March 31st 1981. The coincidence between annual cohorts and the required cohorts for the NHSCR data are illustrated in Figure 4. For each age-group it is necessary, therefore to convert the NHSCR data, recorded in annual cohorts, to Census-period cohorts. Figure 5 illustrates in greater detail the nature of the estimation and conversion routine involved.

Firstly, in order to compare NHSCR with Census data, one needs to estimate the number of moves made by those aged less than one at the end of March 1981 and subtract these from the aggregate NHSCR counts. If the value of t (year) in Figure 5 is taken to be 1980 then those moves to be excluded from the NHSCR inter-FPCA flows are those recorded in the 'younger' section of the Census-period age-time plan of observation (those aged less than one at enumeration). Year of birth is coded in the PUD as the number of years since 1900 (i.e. if year of birth is 1934 then the code assigned is 034). The age at move code ranges from zero, for moves made by those aged under one year of age, to 099 for those aged 99 at time of move. Finally a 'month-of-move' code is contained in the PUD which relates the time of move to the number of moves since January 1970. A simple recoding procedure converts the month-of-move code to one relating to the twelve months of the period of interest (April 1980 = 1 and March 1981 = 12).

Figure 5 illustrates how the ATP for the first age-group can be broken down into a series of sections dependent upon year of birth, age at time of move and month of move. The 'older' section relates to the upper half of the first single-year Census-period cohort (ie.



Probabilities : e.g. $\frac{1}{7}$ = probability that a move is recorded in the younger cohort given that the move was made in April and given the move was recorded in C2

Figure 5: Age-time plan of observation for the Census period

those aged one at the end of the period). The superimposition of the NHSCR annual cohorts onto the Census-period cohort divides the ATP in to three sections - C1, C2 and C3. So, for example, if a move is made by an infant (aged 0) whose year of birth was 1979 then that move must be recorded in section C1. If, however a move is made by an infant whose year of birth was 1981 then that move must be included in section C3. Any other moves by infants (ie. those born in 1980) must be recorded within the boundaries of section C2. All moves made by infants born in 1979 must be in the 'older' section of the age-group. All moves made by infants born in 1981 must be in the younger section of the age-group. It is necessary, therefore, to proportionally assign moves recorded in the C2 section to either the 'younger' or 'older' half. Knowledge of the month of move allows a series of conditional probabilities to be computed based on the probability that a move is recorded in the younger or older half of the Census cohort given that the month of move is known and given that the move was recorded in section C2. Each move is assigned to its appropriate section (C1, C2 or C3) and, if in section C2, proportionally divided,on the basis of the probabilities computed, between the younger and older sections of the first age-group. As the corresponding Census data excludes all transfers made by persons aged less than one at the end of the period, those moves recorded in the younger section of the first age-group will be excluded from the NHSCR inter-FPCA array of flows. Those moves recorded in the 'older' section will be included.

As with the first age-group, each subsequent individual age-group can be split into a 'younger' and 'older' section which relate to a younger and older cohort. The annual cohorts superimposed on the Census-period cohorts again divides the ATP into three sections C1, C2 and C3. As each record is processed it is the comparison of the age at time of move with the year of birth which place each individual move into the relevant section. To make the year of birth directly comparable with the age at move code, it is necessary to convert it to a figure representing the difference between 1980 and the year of birth. So, for example, if a persons year of birth was given as 1950 the new coding would be 30 (1980-1950). This effectively converts the year of birth figure to an alternative measure of a persons age at move. Comparison of this 'age' with the age at the time of move places each individual move into either section C1, C2 or C3 of the ATP. For example, assuming that a coded record gives the age at move as 5, the year of birth of a person moving in the relevant period must, therefore, be either 1974, 1975 or 1976. If the year of birth was 1974 the move must be recorded in section C1 of the ATP. If the year of birth was 1976 the move must be recorded in section C3 of the ATP and finally, if the year of birth was 1980, the move must therefore be recorded in section C2.

Moves assigned to the C2 section can again be distributed proportionally between the younger and older halves of the age-group using conditional probabilities based upon month of move. After the complete processing of PUD for the 1980-81 period each single year age-group will contain a younger and older half. It is then possible to aggregate these halves to create single-year cohorts consistent with those of the Census data, and further to five-year cohorts which are used in the comparative analysis. The origin-age-sex and destination-age-sex arrays will therefore contain migration information with an age-time plan of observation that coincides with that of the Census transition data.

3.2 Assignment of not-stated flows

Procedures are required to estimate not-stated flows in both Census and NHSCR data, and two methods have been adopted.

(a) Assignment of origin not-stated flows in the Census data.

The Census data include only one not-stated category - that of origin not-stated. Those Census flows with unknown origin were assigned as follows :

$$C_{ij}(2) = C_{ij}(1) + C_{?j}(C_{ij}(1) / (\sum_j C_{ij}(1)) \qquad (5)$$

where,

C = Census migrant;

$C_{ij}(2)$ = adjusted Census flow between origin i and destination j;

$C_{ij}(1)$ = recorded Census flow between origin i and destination j;

$C_{?j}$ = Census moves to zone j with origin not-stated.

(b) Assignment of all not-stated categories in the NHSCR data

The generation of age-disaggregated information from the NHSCR primary unit data has required the reassignment of age, sex and origin not-stated categories. There is no destination not-stated category. Flows with origin and/or age and/or sex not-stated have been assigned as follows.

$$M^{as}_{i*}(2) = M^{as}_{i*}(1) \tag{6}$$

$$+ M^{as}_{?*} \left(M^{as}_{i*}(1) \Big/ \sum_{i} M^{as}_{i*}(1)\right)$$

$$+ M^{?s}_{i*} \left(M^{as}_{i*}(1) \Big/ \sum_{a} M^{as}_{i*}(1)\right)$$

$$+ M^{a?}_{i*} \left(M^{as}_{i*}(1) \Big/ \sum_{s} M^{as}_{i*}(1)\right)$$

$$+ M^{?s}_{?*} \left(M^{as}_{i*}(1) \Big/ \sum_{ia} M^{as}_{i*}(1)\right)$$

$$+ M^{a?}_{?*} \left(M^{as}_{i*}(1) \Big/ \sum_{is} M^{as}_{i*}(1)\right)$$

$$+ M^{??}_{i*} \left(M^{as}_{i*}(1) \Big/ \sum_{as} M^{as}_{i*}(1)\right)$$

$$+ M^{??}_{?*} \left(M^{as}_{i*}(1) \Big/ \sum_{ias} M^{as}_{i*}(1)\right)$$

where,

$M^{as}_{i*}(2)$ = adjusted NHSCR moves from origin i for age-group a and sex s;

$M^{as}_{i*}(1)$ = recorded NHSCR moves from origin i for age-group a and sex s;

$M^{as}_{?*}$ = NHSCR moves from unknown origin, for age-group a and sex s;

$M^{?s}_{i*}$ = NHSCR moves from origin i for unknown age-group and sex s;

$M^{a?}_{i*}$ = NHSCR moves from origin i for age-group a and unknown sex;

$M^{?s}_{?*}$ = NHSCR moves from unknown origin unknown age-group and sex s;

$M^{a?}_{?*}$ = NHSCR moves from unknown origin for age-group a and for unknown sex;

$M^{??}_{i*}$ = NHSCR moves from origin i for unknown age-group of unknown sex;

$M^{??}_{?*}$ = NHSCR moves from unknown origin for unknown age-group of unknown sex.

