Inequalities In Life and Death - What If Britain Were More Equal? The Technical Report

Richard Mitchell, Danny Dorling, Mary Shaw

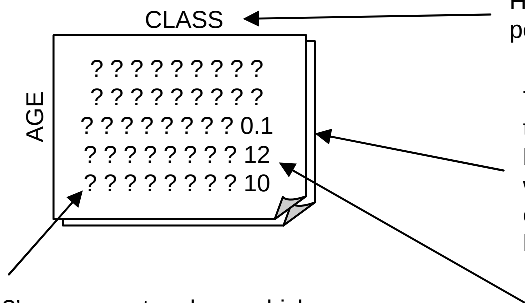
This report accompanise *Inequalities In Life and Death - What If Britain Were More Equal?,* published by Policy Press on behalf of the Joseph Rowntree Foundation. This report is unlikley to make much sense without the main part which explains why this work was undertaken, its context and results. The report is available from Policy Press ISBN number: 1 86134 234 9. Copies of this technical appendix are available free from www.social-medicine.com or by emailing a request to inequality@social-medicine.com

The technical report explains, step by step, how data needed for the work described in *Inequalities in Life and Death...* were obtained, manipulated and analysed.

The first part of the technical report concentrates on the use of various techniques to obtain the data necessary to describe and account for changes in the geography of mortality in Britain between the early 1980s and the early 1990s and to test various social policy scenarios for their impact on the patterns of mortality. The data necessary to achieve these aims and tests were counts of the population falling into every possible combination of age group, sex, social class (I-V) and employment status (employed, unemployed and economically inactive) for every parliamentary constituency in Britain, for two points in time (1983 and 1993). Such data are not available in the public domain but the building blocks required to construct or estimate them are. Sections 1 - 10 describe how these data were obtained. Relatively little about this part of the research is mentioned in the main report owing to its complex and rather tedious nature.

The second part of this technical report describes how these data were then utilised to answer the various questions posed by the main report. Much of this process is touched on in brief in the main report and where basic concepts have been explained or demonstrated in that text, they have not been repeated here. The two reports should thus be read together for maximum understanding.

Throught the technical report we have used diagrams, rather than text, to explain and clarify the techniques and processes used. In those diagrams we have used a number of stylised symbols to represent data and techniques. The key below explains these.

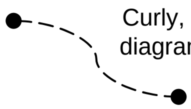


Headings represent the categories into which people have been placed

This symbol has been used to represent tables of data. In reality the data were held by computer, but the paper analogy works well. Where two 'pages' appear on top of each other, it signifies that these data were held separaely for men and women

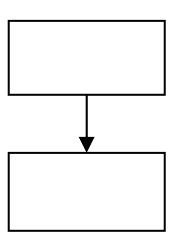
?'s represent values which we needed to know, but did not have available and therefore had to estimate.

Numbers within these tables represent known or estimated values but NB the numbers which appear in the report ARE NOT REAL, they just represent data.



Curly, dotted lines connect a diagram to some explanation

Stick figures represent the population

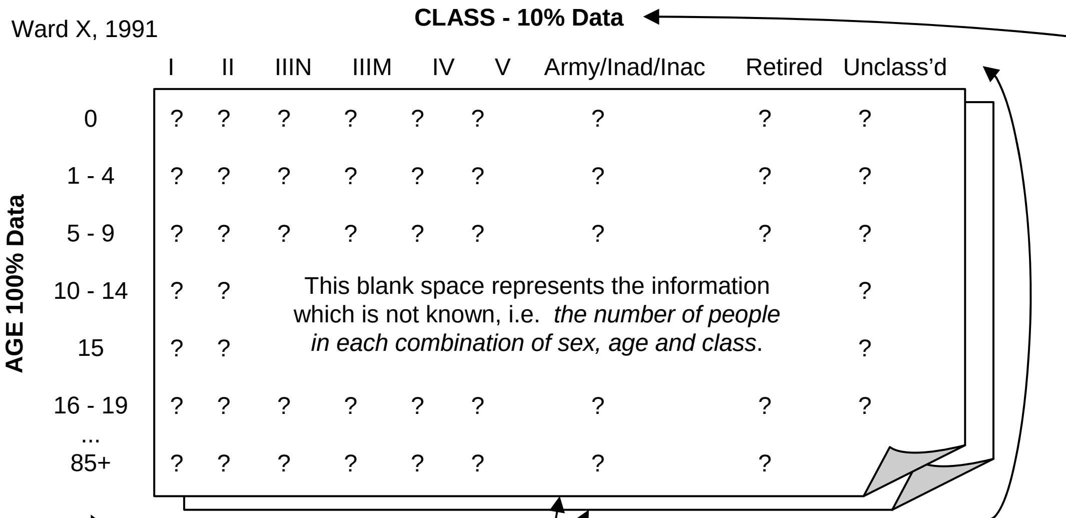


Flow charts represent a sequence of processing undertaken by computer

If you have any difficulties understanding the report please direct your questions to: inequalities@social-medicine.com

How we got Age / Sex / Class / Employment Status Counts Using Iterative Proportional Fitting

1. This is the starting point; the picture shows a hypothetical table labelled with the data which were readily available, and with those which we needed to simulate.



Notice, that at this stage "class" included a category for the retired, those in the army and those for whom the census was unable to assign a class, not just the usual I-V

The total numbers of people in each age group for this ward were known from the census. These values are called the *age marginals*

These two pages represent the fact that we needed to know this information separately for men and women.

The total numbers of people in each class, and those not in any class in this ward were known from the census. These values are called the *class marginals.*

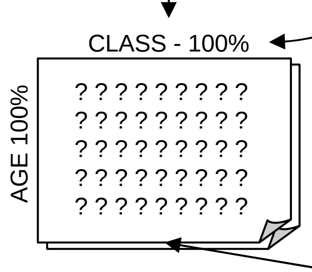
2. There were two preliminary steps we needed to take before we could begin to simulate the data we needed.

(i) To keep things simple, the 10% counts from the census were converted to pseudo-100% counts. This was done by scaling the 10% counts upwards so the total number of people in the class marginals was equal to that in the age marginals.

(ii) We needed to give a social class to the small number of economically active people who were not given a class of any kind in the 1991 census. (A different census system meant that all the economically active people had a class in the 1981 census, so this only applied to 1991).

These people were "given" a class based upon the way in which other people in the same ward, of the same sex, were divided amongst the classes. In effect, we assumed that the unclassified have the same class profile as their neighbours. For example, if 30% of the economically active men in the ward were in class IIIM, and there were 3 unclassified men, 1 of them would be added to the IIIM group.

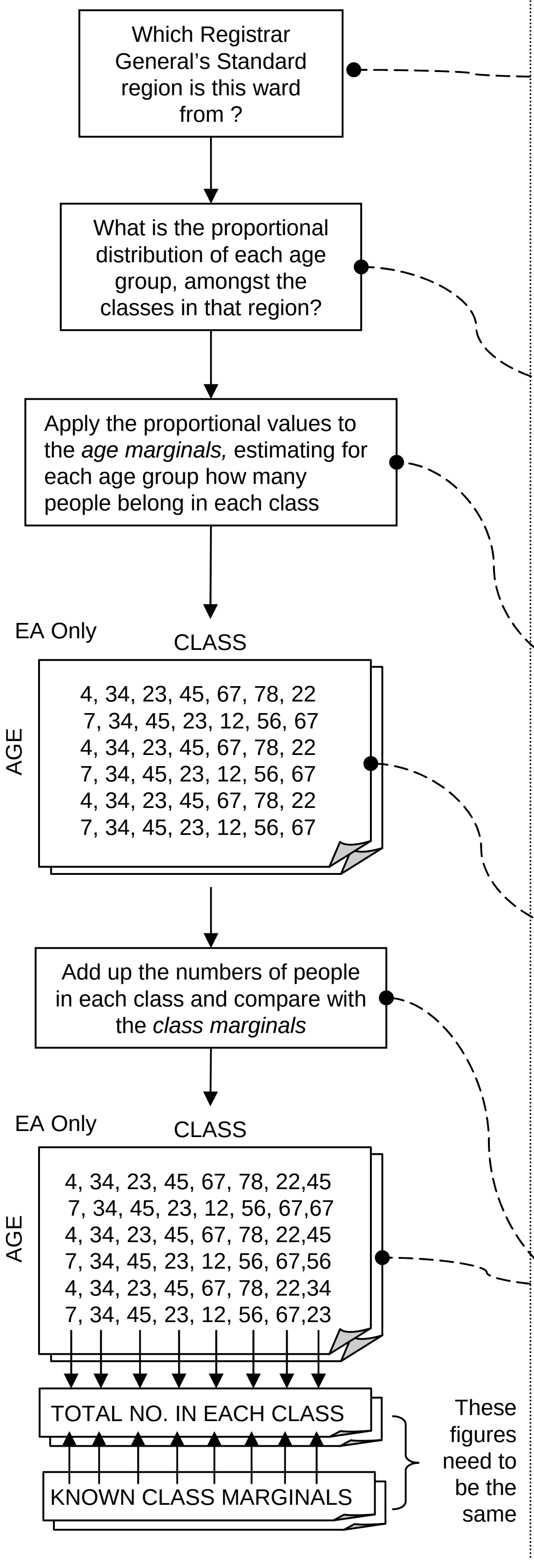
This was a slightly problematic solution since this group of people is more likely to be at the lower end of the social scale but because it is a relatively small number of people (about 2% of the census respondents) our solution was unlikely to impact on the results to a great extent.



These actions done separately for men and women

3. With those preliminary stages complete we could proceed with turning the ?'s shown in stage 1 into estimated values. The following series of diagrams explains how the estimates were derived. A computer program was written (in C++) to execute the stages described.