3.3 Assignment of Armed Forces recruitments and discharges

The NHSCR primary unit data records moves to and from the Armed Forces (ie. recruitments and discharges) but not moves within the Armed Forces (ie. postings). The PUD codes the Armed Forces as a single origin and as a single destination with no information given on the FPCA to which a person is recruited or from which a person is discharged. The computer summaries of NHSCR information used in the initial comparison contained no movement to or from the Armed Forces.

Since the Census migration data includes recruitments and discharges (together with flows within the Armed Forces) it is necessary to assign the available Armed Forces components within the NHSCR array to allow for a more accurate comparison of the data sources. The PUD for 1980-81 includes Armed Forces personnel and their dependents with the same code so that both are assigned by the same process. The assignment of recruitments and discharges is undertaken using usually resident Armed Forces populations obtained from 1981 Census Economic Activity Volumes (OPCS 1984; GRO 1984) based on a 10% sample. The flows are proportionally assigned to individual FPCAs based upon the relative size of Armed Forces population at the origin or destination. The assumption is made that the level of recruitment and discharge to or from an individual FPCA is directly proportional to the size of Armed Forces population in the origin or destination. Such flows within the NHSCR aggregate inter-zonal array are therefore assigned as follows:

$$M_{ij}(2) = M_{ij}(1) + (M_{iAF} \,.\, P^{AF}_{j} / P^{AF}_{*}) + (M_{AFj} \,.\, P^{AF}_{i} / P^{AF}_{*}) \qquad (7)$$

where,

$M_{ij}(1)$ = original NHSCR flow total;

$M_{ij}(2)$ = flow total with Armed Forces assigned;

M_{iAF} = total number of flows from zone i to Armed Forces;

M_{AFj} = total number of flows from Armed Forces to zone j;

P^{AF}_{i} = usually resident Armed Forces population in zone i;

P^{AF}_{j} = usually resident Armed Forces population in zone j;

P^{AF}_{*} = total Armed Forces population.

Armed Forces recruiments and discharges need to be assigned also to the age and sex-disaggregated gross out and inflow NHSCR totals. The reassignment process becomes a little cruder as no age and sex-disaggregation of the usually resident AF population is available and so flows for a paricular age and sex group are distributed to individual FPCAs on the basis of the total AF population in each zone. So, for example, the total outflow for a particular zone i, age-group a and sex s is equal to the outflow total (excluding all AF flows) plus all those recruitments from zone i in age-group a and sex s, plus a proportion of all those AF discharges in age-group a and sex s. The proportion is equivalent to the number of AF personnel resident in an individual FPCA as a prcentage of the total AF personnel. The assignment procedure for outflows, for example, is as follows.

$$M^{as}_{i*}(2) = M^{as}_{i*}(1) + R^{as}_{iAF} + D^{as}_{AF*} \cdot P^{AF}_{i} / P^{AF}_{*} \qquad (8)$$

where,

$M^{as}_{i*}(1)$ and $M^{as}_{i*}(2)$ = original and adjusted outflow totals for zone i, age-group a and sex s;

R^{as}_{iAF} = recruitments from zone i in age-group a and sex s;

D^{as}_{AF*} = total AF discharge in age a and sex s;

P^{AF}_{i} = usually resident AF population in zone i;

P^{AF}_{*} = total AF population in the system.

Adjustments have been made to both Census and NHSCR data sets using the methods outlined in this section and some of the results of comparing the two types of data are now presented.

4. COMPARING CENSUS AND NHSCR DATA FOR 1980-81

4.1 Overall levels of NHSCR and Census migration

The Figures in Table 3 illustrate the differences and ratios that exist between NHSCR and Census flows at a variety of spatial scales. The top half of Table 3 indicates the results of using NHSCR computer summary data in the comparison (Stage 1) whereas the bottom half shows what happens when using NHSCR data directly from the PUD with the inclusion of Armed Forces recruitments and discharges and not-stated flows (Stage 2). Although the NHSCR:Census ratios are higher in the Stage 2 comparison, a scale effect is evident in both tables. The size of ratio decreases systematically as the average distance of migration declines so that the ratio between flows within EC2 regions (a system of metropolitan counties, their region remainders and other regions in the UK) is the closest to unity. The reassignment of Armed Forces and not-stated flows (Table 3b) has the greatest effect upon the regional level ratio, reflecting the importance of longer-distance moves by Armed Forces personnel and their dependents, and the least effect upon flows between EC2 regions within standard regions. The assignment of the origin not-stated category (50,860 moves) and the Armed Forces recruitments (69409) and discharges (62932) greatly increases the overall inter-FPCA ratio between NHSCR and Census flows from 1.25 to 1.37. These Armed Forces flows have been ignored in previous analyses but have an important effect upon the NHSCR-Census relationship. The Census figures include all Armed Forces recruitments and discharges so the only remaining discrepancy regarding Armed Forces movement is that which

Table 3. NHSCR and Census migration flows and ratios at various spatial scales.

(a) Stage 1 comparisons

Migration flows between FPCAs	Total moves (NHSCR)	Total transitions (Census)	Difference	Ratio
(1) Between standard regions	838501	629915	208586	1.331
(2) Between EC2 regions	1148990	881826	267164	1.303
(3) Between EC2 regions, within standard regions	310489	251911	58570	1.233
(4) Within EC2 regions	499757	442918	56839	1.128
(5) All flows	1648747	1324744	324003	1.245

(b) Stage 2 comparisons

Migration flows between FPCAs	Total moves (NHSCR)	Total transitions (Census)	Difference	Ratio
(1) Between standard regions	967224	629915	337309	1.536
(2) Between EC2 regions	1289451	881826	407625	1.462
(3) Between EC2 regions, within standard regions	322227	251911	70316	1.279
(4) Within EC2 regions	524917	442918	81999	1.185
(5) All flows	1814368	1324744	489624	1.370

Source: unpublished NHSCR and Census data supplied by the Office of Population Census and Surveys.
Note: The EC2 regions are those listed in Table 5

involves internal postings of Armed Forces personnel and their dependents (included in the Census but not in the NHSCR). Assignment of such 'intra-Armed Forces' moves would further increase the NHSCR/Census ratio as would the assignment of moves by prisoners and long-term psychiatric patients, were they available. Their effect upon the overall ratio would be effectively balanced by the allocation of an estimated 100 thousand student moves to the Census total (Devis and Mills, Table 3.2)

The discrepancy between flows involving infants is eliminated in both tables using methods outlined in Section 3 with each source recording only flows for those aged greater than one at the end of the 1980-81 period. Furthermore the allocation of not-stated flows in Table 3b is based on flow proportions so does not enhance or reduce the scale effect that is evident. The major discrepancies that exist at these aggregate levels may therefore be strongly influenced by the relative importance of multiple and return moves.