What Was Done



What It Means

(a) The "*standard regions*" are the familiar divisions of Britain such as, "South East", "West Midlands" and "North East". They include the countries Scotland and Wales as whole "regions". Each electoral ward in Britain, nests within one of these regions. It is widely suggested that the regions contain quite distinct types of population, richer, poorer, older, younger, more working class, more professional.

(b) We obtained this information from the ONS Longitundinal Study (LS). For each region we know how a particular age/sex group is divided amongst the various social classes. For example, in the North of England, 31% of all men aged 30-34 are in class IIIM.This is called a *distribution.* Children in the census are given a social class based on their parents. Since the LS does not cover Scotland we have used the *distribution* from the North of England for Scottish wards. This information was used as a starting point for our estimation of the missing data.

(c) Using the appropriate *distribution* for this ward, each age group was divided amongst the classes in the proportions specified by the *distribution.* So if the ward happened to be in the North of England, and there are 100 men aged 30-34 in that ward, 31 of them were placed in class IIIM.

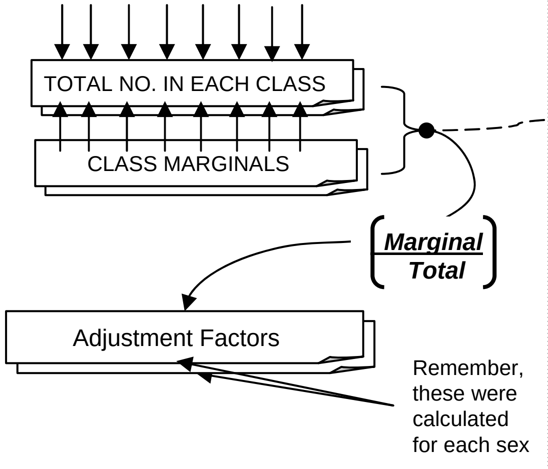
(d) This was completed for all age and sex groups, giving us a first estimate of the number of people in each age/sex/class group. The total *number* of people in an age group, in different classes, matched the *age marginal,* so we had not added or removed anyone. **N.B. *The numbers in all the following diagrams are for illustration only! They do not represent plausible values .*** EA Only means that this table only contains information about people who are economically active.

(e) Since it is very unlikely that the class *distribution* in the ward is identical to that in the wider region, when we added up all the numbers of people in each class, they did not precisely match the *class marginals*. Perhaps, for example, this is a ward with a lot of young professional women living there. In this case, it is likely that we will have fewer women in class I when we add up than we have in our *class marginals.* The *class marginals* come from the census and we assume that they accurately represent the ward's population.

What Was Done

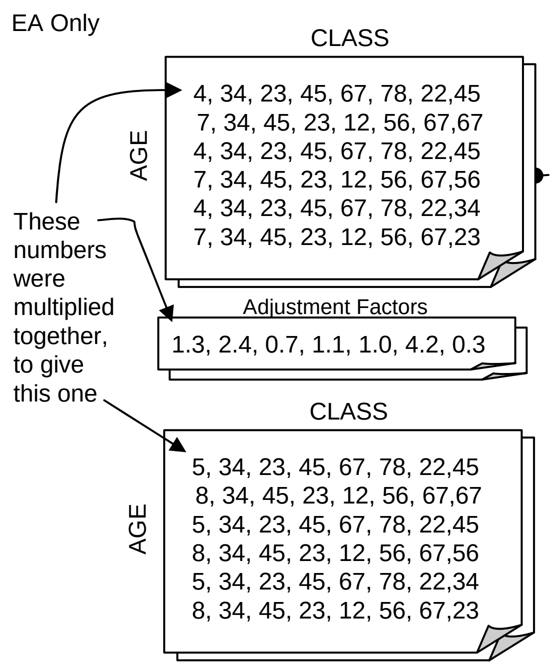
What It Means

(f) In order to adjust the numbers of people in each class so that the totals we had estimated matched the known *marginals*, we computed an adjustment factor for each class. This was simply the *marginal* divided by the total number of people estimated to be in that class. If the *marginal* was 200 and our estimated total was 150, the adjustment factor would be 1.333.



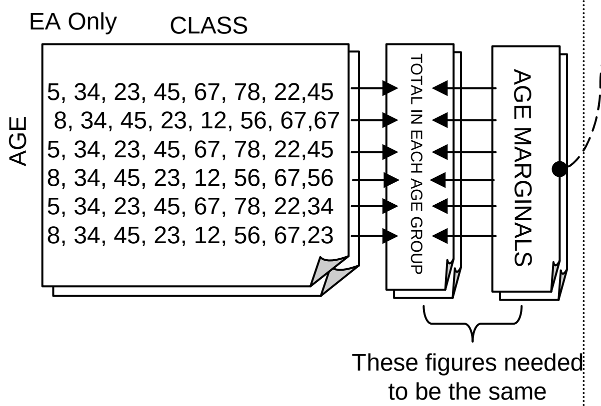
(g) The number of people in each age/sex/class grouping was multiplied by the adjustment factor for that class. The same class/sex specific adjustment factor was used for each age/sex group. So, for example, all the men in class I were subject to the same adjustment factor, regardless of their age, but the factor differed to that for women in class I. In the example given above, we *increase* the number of people in the class by about one third, by multiplying the numbers of people in each age group, in that class by 1.333. So, where as before we only had 150 people in the ward in that class, we now have 200 - the right number.

Once carried out on all the classes (remember this includes the “retired” and “army”), the process resulted in a whole new set of estimates of the numbers of people in each age/sex/class group.

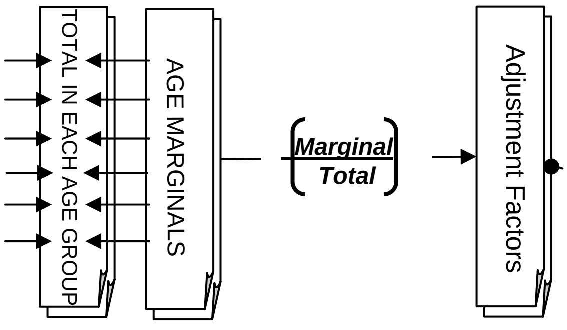


(h) We had now reached a similar situation to that at stage (e), except that we were likely to have small differences between the numbers of people in each *age marginal*, (which we knew from the census is the real number of people in each age group), and the number of people we had in each age group once we add up the table cells after the adjustment described in part (g).

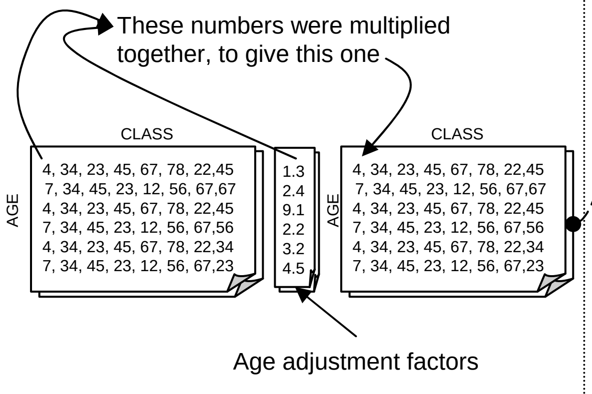
To ensure that we have the right numbers of people in each age group, we carried out the same kind of adjustment as described in parts (f) and (g).



What Was Done	What It Means



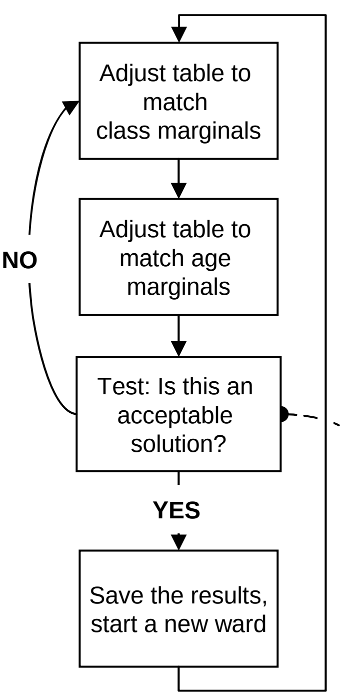
(i) This stage mirrors parts (f) and (g). The adjustment factor, for each age/sex group was computed...



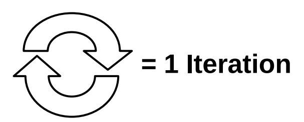
... and then applied to each appropriate cell. For example, the factor for men aged 30-34 is applied to all men aged 30-34, regardless of their class. This process results in a new set of estimates.

Sections (e) to (i) have described the stages whereby our initial estimates of the numbers of people in each age/sex/class grouping, in a ward, were adjusted. The goal was to produce a "solution" in which each *marginal*, value was equal to the number of people in the the appropriate cells within the table. By repeating these stages over & over again we could get a table which exactly (or very nearly) reached that goal. The flow diagram below summarises this process.

In about 11 cases, a ward failed to converge with an acceptable solution. This occurred where the ward was very small, or when it had a highly unusual class/age structure (for example, wards which contain large army barracks). In these cases the table cells were filled as though the ward had an age/sex class structure identical to the Registrar General's Std Region within which the ward is located.



One complete cycle through the class and age adjustment stages is called an *iteration*. Most wards converged after 2 or 3 iterations.



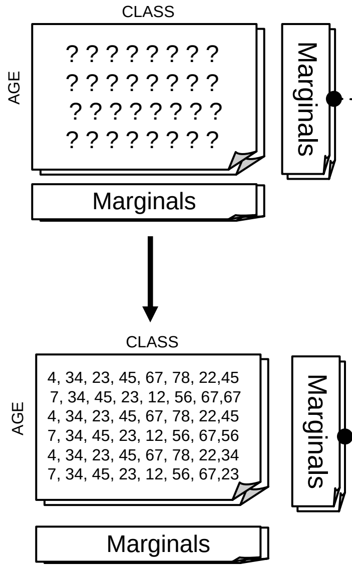
This process is called ***iterative proportional fitting*** or IPF.

The IPF stopped when the known marginals and estimated totals were less than 0.5% apart. That is 0.005 of a person.

4. Summary - A recap of what we had achieved up to this point.

What Was Done

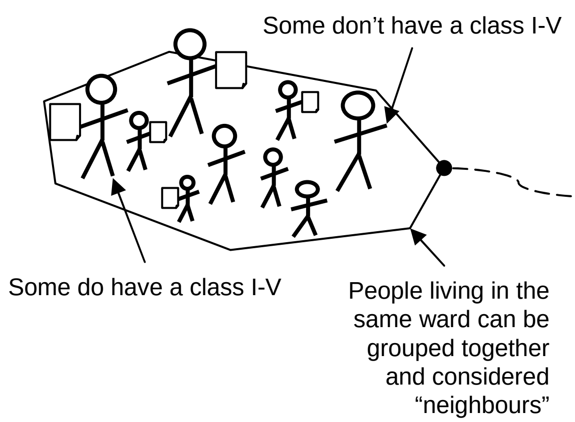
What It Means



At the start we knew the total number of people in each sex/class grouping (*class marginals*). We also knew the total number of people in each age/sex grouping (*age marginals*) but we did not know the numbers of people in each age/sex/class grouping. For example, we knew how many women there were in social class IIIM, in this ward. We also knew how many women aged 20-25 there were in this ward but we did not know how many women aged 20-25, in social class IIIM there were.

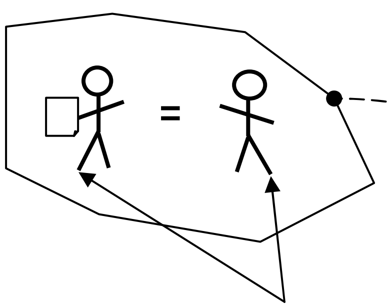
Now we had a good estimate of how many people there are in each age/sex/class grouping. Following the example given above, we had an estimate of how many women aged 20-25 , in social class IIIM live in this ward. The numbers of people in our estimates match the number of people living here, taken from the census. By writing a computer program to automate this, essentially, very simple, task, we created these estimates for every ward in Britain, for 1981 and 1991

5. The next stage was to assign everyone in the ward, a social class between I and V. We needed to do this because the information about death rates which was available to us only differentiated between this set of classes. At this stage, our task got slightly more complex in nature and in solution but it was tackled using the same approach as before; break it down into small sections and use a computer program to automate each section for every ward. We will first describe some concepts which we applied in relation to this task, and then explore how the concepts were operationalised.



When constructing the computer code to give everyone a class between I and V we tried to remember that the numbers in our data sets represent *real people*. The diagram here represents each ward as a group of people of varying ages and sexes. Some of them have a class already, represented by carrying paper, some do not.

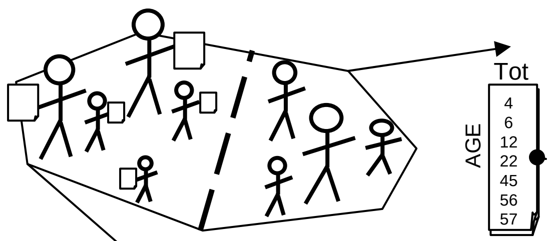
We treated those who live in the same ward as "neighbours", and, as has been found in the real world, we assumed that people who share the same neighbourhood often have similar social and economic characteristics. We also assumed that the characteristics of a ward tend to change quite slowly, and that whilst people may move in and out over time, the type of people who make up the ward's resident population tends to stay the same.



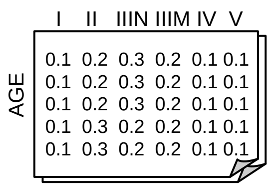
Those who have a class, I-V, define the probability of which class those without, will be given

Our method used the characteristics of those people in the ward who *did* have a class between I&V to assign a class to those who did *not* have a class. Those without a class would get one which made them just like their neighbours of the same age and sex. For example, women, aged 30-34 in this ward who *have* a class between I & V, define the probability of being in class I, II, IIIN, IIIM, IV or V for those women aged 30-34 in the ward who *do not have* a class between I&V.

What Was Done

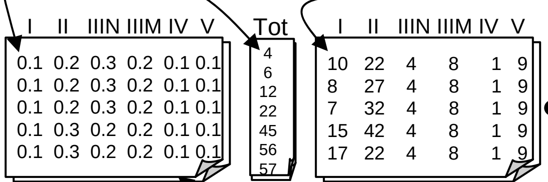


The number of those without a class I-V, in each age/sex group is determined; the "classless"



The distribution of those *with* a class, I-V is calculated. There is a proportional value for each age /sex/ class group and it gets used here

This proportion,
of this number of "classless people"
are added to this class



Remember, we are still doing each stage separately for men and women

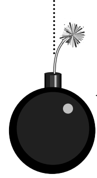
What It Means

(j) The first step was to work out how many people in the ward, in each age/sex group, did not have a class between I & V (let's call them the "classless" for a moment).

Then, the distribution of those with a class between I & V was established. We calculated what proportion of each age / sex group belonged to each class. For example, about 8% of men aged 30-34 were in class I, about 25% in II, about 20% in IIIN, about 30% in IIIM, about 10% in V and about 7% in V.

(k) The "classless" people were placed in a class between I & V according to the proportional distribution determined in step (j). So, following the example above, suppose there were 100 men aged 30-34 who needed to be given a class. 8 of them would be added to the total number of men aged 30-34 in class I, 25 to those in class II, 20 those in class IIIN etc. In this way, the number of people in the ward remained the same, but everyone ended up in a class between I and V.

Applying this method without careful thought about the people our data represent would have produced misleading results

BUT...  **...WARNING!**

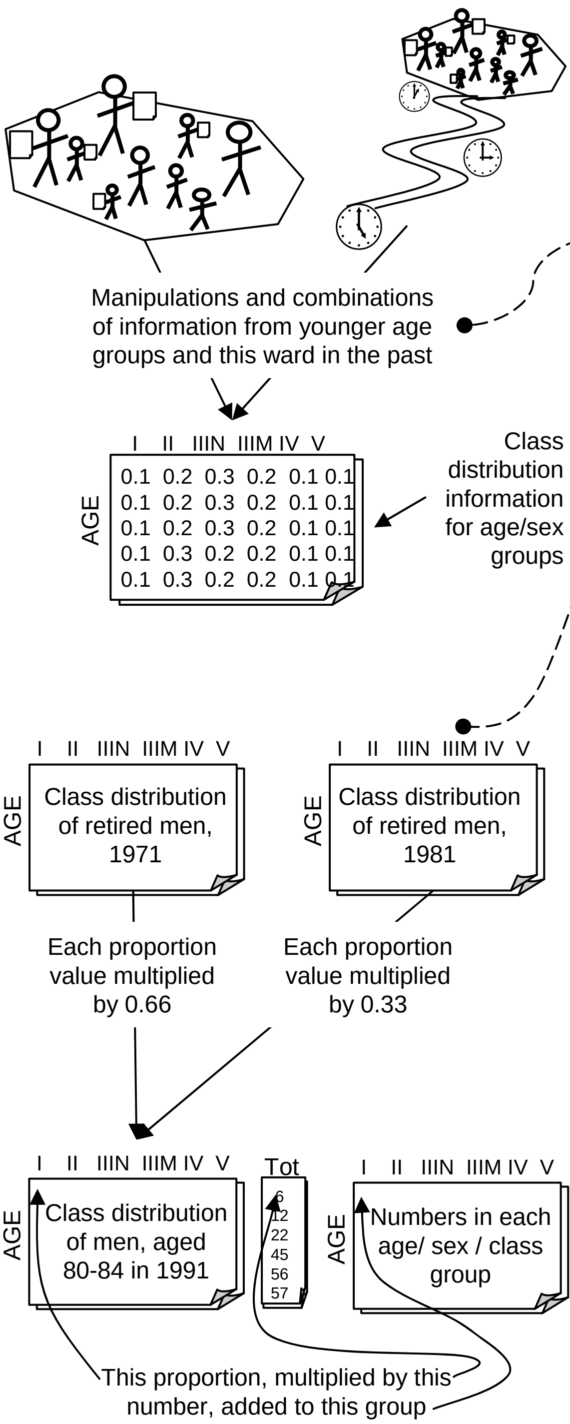
Whilst the method outlined above would have worked well for age groups in which the majority of people had a class I-V and for which there was therefore a reliable distribution of class identity, we knew it would fail when all or most of the members of an age/sex group were "classless" and the vast majority of those without a class I-V are retired people. Unlike the 1981 census, the 1991 census failed to report the previous occupations or classes of the retired. We therefore had no information through which to place the classless retired people into classes I-V.

We also recognised that there has been a radical change in the labour structure of women within the last few decades because so many more women work in the formal labour market, rather than at home as unpaid housewives. Since class, in the census, is based on formal occupation, we anticipated that there would be a strong relationship between being a "classless" woman and age which might distort any class distribution drawn from those women who have a formal occupation, and also that there might be a class bias in the number of women who work in the formal labour market. We therefore needed a more sophisticated version of the technique in order to deal with these extra complexities.