The greatest NHSCR/Census ratio value is found at the regional level, which indicates that it is longer-distance migration that produces the greatest discrepancy between NHSCR and Census figures and shorter-distance flows that have the greatest consistency. Gordon (1975, 1982) has highlighted the multi-stream nature of migration identifying the predominantly long-distance employment-related flows and predominantly short-distance housing-related flows. Transfers that are related solely to a change of house are likely to be more permanent than transfers related to employment if analysed in aggregate terms and therefore the multiple/return move phenomenon will be of least importance at those spatial scales involving the greater proportions of short-distance flows: intra EC2, intra-regional/inter-EC2 and inter-FPCA. Data from the NHSCR and the Census will be most consistent for those flows which are most unlikely to involve more than one change of residence over the period in question. Employment-related flows are likely to be less permanent and will be the predominant component of long-distance migration. Multiple and return movement will therefore be of greatest importance where employment-related moves predominate ie. longer distance flows between the standard regions of Britain. The scale effect illustrated in Table 3 is consistent with this explanation.

There is therefore a marked difference between ratios for these alternative spatial scales at the aggregate level that can be explained partly by the existence of the multiple/return moves phenomenon. It would be unreasonable to 'regard multiple moves as the major explanation of the differences' (Ogilvy,1979), although it is hypothesized that the return move phenomenon will be more important at those scales which involve predominantly long-distance moves. Disaggregation of the migration data should reveal the variation in the effect of the conceptual, population-at-risk and error components upon individual zones, age-groups and sexes. Discrepancies at a disaggregate level are likely to be explained more readily by one single component such as the presence of a large number of Armed Forces personnel or a large educational establishment with the importance of the conceptual components dependent upon the level of disaggregation adopted.

4.2. Statistical relationships and spatial differences between NHSCR and Census flows

The total NHSCR and Census figures in Table 3b can be disaggregated into zonal outflows and inflows at three alternative spatial scales and usually resident end-of-period populations are used to compute rates each type of migration data. Correlation coefficients and least squares regression parameters indicate the nature of the relationship betwen the sets of migration rates computed using data with Stage 2 adjustments. Figure 6 illustrates the scatterplots produced at the standard region, EC2 region and FPCA levels. The correlation is generally good with the strongest relationship in evidence for inflow rates at the standard region level (0.997) and the weakest for netflows at the FPCA level (0.896). The correlation between netflow rates is, at each scale, the weakest. At the standard region and EC2 region scales, inflow rates exhibit a stronger relationship than outflow rates although this is not the case at the FPCA level. It is difficult to establish a pattern in the regression parameters although intercept values are larger at the more disaggregate spatial scale. Regression coefficients, when the intercept approaches zero, do reflect the overall ratio between the NHSCR and the Census.

Table 4 provides a set of statistics for summarizing the overall relationship between individual inter-zonal flows from the NHSCR and the Census at three spatial levels. The information gain statistic (IGS) has the major drawback of only comparing non-zero values (the number of zero elements being quite substantial in an array

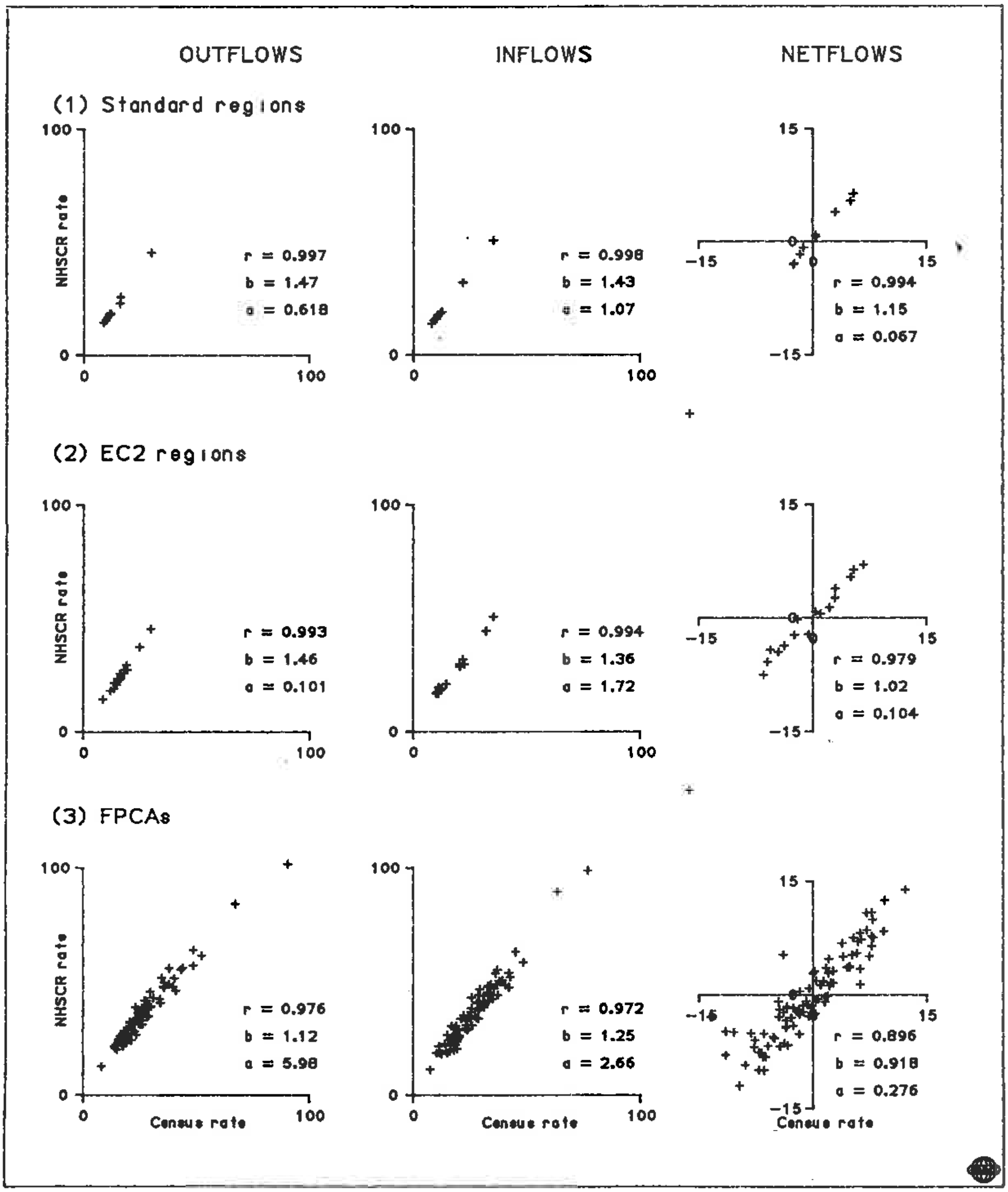


Figure 6: Scatterplots of NHSCR re-registration rates against Census migration rates at three spatial scales

Table 4. Statistics comparing NHSCR and Census inter-zonal flows

Statistic	Regions	Spatial Scale EC2s	FPCAs
Information Gain	0.004	0.009	0.047
Mean Absolute Deviation	34.89	31.37	29.95
Index of Dissimilarity	3.3	5.1	11.4
Correlation Coefficient	0.997	0.997	0.982

containing 97 FPC origins and 95 FPC destinations). It is included here, however, to indicate the increase in information gain as the scale becomes more disaggregate but with a strong relationship between flows at all three levels. The mean absolute deviation (MAD) gives a measure of the 'distance' between the two arrays in terms of absolute numbers. The MAD decreases as the scale becomes finer - from 35% at the standard region level, to 31% at the EC2 level to 30% at the FPCA level. The index of dissimilarity (IOD) measures the degree to which the spatial distibution of the two arrays are dissimilar. The value of the IOD can range from 100, indicating complete dissimilarity, to zero, indicating perfect correspondence. The statistic compares the two arrays by computing the sum of deviations between cell proportions. The value increases as the scale becomes finer indicating that although the MAD statistic computes a relatively small absolute difference at the FPCA scale the IOD value shows that the arrays at this level are least similar. The IOD values are relatively low, however, at all three spatial scales indicating a fair degree of similarity between the inter-zonal flow matrices of the NHSCR and the Census. The final statistic computed is the correlation coefficient (R) ranging from zero to one where one indicates perfect correlation. The R value is high in each case with the correlation decreasing with scale.

Table 5 illustrates the spatial variation in the ratio of NHSCR to Census flows at the EC2 region level. The assignment of Armed Forces flows and not-stated flows in the second stage of the analysis has the effect of increasing each outflow ratio value but generally increases the value of non-metropolitan zone ratios to a greater

Table 5. NHSCR/Census outflow and inflow ratios at the EC2 region scale

EC2 region	Outflow ratio Stage 1 1a	Outflow ratio Stage 2 2a	Inflow ratio Stage 1 1b	Inflow ratio Stage 2 2b
Northern Ireland	1.07	1.54	-	-
Scotland	1.26	1.50	1.24	1.46
Tyne and Wear	1.30	1.41	1.58	1.74
Northern Rem	1.38	1.53	1.29	1.44
South Yorkshire	1.27	1.38	1.46	1.59
West Yorkshire	1.44	1.56	1.55	1.68
Yorks & Humbside	1.25	1.51	1.24	1.46
East Midlands	1.30	1.50	1.26	1.44
East Anglia	1.30	1.60	1.26	1.53
South East Rem	1.30	1.47	1.23	1.38
Greater London	1.30	1.37	1.39	1.50
South West	1.28	1.55	1.23	1.46
West Midlands	1.31	1.39	1.48	1.61
West Midlands Rem	1.24	1.39	1.20	1.32
Greater Manchester	1.35	1.46	1.53	1.66
Merseyside	1.35	1.45	1.51	1.64
North West Rem	1.36	1.48	1.30	1.39
Wales	1.30	1.48	1.34	1.50
All regions	1.30	1.46	1.30	1.46

degree due to the importance of the Armed Forces in these regions. The region of East Anglia has the highest outflow ratio in column 2a with the majority of above average ratios being in non-metropolitan zones. All metropolitan EC2s with the exception of West Yorkshire exhibit below average outflow ratio values.

The dominant feature of the inflow ratios in both columns 1b and 2b is the above average figures for all metropolitan FPCAs. The assignment of Armed Forces flows does little to alter the ranking of the ratio values although areas such as East Anglia, the South West and the East Midlands, with large Armed Forces populations, do experience significant increases in their ratio values relative to other EC2 regions.

The ratio values in columns 2a and 2b can be further disaggregated spatially to highlight variations at the FPCA level (Figure 7). Of the 47 metropolitan FPCAs illustrated in Figure 7a, 38 have below average (1.37) outflow ratios and of the remaining 9 FPCAs 4 are from the county of West Yorkshire (Kirklees, Calderdale, Leeds and Bradford). Of the 40 lowest ranked outflow ratios, 36 are for metropolitan counties. Of the 10 non-metropolitan FPCAs with ratios below the national figure, 5 are in the South East region. The range of the ratio values increases considerably at this scale, from 1.103 (Solihull) to 1.638 (Devon), compared with a range from 1.369 to 1.600 at the EC2 level. There is, therefore, greater variation between outflow ratios at this more disaggregate scale and a clear distinction between values for metropolitan and non-metropolitan zones.

Figure 7b illustrates the variation in the inflow ratios at the

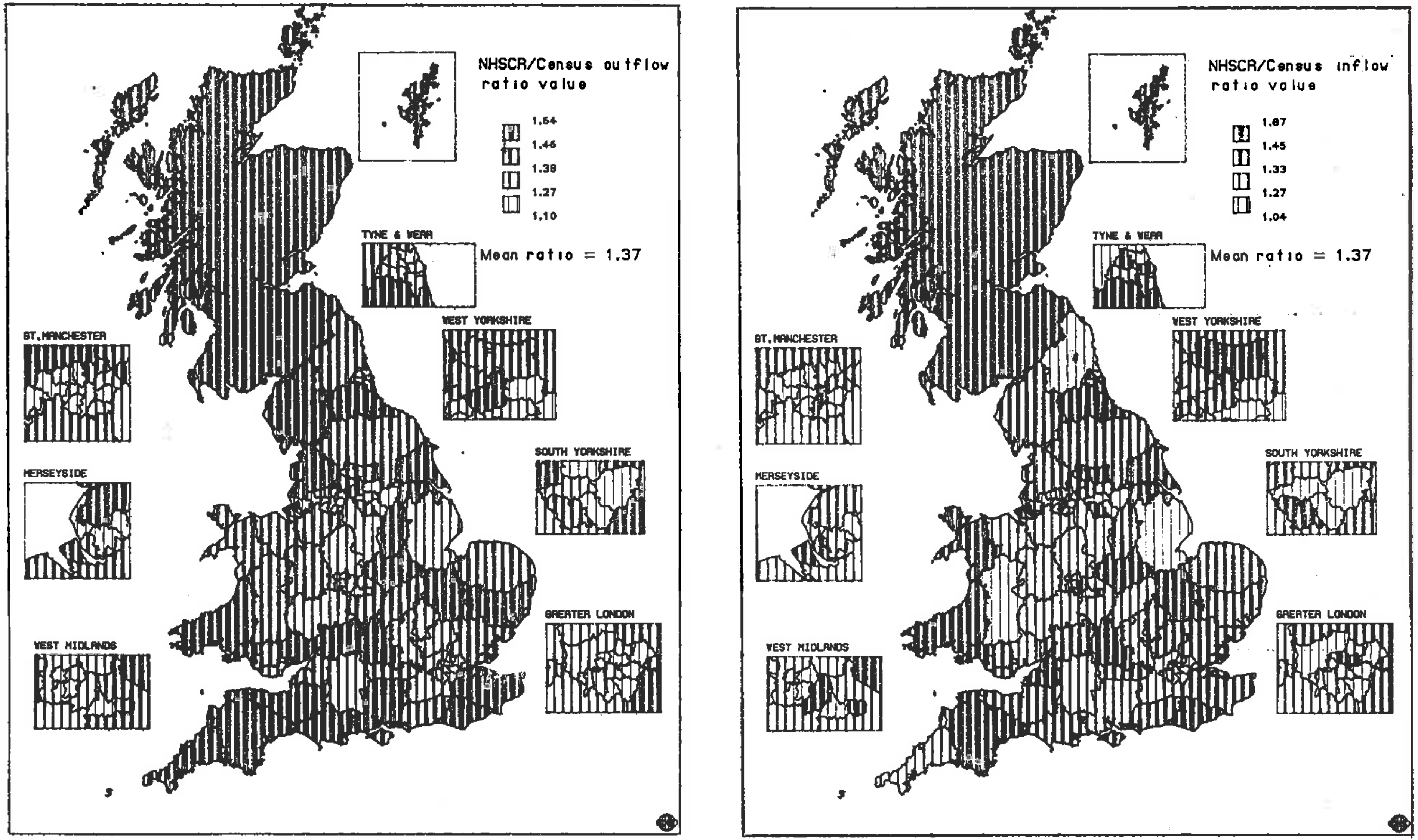


Figure 7: NHSCR: Census migration ratios for FPCAs, outflows and inflows, 1980-1981