This matters because of the nature of our interest in identifying people's class. We know from previous research that class plays a large role in determining an individual's chance of dying. Just because a woman does not qualify for a class in the UK census system because she is not in the formal labour market, does not mean that class holds no influence on her health. We needed to infer her position within the class hierarchy to estimate that influence on her life chances. The same argument holds for retired women *and* men. Just because a man is no longer economically active does not mean that his class, perhaps determined by the job he used to do, holds no influence over his chances of dying. The problem grows for older retired people, since their class identity may be based on a labour market which has dramatically changed since their retirement. Those newly retired might have a very different class structure to those retired 10 or 20 years before.

Where data are available about individuals, inference about class identity can be drawn from work history, partner's class etc. However, we were restricted to using aggregate data which described the nature of the area's population. We employed two solutions to the problems described.

What Was Done



What It Means

(j) The key to dealing with this problem was the source from which the distribution information for classes I-V was drawn / built. We had two sources of information available. We knew the I-V class distribution amongst the younger age groups within the ward, and we also had access to the census data for the ward from 1981 and 1971. We could therefore look "back" to younger groups and/or look back in time to see what the class profile for older men and women in the 1980s and 1990s was like.

(k)These different sources of information were blended together by adjusting their relative contributions according to the group of people for whom we were trying to produce the class distribution information. To do this we multiplied the different distributions by proportional weighting factors.

In the example shown we combined $^1/_3$ of the distribution of retired men for this ward in 1981 and $^2/_3$ of that for retired men in 1971to give an appropriate distribution for men, aged 80-84 in 1991. Consider the group in question; nearly all of them will have been very close to retirement in 1971 and (we assumed), will then have exhibited a similar class profile to their slightly older, retired peers. However, since we only had information on the retired male population as a whole for 1971, and this included very much older men, with perhaps a different structure again, we temper the influence of the 1971 retired with a contribution from the 1981 retired, which is also likely to include a good proportion of the group in question.

In some cases we used the class distribution of a younger age group to give class to an older group. This was particularly useful in the late 50s-early 60s age groups where there is a class bias amongst those who have taken early retirement. If we had classified those who were retired at 55 (and were thus classless), using those who were still working, we would have created too many working class people, because a higher proportion of those who retire early are drawn from the higher classes.

Our choice of weighting factors was based on intuition rather than any established criteria but it was tested and fine tuned, as will be explained later.

What Was Done

These rules were used to give a class to people in the *1981* Census

For ages < 55, both sexes, class assigned using the distribution from those of the same age and sex.

men, aged 55-59, class assigned using the distribution of working men of same age

men, aged 60-64, class assigned using the distribution of working men aged 55-59

men, aged 65-69, class assigned using the distribution of all men aged 60-64, calc'd above

men, aged 70-74, class assigned using the distribution of all men aged 65-69, calc'd above

women, aged 55-59, class assigned using the 0.75 of distribution of working women of same age and 0.25 of working men, same age

women, aged 60-64, class assigned using the 0.69 of distribution of working women aged 55-59 and 0.31 of working men, aged 55-59

women, aged 65-69, class assigned using the 0.69 of distribution of all women aged 60-64, calc'd above and 0.31 of men, retired in 1981

women, aged 70-74, class assigned using the distribution of all women aged 65-69, calc'd above

men and women, aged 75-79, class assigned using 0.66 of the distribution of men retired in 1981, and 0.33 of the distribution of men, retired in 1971

men and women, aged 80-84, class assigned using 0.33 of the distribution of men retired in 1981, and 0.66 of the distribution of men, retired in 1971

men and women, aged 85+, class assigned using the distribution of men retired in 1971

What It Means

Computer code was written to do these calculations, for every ward in Britain.

(l) We found that different sets of rules were needed to produce our estimates for 1981 and 1991. This is because of the limited nature of the data available to us, and because of the aforementioned changes in women's relationship with the labour market.

(m) The computer was programmed to determine the age / sex group in question and follow a particular set of rules, as shown in the flow chart. Remember, only the people without a class between I&V were assigned one. Those with a class between I&V kept it, no matter how old they were. Remember too, that children were assigned a class by the census, based on their parents. If their parents were been "classless", our program, in effect, assigned a class based on the parents of children the same age and sex, living in the same ward, who did have a class between I&V.

The proportional values (e.g. 0.69, 0.75) were derived using LS data as a guide.

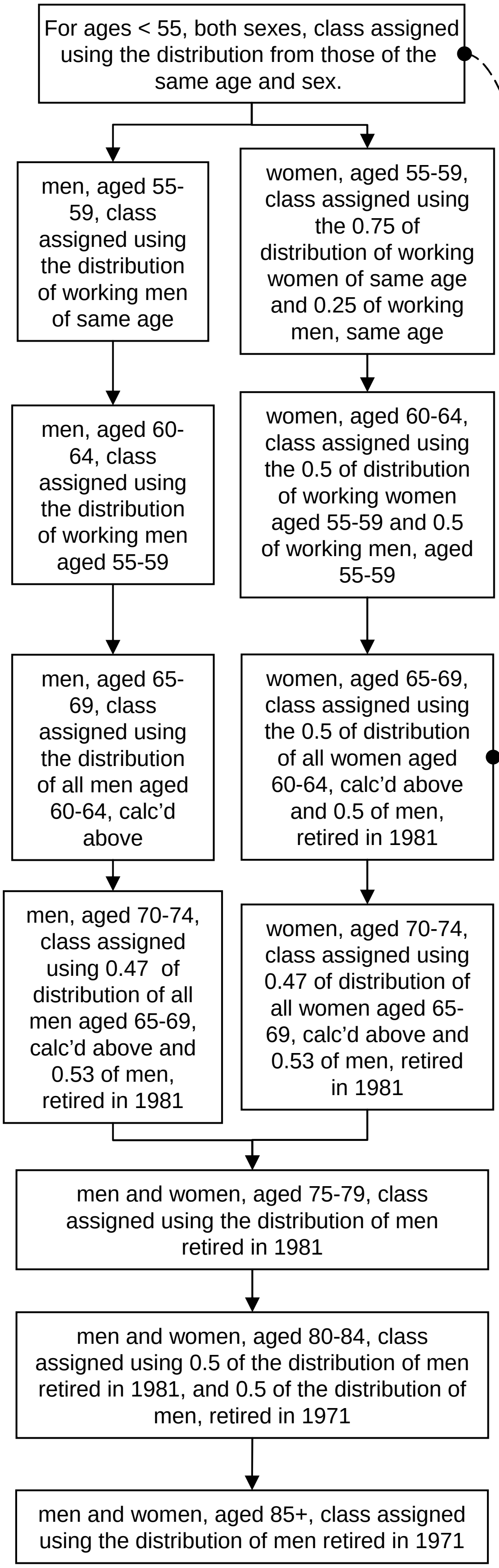
Notice that the program used the class distributions which resulted when everyone in an age/sex group had been assigned a class, to influence older age groups.

There were a few instances in which there was no one with a class in a particular age/sex group, but there were people without a class in that age/sex group. These people were assigned a class using the national distribution as a guide.

In fact, the 1971 census does not provide a table which tells us the social class of retired men, but it does provide their socio-economic group (SEG). Class and SEG are very closely related and by using the Samples of Anonymised records to tell us which SEGs belong to which classes, we were able to estimate a class distribution for men, in 1971.

What Was Done

These rules were used to give a class to people in the *1991* Census



What It Means

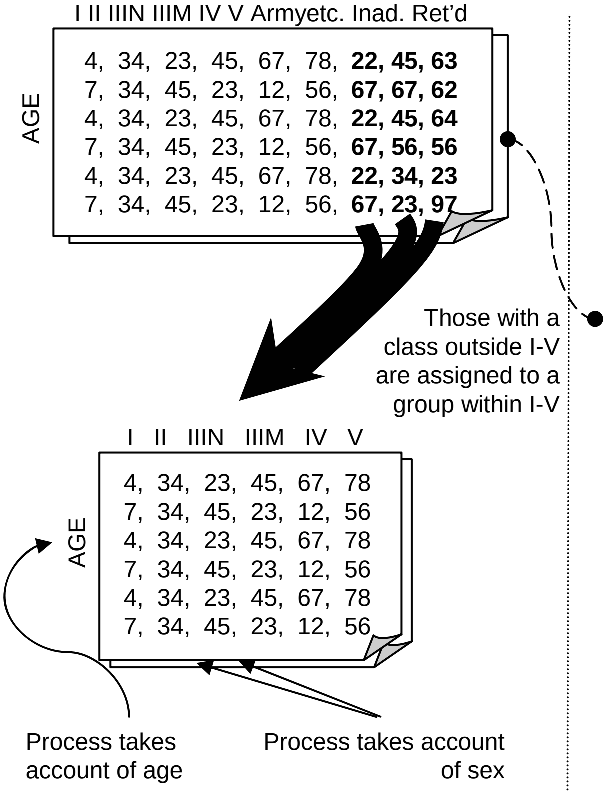
(l - ii) We found that different sets of rules were needed to produce our estimates for 1981 and 1991. This is because of the limited nature of the data available to us, and because of the aforementioned changes in women's relationship with the labour market.

(m - ii) The computer was programmed to determine the age / sex group in question and follow a particular set of rules, as shown in the flow chart. Remember, only the people without a class between I&V were assigned one. Those with a class between I&V kept it, no matter how old they were. Remember too, that children were assigned a class by the census, based on their parents. If their parents were been "classless", our program, in effect, assigned a class based on the parents of children the same age and sex, living in the same ward, who did have a class between I&V.

Notice that the program used the class distributions which resulted when everyone in an age/sex group had been assigned a class, to influence older age groups.

There were a few instances in which there was no one with a class in a particular age/sex group, but there were people without a class in that age/sex group. These people were assigned a class using the national distribution as a guide.

6. Summary - A recap of what we had achieved up to this point.

What Was Done	What It Means
	We started this section with an estimate of how many people in a ward belonged in each age/sex/class group, but amongst the possible class groups were a number outside the standard I-V. These included retired people. We needed to assign those outside the I-V to an appropriate class within the I-V scheme because we know the effect of being within the I-V hierarchy on the chances of dying. We drew on information about those who were within the I-V scheme and about what kinds of people were living in this area 10 and 20 years ago. We wrote more computer code to follow a set of rules for assigning these groups to the right class, based on their age and sex.

7. How did we know if we had got our rules right and given people a class between I & V, in a plausible way?

In order to make sure that we had got our rules right, and to fine tune them, we compared our results with known values from the LS. The LS was able to give us information on the distribution of people in age/sex/class groups and uses the I-V class scheme we were trying to estimate. Researchers using the LS have access to a great deal of information which allows them to make fairly accurate allocations of individuals to a class within the I-V scheme, even if they are retired or outside the formal labour market. The process uses information about their families, job history and the people they share homes with. However, this information was only available in aggregate for England and Wales. To test our program, we aggregated our estimates to mimic the LS data, by adding together, our estimates for wards in England and Wales.

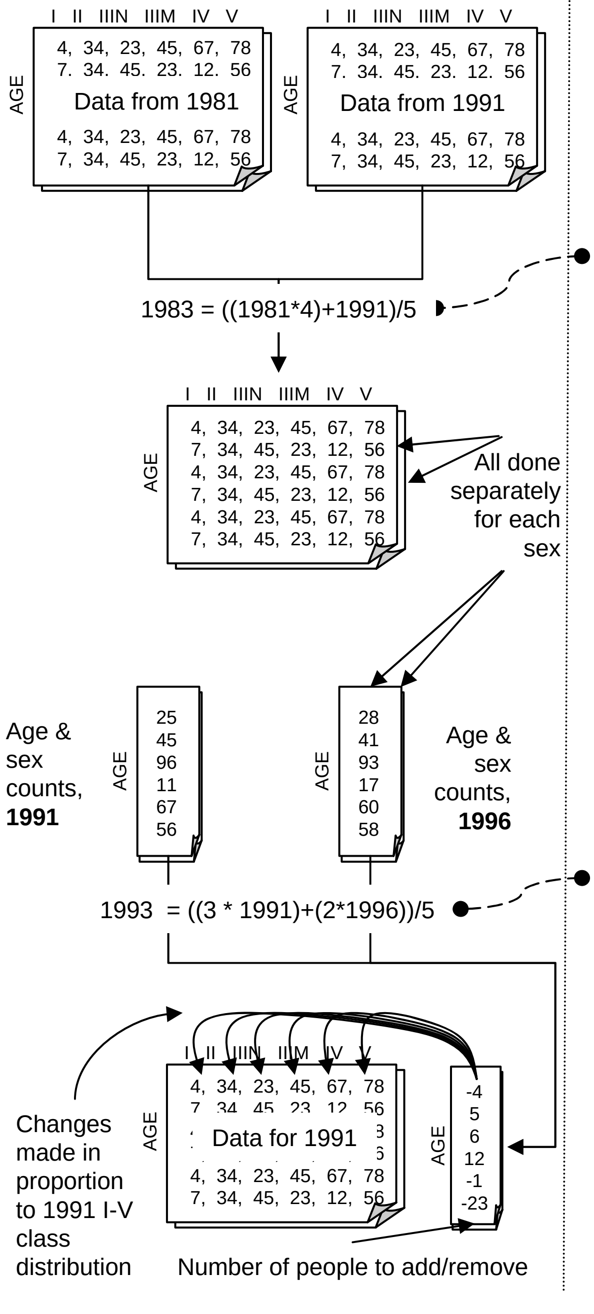
We tested to see whether, after we had allocated everyone a place within the I-V scheme, our estimates of the distribution of men and women across age and class groups was close to that from the LS. Initially we found some discrepancies between our own estimates and the LS figures. However, these were primarily due to the different way in which the LS allocated a class to women outside the I-V scheme. Our system gave women a class in such a way as to make them appear similar to other women of the same age, living in the same area. The LS system gives these women a class based more firmly on their immediate partners and family. It was important to try and get our data to match as closely as possible to the LS data, since this was also the source of our information about how class affects the chances of dying. In the end, we compromised on a set of rules by which our figures match those from the LS very closely amongst older people (when rates of death are much higher and therefore carry more importance) but did not match so closely at younger ages. This involved raising the influence of the class distribution as defined by men, in a ward, on older women who need to be allocated a class within the I-V scheme.

8. The final stage was produce estimates of the numbers of people, in some age/sex/class groups, who were unemployed, employed or economically inactive (i.e. their employment status). This stage only applied to people aged between 16 and 64, since by definition, those younger than 16 cannot be unemployed, and nearly all those over 64 are retired.

Before we did this however, we made a number of small adjustments to the data we had already produced. These adjustments had two purposes; i) to prepare the data for use in our analyses of death across Britain, through time and ii) to render them more useful to us in producing estimates of employment status.

Two adjustments were made to the data for both 1991 and 1981. The first projected the data forward in time by two years, to 1993 and 1983 respectively. We needed our data for these specific years because they fall in the middle of the two, five year periods for which we had death records available. We study deaths in five year periods because they form a more stable data set. If we simply took deaths for one year, our results might be distorted by time specific events such as an outbreak of flu amongst the elderly, or meningitis amongst the older teenagers. The five year death periods ran from 1981 to 1985 and 1991 to 1995. The adjustment procedure was slightly different for the two time periods.

What Was Done



What It Means

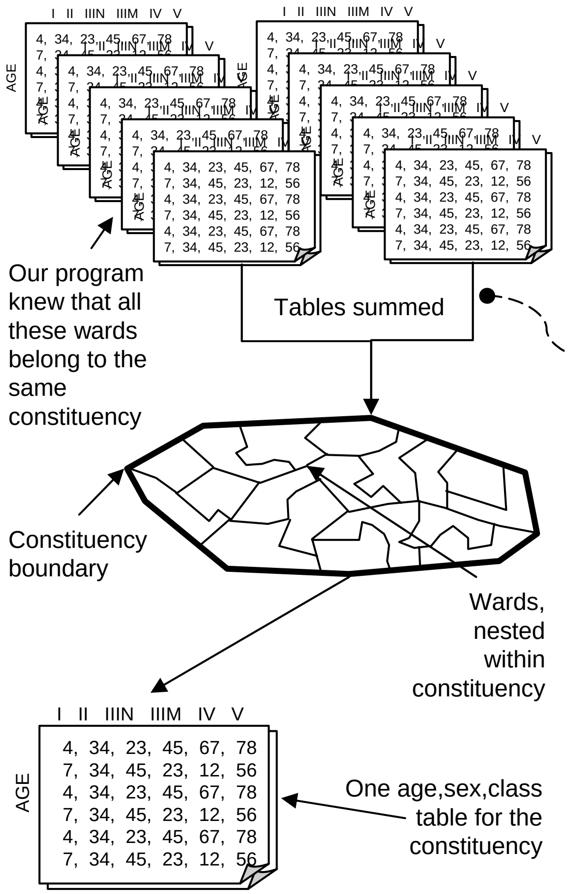
(n) To adjust the 1981 data to 1983, we combined the estimates from 1981 and 1991 together using a simple formula. The number of people in a particular age/sex/class group, in 1983 was estimated as being four times that in 1981, plus that in 1991, divided by 5. The formula has the effect of assuming that the changes which took place between 1991 and 1981, happened at the *same rate* each year. Thus, the size of the change between 1981 and 1982 and that between 1990 and 1991 was assumed to be the same. We literally “froze” that process of change in 1983.

(o) To adjust the 1991 data to 1993 was slightly more difficult, since we did not have the luxury of a data set with identical structure, for 2001. Instead, we were able to utilise ONS estimates of the numbers of people in each age/sex grouping, in each ward, for 1996.

An estimate of the numbers in each age/sex group in 1993 was calculated using the same kind of simple formula described above; again assuming equal change in every year between 1991 and 1996. Then, the difference between the numbers of people in each age/sex group in 1991 and in 1993 were calculated. This gave a number of people who needed to be added or removed from each age/sex group so that the number of people in that group would equal our estimates for 1993. The people were added or removed in proportion to their distribution amongst the classes, so we effectively assumed that the relationship between age/sex and class membership remained unchanged between 1991 and 1993.

For example; suppose the number of men aged 30-34 was estimated to have risen by 20, between 1991 and 1993, and suppose also that the proportion of men aged 30-34, in class II was 25%, in 1991. To adjust our 1991 figures to an estimate for 1993, we would add an extra 5 people to that group (5 is 25% of 20). If the number had fallen by 20 instead of rising, we would remove 5 people from that group.