FPCA scale The largest ratios are exhibited by inflows to West Glamorgan (1.875), Coventry (1.783), Cleveland (1.757), Sheffield (1.744), Newcastle (1.666) and Leeds (1.619). Other major cities also exhibit high inflow ratios: Manchester (1.590), Birmingham (1.531) and Liverpool (1.462). Even for inflows, however, 32 out of 47 metropolitan inflow ratios are below the average figure. It is the big cities within the metropolitan counties that have the highest inflow ratios with other metropolitan FPCAs in the same county having much lower ratio values, with some exceptions (South Tyneside, Bradford and Calderdale). The 12 individual FPCAs of Greater London exhibit considerable variation in their ratio values with only two having a value above the national figure - the FPCAs of City, Hackney Newham and Tower Hamlets (1.456) and Camden and Islington (1.399). The remaining 10 London FPCAs have inflow values below the mean. An important point to note is the increase in the ratio of outflows and inflows from and to FPCAs with large Armed Forces populations such as Hampshire, Lincolnshire and Wiltshire. All three were shown in the initial comparison to have very low inflow and outflow ratios (Boden, Stillwell and Rees 1987a) but the reassignment of the Armed Forces recruitments and discharges gives a truer picture of their levels of migration and thus their ratio values.

The distinct variation between outflow and inflow ratios for metropolitan and non-metropolitan zones has been emphasised in this section using data with Stage 2 adjustments. Table 6 illustrates the aggregation of inter-FPCA migration to give ratios between inter-metropolitan/non-metropolitan flows. An interesting result at both scales is the ratio value for inter-metropolitan flows. At the

Table 6. Ratios and differences between inter metropolitan / non-metropolitan flows at the EC2 and FPCA level

(a) EC2 level

Type of flow	NHSCR	Census	Difference	Ratio
Metropolitan to metropolitan	75496	43376	32120	1.74
Metropolitan to non-metropolitan	376092	277454	98638	1.36
Non-metropolitan to non-metropolitan	545794	371106	174688	1.47
Non-metropolitan to metropolitan	292069	189890	102179	1.54
Total	1289451	881826	407625	1.46

(a) FPCA level

Type of flow	NHSCR	Census	Difference	Ratio
Metropolitan to metropolitan	337424	285237	52187	1.18
Metropolitan to non-metropolitan	376092	277454	98638	1.36
Non-metropolitan to non-metropolitan	808783	572163	236620	1.41
Non-metropolitan to metropolitan	292069	189890	102179	1.54
Total	1814368	1324744	489624	1.37

EC2 level, the ratio is very high (1.74) relative to ratios for other types of flows ,whereas at the FPCA level the value (1.18) is below other ratios. At the EC2 level, inter-metropolitan flows constitute only 6% and 5% respectively of total NHSCR and Census flows. The nature of the spatial distribution of EC2 metropolitan zones ensures that inter-metropolitan flows at this level are predominantly inter-regional longer distance transfers. Excluded from the inter-metropolitan totals will be important short-distance intra-metropolitan flows for which the Census and NHSCR are most consistent - ie. those flows within EC2 metropolitan regions which are likely to involve a more permanent change of residence and which are unlikely to be accompanied by further moves in the period of observation. Short-distance, predominantly housing-related flows, are those which are likely to correspond most closely between datasets but are also those which make up only a very small proportion of inter-metropolitan flows at the EC2 level. The predominance of longer-distance flows, a major component of which will be unstable employment-related flows therefore gives the large ratio value exhibited by flows between metropolitan zones at this level.

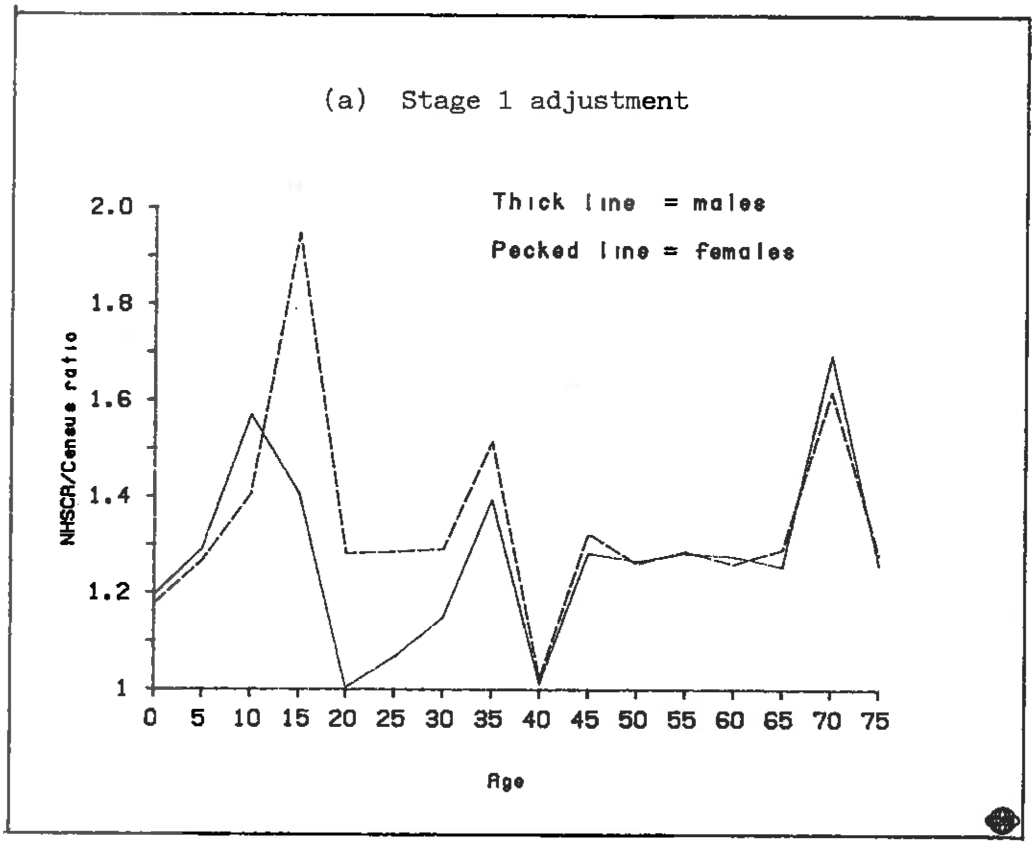
At the FPCA scale, however, the corresponding ratio drops considerably to 1.18, with the proportion of inter-metropolitan flows increasing to 19% and 22% of total NHSCR moves and Census migrants respectively. At this level of spatial disaggregation, short-distance flows between metropolitan FPCAs contained within the larger EC2 metropolitan zones will be recorded. Inter-metropolitan flows at the FPCA level will, therefore, be predominantly short-distance flows.