What Was Done

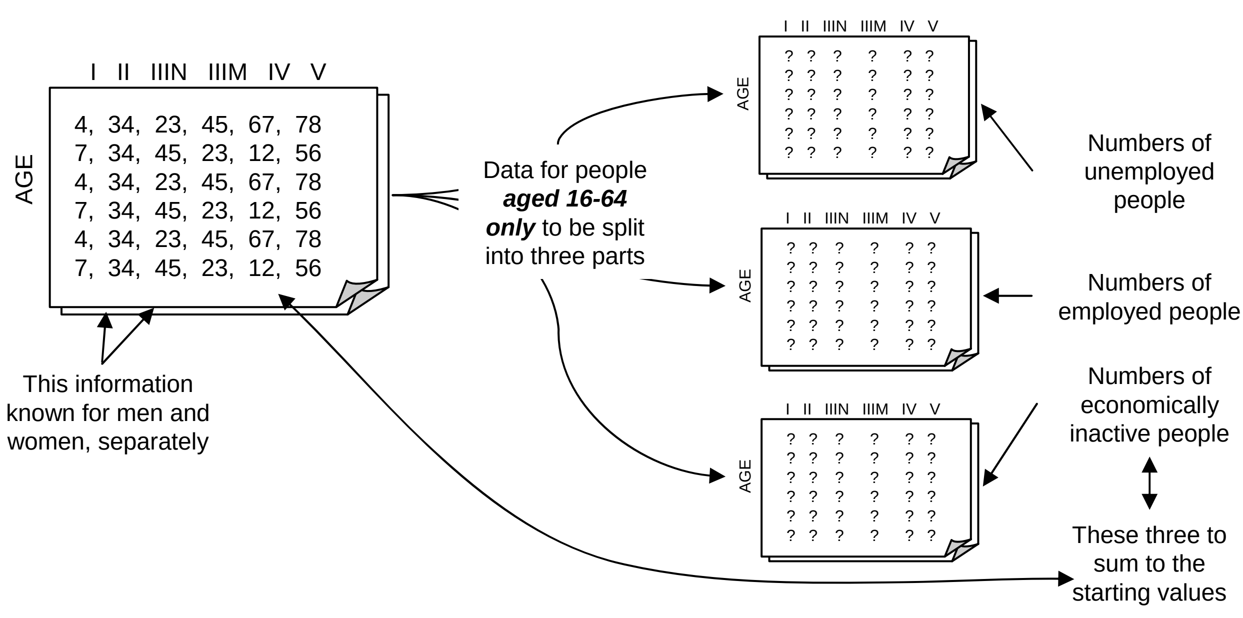


What It Means

(p) The second adjustment was identical for the (now) 1983 and 1993 data. Up to this point, our computer had carried out all of its work at *ward* level. We now aggregated our results to 1997 Parliamentary constituency level, which is the spatial unit with which we wished to analyse patterns of death.There are 641 constituencies in Britain. The computer followed a lookup table which told it which wards belong to the same constituency. We had used the same set of ward boundaries for our 1983 and 1993 data so the wards which belonged together as a constituency in 1983 were the same as those forming the constituency in 1993. The people in each age/sex/class group in wards which belonged to the same constituency were simply added together.

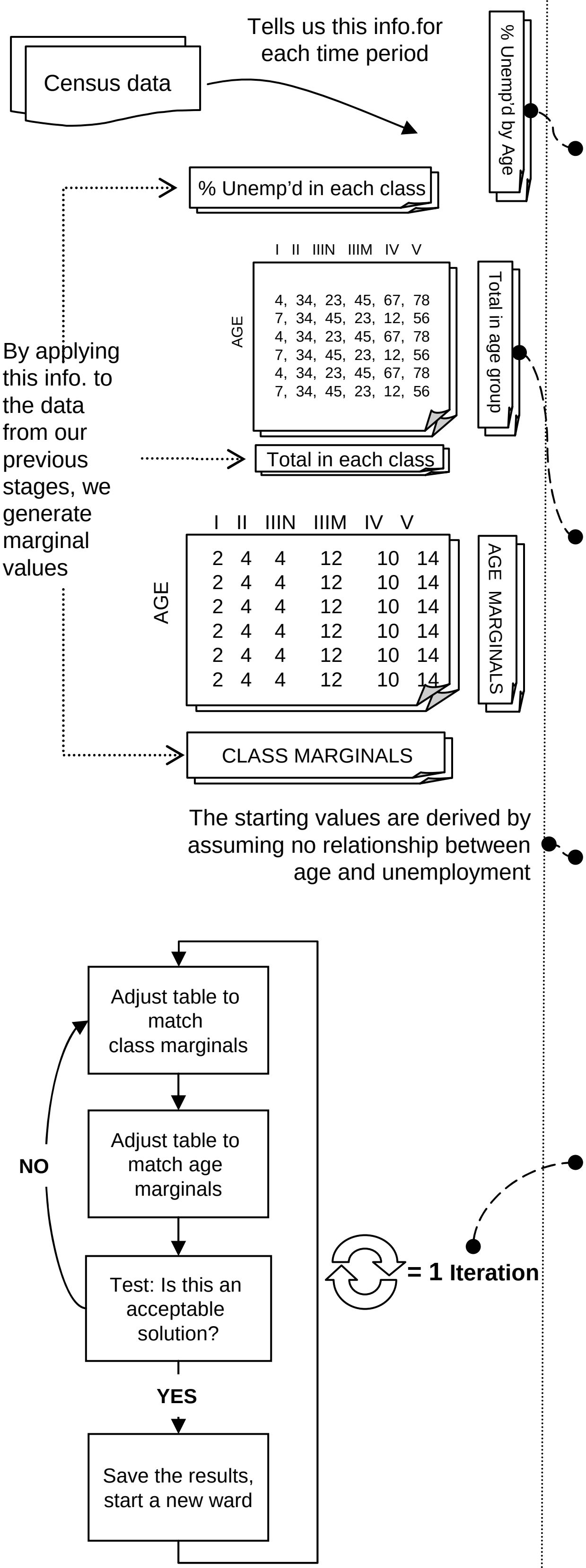
With these adjustments completed, we were ready to create estimates of the numbers in age/sex/class/employment status groups. This process was very similar to that with which we created the age/sex/class estimates. We used the same technique, *iterative proportional fitting (IPF),* and in fact, we re-used some of the computer code we had written before. We will outline the task in the same way as before and then explain how we found a solution.

9. This is the starting point; the picture shows a table, labelled with the data which we had already estimated, and tables which we needed to simulate, for both 1983 and 1993.



For ease of explanation, we will illustrate the first stages of the process just for unemployment. These steps were exactly the same for employment and economic inactivity.

What Was Done



What It Means

(q-i) To estimate the data we needed, we repeated the IPF process. To make this work, we derived two sets of marginals from the census (using the 1981 census for our 1983 data and the 1991 census for our 1993 data). The census data told us the proportion of men / women who were unemployed, employed or economically inactive for each age group, and for each class I-V and we assumed no change in these proportions between 1991 and 1993.

(q-ii) We knew the total number of men / women in each age group, and in each class I-V from our previous work. Our computer code applied the proportions derived in (q-i) to the 1993 and 1983 data to give us *age marginals* and *class marginals* as shown. For example, if the 1991 census data told us that 12% of men aged 20-24 were unemployed, and our estimates showed that in this constituency, in 1993, there were 100 men aged 20-24, the age group marginal value for men aged 20-24 was set at 12.

(q-iii) Before *IPF* could begin we needed a set of first estimates, from which the *IPF* process could proceed. These were derived by assuming that the number of unemployed people in a class was spread evenly across all age groups. So, for example, if the marginal value for men in class II was 30, the starting value for men in each of the 10 age groups was set at 3.

The IPF process proceeded just as before.

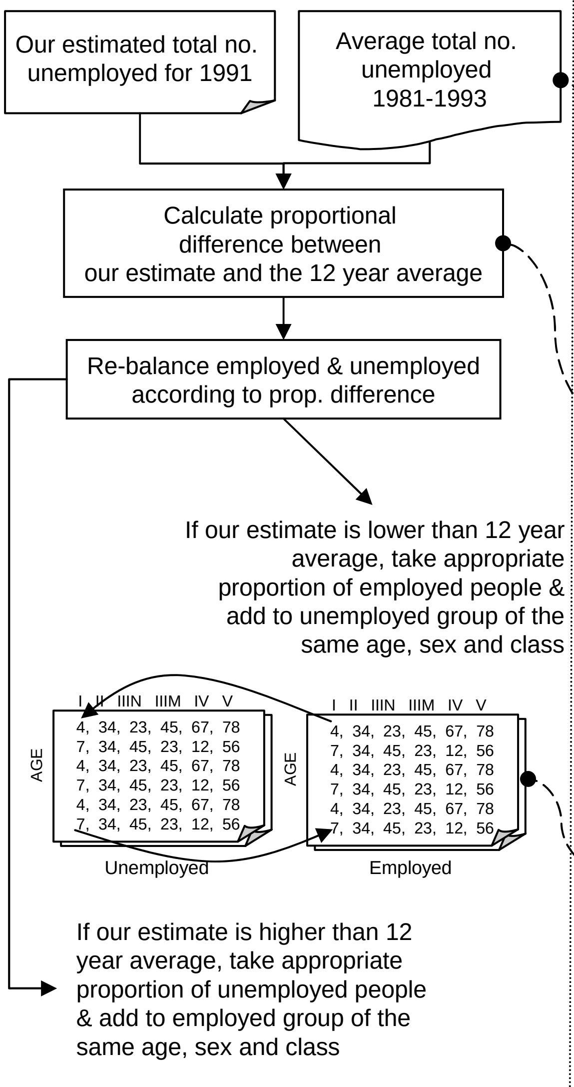
The IPF stopped when the known marginals and estimated totals were less than 0.5% apart. That is 0.005 of a person. Most constituencies converged on a solution after 2 or 3 iterations. The computer program completed the IPF for both men and women, for all constituencies in Britain.

The same process was repeated by the computer program, to produce tables for the employed and the economically inactive. Remember, this process did not change the numbers of people in each age/sex/class group - it simply further divided those aged 16-64 into age/sex/class/employment status groups.

The next stage introduced an adjustment to the figures which was related to the theoretical justification for including employment status in our analysis of deaths in Britain. Research has shown that the detrimental effect of unemployment on an individual's health does not occur after one day, or even one month of unemployment, but after longer spells out of work. The effect may also not have an impact until much later in life, possibly long after the period of unemployment has finished. However, the influence of unemployment on death rates is substantial and it was important that we modelled the numbers of people who might have experienced longer term unemployment in their area, and the change in those numbers over time, as accurately as possible. The great disadvantage of using the census to estimate those numbers is that it is only a snapshot in time. For all we know, all the people in a ward, who said that they were employed on census night in 1981 might have been made redundant the day after the census. These snapshots also occur at particular points in the country's economic cycle. The 1991 census coincided with a particularly deep recession in the south . Had we used it as an indicator of how many people had been unemployed since the 1981 census, our estimates would have been far too high.

We wanted to try and alter our estimates to better reflect the chances of unemployment for people living in each constituency over the 12 years before the census because research on the LS, from which we were going to draw our death rates, uses a 12 year time lag between recorded unemployment and mortality. To do this, we compared the total number of people we had estimated were unemployed, with a preceding 12 year average number, drawn from the figures published by NOMIS;

What Was Done



What It Means

(r) The computer program added up the number of unemployed men and women, of all ages & classes, to give a total number for the constituency. This figure was compared with the average total over the previous 12 years, calculated from the NOMIS figures.

(s) The proportional difference between these figures was calculated. For example, if the 12 year average figure was 3800 and the total estimate for 1993 was 4200, the proportional difference was .906. The value is *less than 1.0* because the number of unemployed in 1993 is too high and needs to be reduced. If the 12 year average figure was 3800 and the total estimate for 1993 was 3000, the proportional difference would have been 1.26. The value is *more than 1.0* because the number of unemployed in 1993 is too low and needs to be increased.

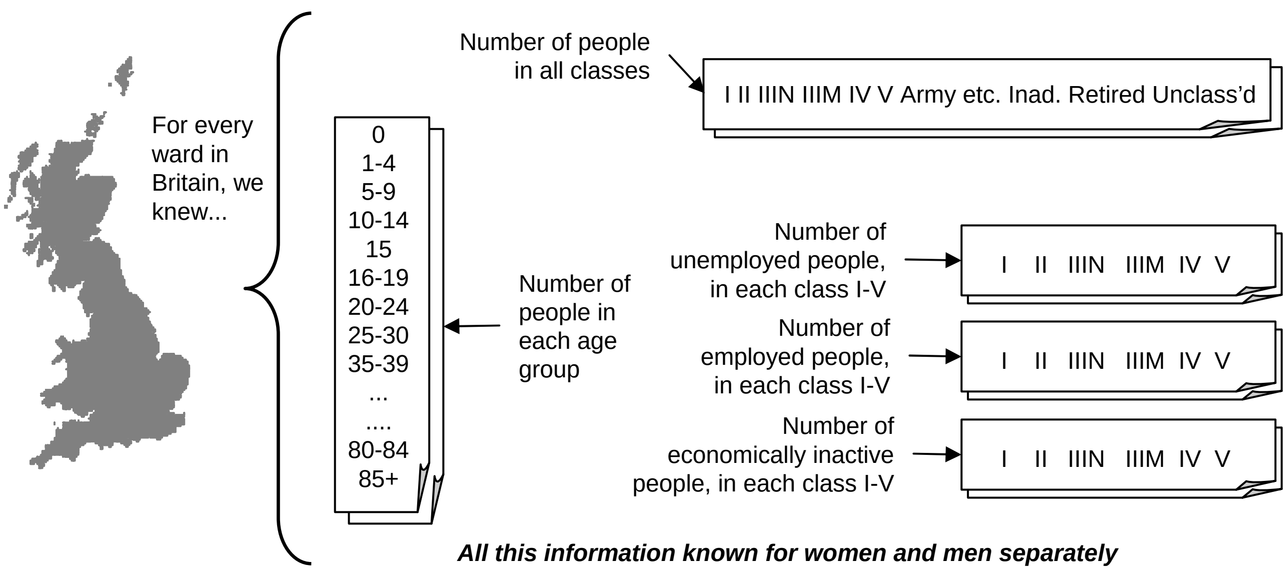
(t) Suppose the proportional difference was 0.90. Each age/sex/class grouping of *unemployed* people was increased by adding enough people to raise the number in the group by 10%. The same number of people were removed from the same age/sex/class group who were *employed.* In this way, the number of people in the constituency stayed the same - people were just swapped between the unemployed groups and the employed groups.

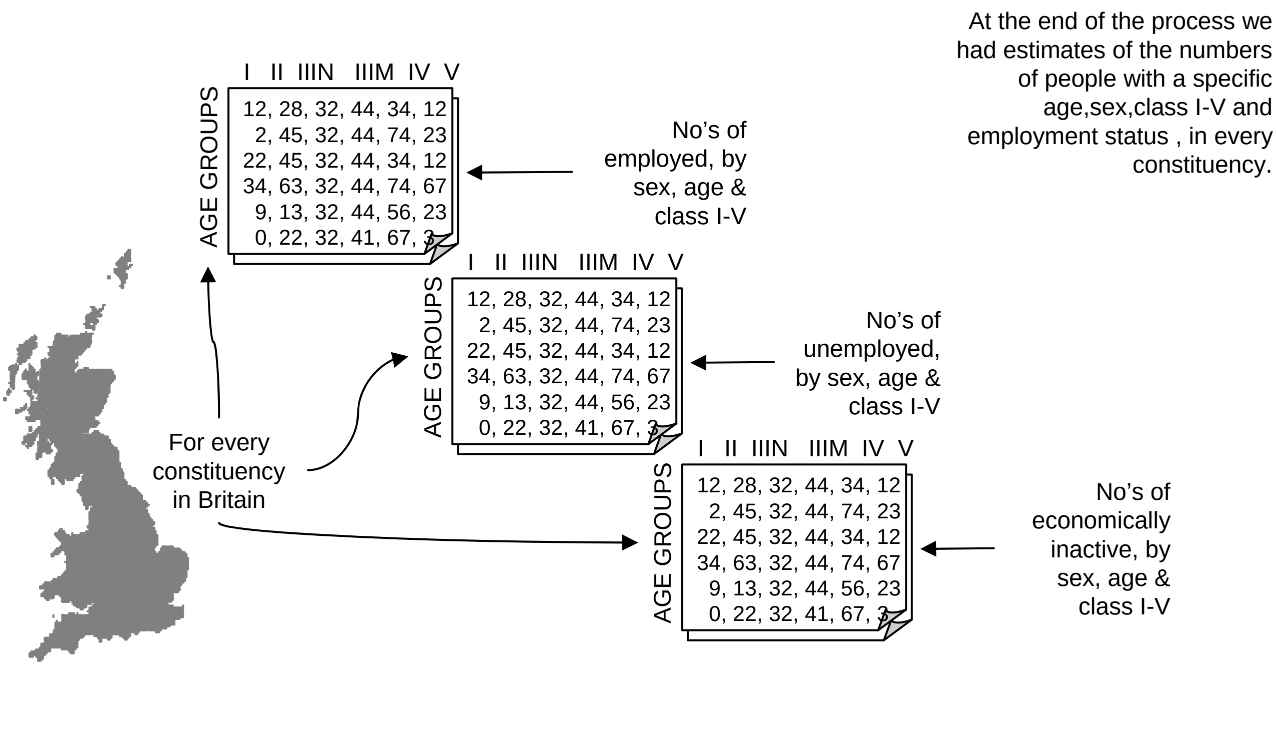
The very last step in this process was to convert the numbers in our age/sex/class/employment status tables to a proportional distribution and apply those proportions to our age/sex/class estimates from the previous section. In plain terms, this means that for each age/sex/class group, we worked out what *proportion* were unemployed, employed or economically inactive and then used those proportions to further divide the groups we had created by stage (p) accordingly. So, if the results from section (q) to section (t) showed that for men, aged 20-24, in class IIIM, 10 were unemployed, 10 were economically inactive and 110 were employed, we took the count for men, aged 20-24, in class IIIM given by the results of section (p) and made 8% unemployed, 8% economically inactive and 84% employed.

Why did we do this, when our estimates of age/sex/class/employment status already represented numbers of real people?

This stage was included because of differences in the way we used *IPF* in stages (e) to (i) and in stage (q). When we created our original age/sex/class counts, we began the process using a known distribution to create the first estimates (see sections (a) to (d)). In this case, we can be very confident that we have obtained plausible estimates of the distribution of people amongst age/sex/classes because we started the *IPF* process from a set of values likely to be quite close to the real distribution. In stage (q) we began the *IPF process without a known distribution* to create the first estimates of age/sex/class/employment status counts. To maximise the quality of our estimates, we therefore treated the results of our work described in sections (q) to (t) as a distribution, rather than as the best estimates of the true *numbers* of people in each age/sex/class/employment status groups

10. Summary - This, quite lengthy, process has moved us between these two situations:



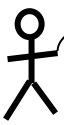


How We Used The Age / Sex / Class / Employment Status Counts To Analyse The Geography of Mortality In Britain And The Changing Power Of Socio-Economic Status To Account For That Geography

11. What did we do with these figures?

The next stage of the research was to use these figures to explore the relationship between characteristics of the population and the numbers of deaths which took place in each constituency. This analysis was based on the relationship between the numbers of deaths which we knew to have taken place in a constituency, and the number which we *expected* to take place.

This ***man*** is aged ***64***



As an individual, some of your personal characteristics are good predictors of your chances of dying. Your *age* and *sex* are the two most commonly used predictors in work of this kind. People of different ages and sexes die, on average, at different rates. If we know the number of people with a given set of characteristics, and the rate at which, on average, people with those characteristics die - we can produce an estimate for the number of deaths amongst that group we would expect if they die at the average rate. For example; if 3 out of every 100 men, aged 64 die per year in Britain and a constituency contains 400 men aged 64, we would expect 12 of them to die in a typical year.

It is usually sufficient to place people in a 5 year age group (i.e. to know that this man is aged between 60 and 64), rather than work with individual years of age when calculating expected deaths.