The magnitude of these flows, due to their more permanent nature (ie. they are less likely to involve multiple or return moves in the single year of observation), will be similar in both the Census and the NHSCR. The ratio between inter-metropolitan flows at this level will therefore be lower.

4.3 Age and sex variation in the NHSCR:Census flow ratio

The characteristic age-specific migration profile that exists at all spatial scales is likely to have a considerable effect upon the ratio values. NHSCR and Census migration data will exhibit similar national and sub-national profiles but at different levels and it is this variation in the level of migration by age and sex group at the FPCA scale that we are interested to observe. The alternative adjustment procedures outlined in previous sections have attempted to align NHSCR data as accurately as possible with the Census data in terms of age-time plan of observation. We are thus comparing cohorts defined by end-of-period age-groups.

Figure 8 indicates the overall variation in the NHSCR/Census ratio by age. The ratio profiles are at a generally higher level in Figure 8b but the most noticeable difference is the increase in the ratio value for males aged 15-19. The initial computer summary of NHSCR data considerably under-counted the number of 15-19 year-old moves involving the recruitment and discharge of Armed Forces personnel and their dependents and therefore surpressed the ratio value in this age category. The increase in the 20-24 age-group for males is also considerable.



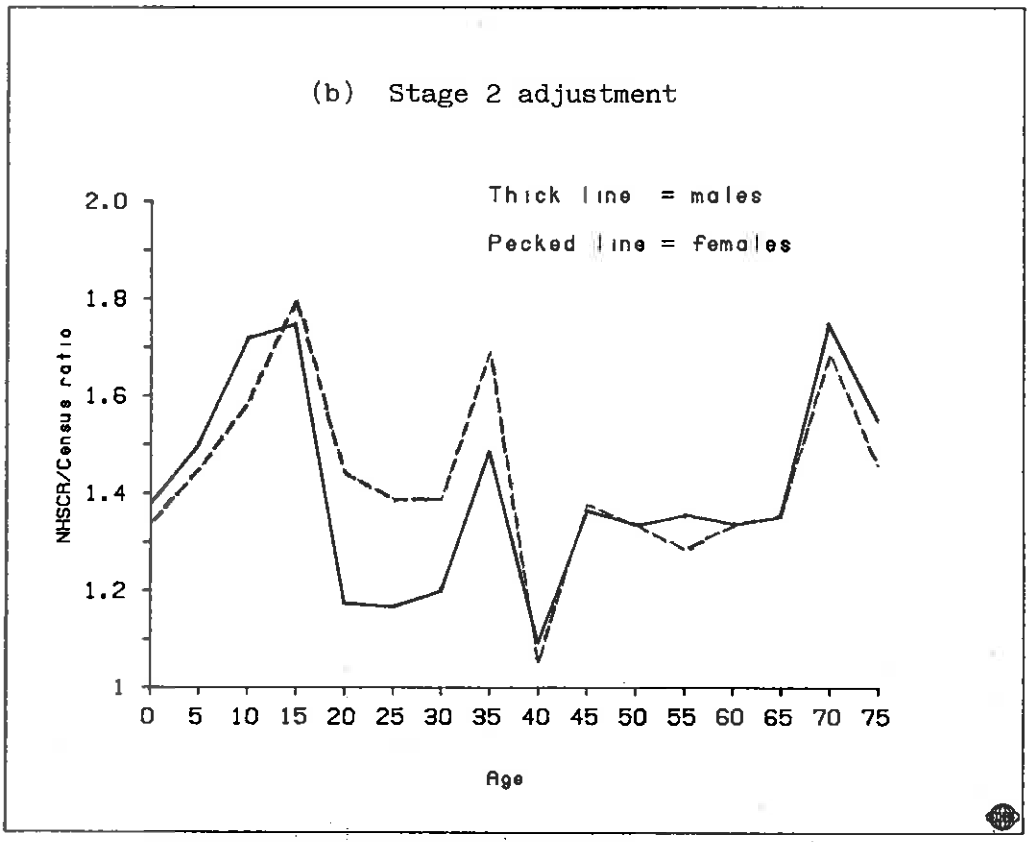


Figure 8: NHSCR/Census ratios by age group, males and females, alternative adjustments

In Figure 8b males and females both have high ratio values at ages 15-19, 35-39 and 70-74. In the male ratio profile the early peak includes the 10-14 age-group. The female ratio value is higher than the male in the 15-19 and 35-39 age-groups but not in the 70-74. The largest ratio value for both males and females is found in the 15-19 age-group. Both schedules reach a low-point between the ages 40-44 with ratios from 45-69 being fairly similar for both sexes. The greatest discrepancy between the male and female ratio exists in the 20-34 age-range with the male ratio being much lower than the female, especially in the 20-24 age-group. The male ratio is higher than the female ratio in the three youngest age-groups.

The age-specific differences may be explained by a number of factors. Non-surviving migrants, ie. those persons making a move but not surviving to the end of the period, will be particularly important in the older age-groups. Devis and Mills (1986, Table 3.8, p17) estimate that approximately 5% of migrants in the 75+ age category do not survive to the end of the period. This component will be particularly important for flows to and (from) 'retirement' areas on the south coast of England.

Student moves will accentuate the discrepancy in the 10-29 age-range with the majority of student flows taking place in the 15-19 age-group. The exclusion of such moves in the Census data is a major component of the large discrepancy between male and female flows in the 15-19 age-group.

Multiple moves and return moves are likely to be of greatest importance for migrants with the highest mobility levels, ie. the 15-29 age range. Devis and Mills (table 3.7, p16) estimate that multiple moves are most prevalent in the 15-19 and 20-24 female age-groups particularly in the former. The relatively low ratios for males in the 20-29 age-range may be due to the non-registration by such migrants after each move. Males aged 20-29 are those least likely to require the services of a doctor and therefore least likely to register with the NHS upon each change of residence.

4.4 Multiple regression models to predict NHSCR inflows

In this final section multiple regression techniques are used to construct linear equations relating NHSCR inflow totals (Stage 2 adjusted) for males and females to a number of independent variables: the corresponding Census flow total, the Armed Forces usually resident population and an estimate of the number of inward and outward student re-registrations. Armed Forces populations are used in the absence of an accurate estimate of inter-FPCA Armed Forces movement and are obtained from 1981 Census Economic Activity volumes based on 10% processing (OPCS, 1984). The estimates of student moves are taken from Devis and Mills (1986, Appendix Table 2). No estimate is available of student re-registrations to and from the FPCAs of Greater London as the non-correspondence of FPCAs and LEAs provides particular problems for estimating student movement. Furthermore, no estimates of student re-registration are available for Scotland. The multiple regression analysis is, therefore, restricted to non-London FPCAs in England and Wales.

Table 7 presents Pearsons coefficients for male and female inflow figures. A strong correlation is evident between NHSCR and Census inflows for both sexes. The correlation between NHSCR flows

Table 7. Correlation matrix for all variables, male and female inflows

	NINM	CINM	AFPOPM	STRGIM
NINM	1.000	0.988	0.679	0.626
CINM	-	1.000	0.718	0.545
AFPOPM	-	-	1.000	0.343
STRGIM	-	-	-	1.000
	NINF	**CINF**	**AFPOPF**	**STRGIF**
NINF	1.000	0.993	0.661	0.436
CINF	-	1.000	0.646	0.370
AFPOPF	-	-	1.000	0.272
STRGIF	-	-	-	1.000

NINM (NINF) = NHSCR male (female) inflow
CINM (CINF) = Census male (female) inflow
AFPOPM (AFPOPF) = Armed Forces male (female) population
STRGIM (STRGIF) = Student male (female) inflow

and the Armed Forces population variable is also relatively strong. Correlations between the student variable and NHSCR female inflows are the weakest of the three. Table 8 indicates the correlation between NHSCR/Census inflow ratios and the Armed Forces and student variables. It can be assumed that any Armed Forces variable should have a negative effect upon the ratio value. This is the case for the ratio between male non-metropolitan inflows and between all male inflows but is not true for other relationships between the NHSCR/Census ratio and the Armed Forces population. A negative coefficient indicates that, due to the fact that Armed Forces movement is excluded from the NHSCR but included in the Census, the ratio will decrease as the size of the Armed Forces population increases. For metropolitan flows, however, the correlation is strongly positive indicating an increase in the ratio value as the size of the Armed Forces population increases. The problem here is the relatively small numbers of Armed Forces personnel present in metropolitan areas (4%) compared to non-metropolitan areas (96%). For the student variables all correlations are positive. The Census records students as living at home and does not register any move to place of education. The NHSCR, however, will record all such moves, assuming the student re-registers with a new FPC. The greater the number of estimated student inward re-registrations, therefore, the greater the NHSCR/Census ratio.

Multiple regression analyses were undertaken to establish linear relationships between NHSCR inflows and the corresponding Census flows, Armed Forces populations and estimated student re-registrations using the 'stepwise' entry method. Only non-London

Table 8. <u>Pearson correlation coefficients between NHSCR/Census ratio and the Armed Forces and student variables, male and female, inflows</u>

Correlation	All FPCAs	Metropolitan FPCAs	Non-metropolitan FPCAs
RATIOINM with :			
(a) AFPOPM	-0.082	0.602	-0.207
(b) STRGIM	0.458	0.675	0.308
RATIOINF with :			
(a) AFPOPF	0.160	0.359	0.028
(b) STRGIF	0.570	0.781	0.275

RATIOINM = NHSCR male inflow total/Census male inflow total
RATIOINF = NHSCR female inflow total/Census female inflow total

Table 9. Multiple regression equations to predict NHSCR inflows to non-London FPCAs.

All non-London FPCAs

(a) PNINM = 91.3 + 1.28(CINM) + 1.29(STRGIM) - 0.07(AFPOPM)

Step	Enter	R	R-squared	R-squared change
1	CINM	0.989	0.977	0.977
2	STRGIM	0.994	0.988	0.011
3	AFPOPM	0.995	0.989	0.001

(b) PNINF = -92.8 + 1.38(CINF) + 1.04(STRGIF) + 0.78(AFPOPF)

Step	Enter	R	R-squared	R-squared change
1	CINF	0.993	0.986	0.986
2	STRGIF	0.996	0.992	0.006
3	AFPOPF	0.997	0.992	0.001

Non-metropolitan FPCAs

(c) PNINM = 385 + 1.26(CINM) +1.17(STRGIM) -0.06(AFPOPM)

Step	Enter	R	R-squared	R-squared change
1	CINM	0.987	0.974	0.974
2	AFPOPM	0.992	0.984	0.010
3	STRGIM	0.993	0.986	0.002

(d) PNINF = 189 + 1.36(CINF) + 1.03(STRGIF) + 0.76(AFPOPF)

Step	Enter	R	R-squared	R-squared change
1	CINF	0.991	0.982	0.982
2	STRGIF	0.994	0.988	0.006
3	AFPOPF	0.995	0.989	0.001

Metropolitan FPCAs

(e) PNINM = -575 + 1.29(CINM) + 1.12(STRGIM) + 1.80(AFPOPM)

Step	Enter	R	R-squared	R-squared change
1	CINM	0.977	0.955	0.955
2	STRGIM	0.990	0.980	0.024
3	AFPOPM	0.993	0.986	0.007

(f) PNINF = -258 + 1.40(CINF) + 1.05(STRGIF)

Step	Enter	R	R-squared	R-squared change
1	CINF	0.981	0.962	0.962
2	STRGIF	0.993	0.987	0.025

PNINM (PNINF) = Predicted NHSCR inflow males (females)
CINM (CINF) = Census inflow males (females)
AFPOPM (AFPOPF) = Armed forces population males (females)
STRGIM (STRGIF) = Estimated student inflow re-registrations males (females)

FPCAs were included in the analysis. Table 9 illustrates the regression equations to predict male or female NHSCR inflows from the independent variables. R is a measure of the correlation between the dependent and the independent variables and R-squared measures the goodness of fit of the linear model to the observed data. The R-squared change figure gives an indication of the increase in the goodness of fit as successive independent variables are added to the model. For inflows the regression model best fits the data for female inflows to all non-London FPCAs although the effect of the student and Armed Forces variables upon this model is negligible (R-squared change = 0.006 and 0.001). For male inflows to all non-London FPCAs the Armed Forces variable again has little effect upon the fit of the model although the estimate of student inward re-registrations produces an R-squared change of 0.011.

Models (c) to (f) in Table 9 illustrate the generally poor effect of the Armed Forces variable on the predictive equation with the exception of model (c) where the variable produces an R-squared change of 0.01 for non-metropolitan male inflows. Non-metropolitan male flows are likely to be those most influenced by the Armed Forces variable as the large majority of Forces personnel are male and contained in non-metropolitan areas. In general, however, the proxy variable of usually resident Armed Forces population appears to be a poor substitute for the estimation of the level of movement between FPCAs within the Forces. In model (f) the variable does not meet the internal tolerance levels of the regression routine and is thus excluded from the model. The student variable appears to be of greatest value in the prediction of metropolitan inflows for both sexes (R-squared change = 0.024 and 0.025). With the importance of the Armed Forces variable in the prediction of non-metropolitan inflows the effect of the student variable upon the fit of the model to observed data is negligible.

These analyses demonstrate that although the fit of the models to the observed data is generally good, the importance of the independent variables varies. The Armed Forces variable is of questionable value as a substitute for the estimation of inter-FPCA Forces movement, although it does have a considerable importance within the model for predicting non-metropolitan inflows. The student variable appears to have the greatest effect upon the prediction of metropolitan inflows and it is the model to predict female outflows from non-metropolitan zones which gives the poorest fit to the data. Improvements to this basic model could be made with the addition of further independent variables such as a more accurate measure of inter-FPCA movement within the Armed Forces, a count of moves involving prisoners and long-term psychiatric patients or, more importantly, a measure of the effect of multiple and return migration. The problem is how to adequately estimate the spatial variation in the effect of such multiple moves.

5. CONCLUSION

The preceding sections of the paper have attempted to highlight some of the major similarities and differences that exist between Census and NHSCR data at a number of spatial scales and different aggregations. Here we summarize the findings and provide an interpretation of the major characteristics in terms of the components identified in Section 2 of the paper.

Figure 6 emphasised the strong correlation between Census transitions and NHSCR moves at three spatial scales. At the standard region level correlation coefficients of over 0.99 are observed. The coefficients drop to approximately 0.97 when we break up these totals into flows between FPCAs. It is only when the net-flows are considered that the correlation drops to any degree - to 0.896 in the case of net migration at the FPCA level. The general patterns of migration captured by the two data sources are thus very similar.

However, important differences do exist between NHSCR and Census figures. Table 3b illustrates that with the assignment of the Armed Forces and not-stated flows the total NHSCR inter-FPCA moves exceeds those of the Census by 37%. A scale effect is clearly evident, with a ratio of 1.536 between standard region flows highlighting a greater discrepancy between longer-distance migration.

When we compare NHSCR and Census counts across spatial zones we observe substantial differences. A distinct variation between outflow and inflow ratios for metropolitan and non-metropolitan zones is observed in both Table 5 and Figure 7. The assignment of Armed Forces moves in the NHSCR data increases the ratio value particularly in non-metropolitan zones with large Forces populations. This is emphasised by the sharp increase in the ratio of inflows and outflows to/from East Midlands, East Anglia and the South West after assignment illustrated in Table 5. The estimation and assignment of moves within the Armed Forces (postings) would further increase the ratio value in these non-metropolitan zones. At the FPCA level the highest inflow ratios appear in metropolitan districts - Sheffield (1.74), Newcastle (1.67), Leeds (1.62), Manchester (1.59), Birmingham (1.53) and Liverpool (1.46) - illustrating the importance of the in-movement of students to higher education establishments.

Table 6 further emphasised the metropolitan/non-metropolitan discrepancy and reinforced the hypothesis that it is shorter-distance flows which exhibit the greatest degree of consistency between the Census and the NHSCR. Return migration was held to play some part in the higher NHSCR/Census ratios observed for inflows to metropolitan areas as a whole: areas unattractive to migrants were associated with returns to more attractive zones (non-metropolitan areas). Such return flows depressed the observed Census count while not affecting the NHSCR count.

Section 4.3 outlined the important discrepancies that exist between age and sex groups, noting particularly the under-recording of moves by 15-19 year-olds in the computer summary data (Stage 1 comparison) due to the exclusion of Armed Forces information. The greatest ratio values are exhibited by the 15-19, 30-34 and 70-74 age-groups for both sexes and also for males aged 10-14. Student distortion will have a large positive effect on the 15-19 age-group ratio value. The greatest consistency between the Census and the

NHSCR is observed between male flows in the 20-29 age-range where non-registration by the more mobile members of the population is of greatest importance and gives rise to a major difference between male and female ratio values.

It is clear that the non-common components of the two migration measures act together and that the observed NHSCR/Census differences for any one area are the net outcome of the various effects outlined above. The results presented here have merely highlighted areas where one effect was clearly dominant. The expansion of our bivariate regression equations relating NHSCR and Census migration, to multivariate models for FPCA total inflows demonstrates the importance of Armed Forces in non-metropolitan area in-migration and of students in in-migration to metropolitan areas.

The comparisons reported here and in Working Paper 495 have therefore established a number of possible guidelines regarding the use of migration data in population analysis. Firstly, a record of Armed Forces recruitment and discharge moves together with Armed Forces dependants moves is available from the PUD and this paper has illustrated a method for re-assigning the flows to individual origins and destinations. However, further information is needed from the Ministry of Defence in FPCAs with large Armed Forces populations to account for the internal transfer of service personnel within the Armed Forces (postings).

Secondly, the NHSCR has the advantage of locating students at their places of education in contrast to the Census which records the usual residence of students as their home address and thus excludes moves to University or College from its tabulations. Large ratio values for inflows to metropolitan districts have been illustrated thus supporting the argument for using NHSCR figures.

Thirdly, there are a considerable number of 'not-stated' flows in both the Census and the NHSCR which require re-distribution in order to establish the correct level of migration. Methods for re-assigning these flows to individual FPCAs, age-groups and sexes are outlined in this paper.

Fourthly, the conceptual differences between the two sources of migration data have been outlined emphasising the need to match the population model utilised to the type of migration data available. Census data entails a transition model, NHSCR data a movement model.

Fifthly, moves by the under 1's should be included in any population model. This again favours the NHSCR as the Census excludes all infant migrants.

Finally it is probable that the undercounting of moves by the NHSCR is not as serious as the considerable under-enumeration evident from the 1981 Census data, again supporting the use of the re-registration information instead of the Census.

Taking all these factors into account it appears better to use NHSCR data streams directly for inter-censal years and use Census data to fill the spatial gaps than to use Census data directly and NHSCR indirectly to fill the temporal data gaps, adopting a movement model for projection in favour of a transition model.

References

Boden, P., Stillwell, J.C.H. and Rees, P.H. (1987a) Migration data from the National Health Service Central Register and the 1981 Census: further comparative analysis, Working Paper 495, School of Geography, University of Leeds

Boden, P., Stillwell, J.C.H. and Rees, P.H. (1987b) Migration data from the NHSCR and the 1981 Census: further comparative analysis. Paper presented at the IBG Population Geography Study Group conference, Mansfield College, Oxford, September 1987

Britton, M. and Birch, F. (1985) 1981 Post-Enumeration Survey, HMSO, London

Courgeau, D. (1973) Migrants et migrations Population, 28, 95-129

Devis, T. (1984) Population movements measured by the NHSCR, Population Trends, 36, pp 15-20

Devis, T. and Mills, I. (1986) A comparison of migration data from the National Health Service Central Register and the 1981 Census, Occasional Paper 35, OPCS, London

Gordon, I. (1975) Employment and housing streams in British inter-regional migration, Scottish Journal of Political Economy, Volume 22, pp 161-177

Gordon, I. (1982) The analysis of motivation-specific migration streams, Environment and Planning A, Volume 14, pp 5-20

Ledent, J. (1980) Multi-state life tables - movement versus transition perspectives, Environment and Planning A, 12, 533-562

Ogilvy, A.A. (1980a) Inter-regional migration since 1971: an appraisal of data from the NHSCR and Labour Force surveys, Occasional Paper 16, OPCS, London

Ogilvy, A.A. (1980b) Migration figures from the NHSCR: adjustments for differences in age/sex recording. Building Research Establishment Information Paper 16/80

OPCS (1981) Census 1981: Definitions, Great Britain, HMSO, London

OPCS (1984) Population projections, area. Series PP3, No.5 HMSO, London

Rees, P.H. (1981) Accounts based models for multi-regional population analysis: methods, program and users' manual. Working Paper 295, School of Geography, University of Leeds

Rees, P.H. (1984a) Spatial population analysis using movement data and accounting methods: theory, models, the MOVE program and examples. Working Paper 404, School of Geography, University of Leeds

Rees, P.H. (1984b) Does it really matter which migration data you use in a population model? Chapter 5 in White, P. and van der Knaap, B. (eds.) Contemporary Studies of Migration, Geo Books, Norwich, pp 55-77

Rees, P.H. (1986) Developments in the modelling of spatial populations, in Woods, R. and Rees, P.H. (1986) Population structures and models, Allen and Unwin, London

Rogers, A. and Castro, L.J. (1986) Migration, in Rogers, A. and Willekens, F.J. (1986) Migration and settlement - a multi-regional comparative study

Thomson, P. (1984) A comparison of the 1981 Census and the NHSCR as sources of migration data. Paper presented at the conference on Census Analysis and Applications, Sheffield, 20-22 September 1984.