However, it is not only your age and sex which are good indicators of your chances of dying, your social class and employment status are too. We can (usually) use these extra characteristics to make an even better estimation of the chances of dying than we could just knowing an individual's age and sex. Together, age, sex, social class and employment status constitute a very good base from which to predict the risk of someone dying in a typical year. That is why we spent so much time generating counts of people in each constituency in Britain, with each possible combination of these four characteristics.

This **working class**, **employed** ***man*** is aged ***64***

12. Where did we get the appropriate average death rates from to apply to our data?

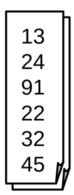
The death rates we used to generate our *expected* numbers of deaths came from a variety of sources. Rates of death based just on knowing a person’s age and sex are published by the ONS. Rates which are based on knowing a person’s age, sex and social class were taken from work using the ONS Longitudinal Study (see section 7). Rates which are based on knowing a person’s age, sex, social class and employment status were calculated by us, using information from the LS and from published tables. Although we refer to these death rates as the “national average”, they are in fact based only on data for England and Wales. This is because the information we need to calculate the figures is not available for Scotland. In fact we know that death rates in Scotland (as a whole) are slightly higher amongst some groups but without comparable data, the best option available was to utilise rates based on the English and Welsh and remain aware of the potential consequences for the results. The effect of doing this will be to make all the findings for Scotland very slightly conservative. It is important to understand that the mortality rate for each group is the *average* for England and Wales together and that this has been presented here as the ‘national’ average.

13. Why have three different sets of death rate, and how were these used?

The significance of having three different sets of death rate, based on differing amounts of information about a person’s personal characteristics is that it allows us to see how important those extra bits of information are in determining the numbers of deaths which take place in a particular constituency. Remember, the deaths rates we use are only the “national” average, and that these may actually differ from place to place *within* Britain. By using different sets of death rate which allow for people’s own characteristics we can see which characteristics are more (or less) important in determining the number of deaths which happen in different places across the country. How do we do that? The best way to explain it is to use an example. Let’s use Sedgefield, in the North East of England.

We are going to follow a step by step example of how these death rates and all the population figures we have calculated are used. We will just look at the figures for ***men*** aged less than 65 in Sedgefield. The actual number of deaths amongst men in Sedgefield, aged under 65, for the five years between 1991 and 1995 was 630 (that is an average of 126 per year). We know this number from the death records which Britain keeps. We tend to work with 5 years worth of deaths because it irons out any peaks caused by things like ‘flu epidemics.

The first step is to obtain an analysis of the deaths, just using the information we have about age and sex.



We already know how many men there are in each age group

... and published research tells us the national average death rate for each age group

Age Group	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	Total
No. men of that age in 1993	574	2199	2866	2879	2726	2809	3251	3049	2903	3151	3102	2617	2345	2181	36652
Number in each group multiplied by national average risk of death over 5 years based on age and sex, to give:															
Expected no. of deaths in that group	20	4	2	3	8	12	14	15	19	30	50	71	111	181	540
Actual No. of Deaths (5yrs)	13	2	1	4	6	16	3	15	22	36	44	100	142	226	630
SMR															117

The number of men in each age group was known and was multiplied by the national average death rate to get an expected number of deaths for the 5-year period, for each group. Expected deaths were then totalled and that number was compared to the actual number which occurred (known as the **observed** number of deaths). So, in Sedgefield in the early 1990s (1991-95) there were actually 630 deaths amongst men aged under 65 but using the national average death rates and the numbers of men in each age group, only 540 were ‘expected’. This gives a Standardised Mortality Ratio (SMR) for Sedgefield's men of 117.

N.B. At this point in the process we had an opportunity to check and correct our calculations. In theory, the total number of **observed** deaths should match the total number of **expected** deaths for Britain as a whole. When we worked out our expected number of deaths for each constituency, these could be summed to a national total and checked against the known values. We typically found a 1-2% difference between the totals. To make the results as accurate as possible, the counts of expected deaths for each age / sex group were corrected so that they summed precisely to the nation total of observed deaths.

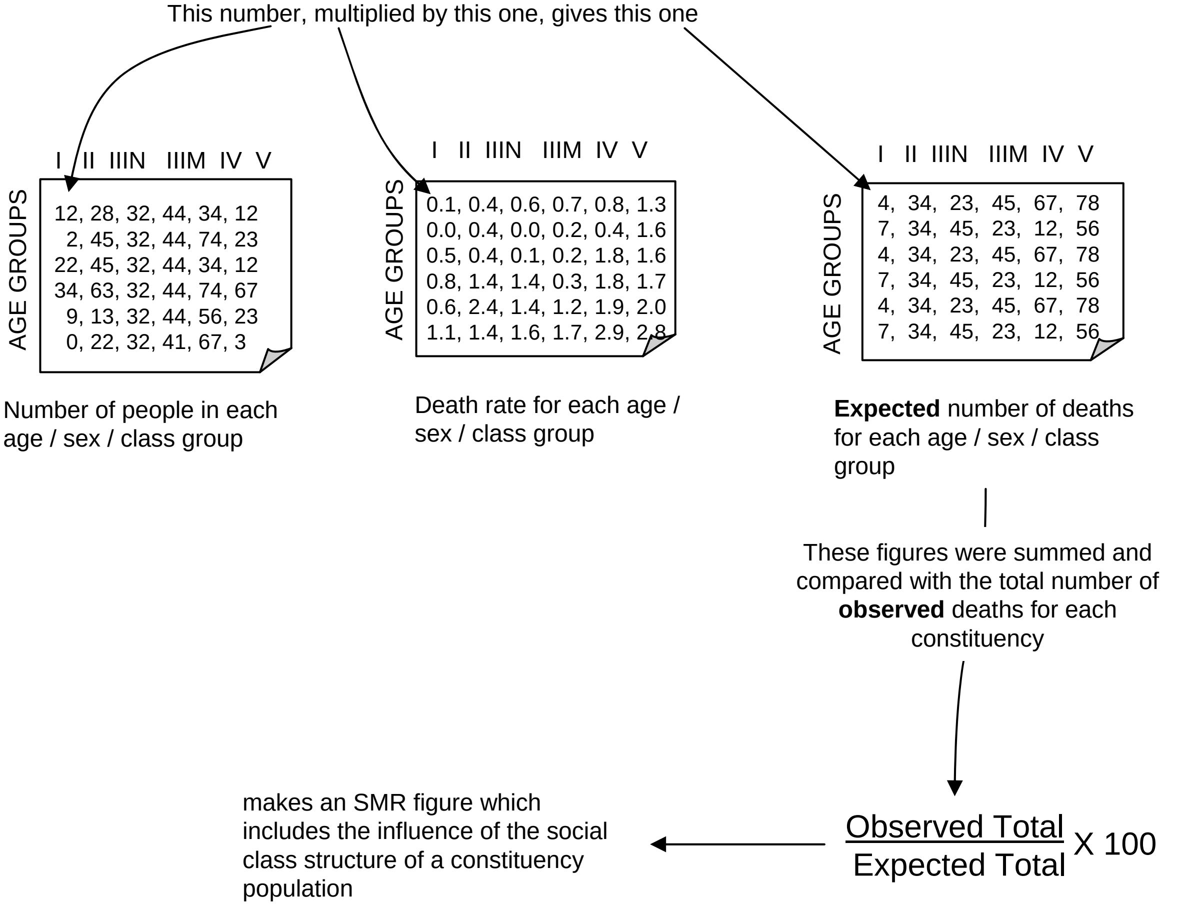
14. What Does SMR Mean?

SMRs are a measure of the population's chance of dying each year, relative to the national average. They allow the mortality rates in different areas with different population sizes and compositions to be directly compared. An SMR below 100 indicates that the area has mortality which is relatively low compared to the national average, whereas areas with SMRs greater than 100 have relatively high mortality. An SMR of 200, for example, would mean that the population experienced mortality at rates twice that which was normal for the nation as a whole. An SMR of 50 would mean that the population experienced mortality at rates half that of the national average. The figure of 117 means that Sedgefield's men have a risk of death at ages below 65 which is 17% higher than the national average. To calculate the SMR you need to know the **expected number** of deaths and the **observed** number of deaths. The formula is easy:

$$\frac{\text{Observed Total}}{\text{Expected Total}} \times 100$$

15. Adding Social Class To This Model

To include information about the social class composition of Sedgefield's men, the process was almost exactly the same as before, but the number of groups was increased so that there was one for every possible combination of age and social class. For example, Sedgefield's 3049 30-34 year old men are made up of 153 in class I, 718 in class II, 231 in class IIIN, 1221 in class IIIM, in 602 class IV and in 123 class V. A slightly different death rate was applied to each of these groups because, on average, the risk of death increases with lower social class. Otherwise, the process was exactly the same as before, and of course the *observed* number of deaths was the same. When class information was added to the model, the expected number of death rises to 551. This is because the calculation was better able to take account of the higher death rates amongst Sedgefield men due to relatively high numbers of them being working class. The SMR (adjusted for class) now came down to 114.



16. Adding Employment Status to the model

To include information about employment status, the number of groups was increased again so that there was one for every possible combination of age, social class and employment status. A slightly different death rate was applied to each group, reflecting the fact that the national average death rate is higher for the unemployed than the employed – regardless of social class (although interacting with it somewhat). The difference in mortality rates between the employed and unemployed varies with social class and age.

how we calculated the death rates here DANNY!!

When employment status information was added to the model for Sedgefield's men aged under 65, the expected number of deaths rose to 581, making an SMR (adjusted for class and employment) of 108.

This means that even after allowing for the employment and class structure of this population, their chances of death are still 8% higher than the national average. On the other hand, without the information about class and employment status, those chances of death would appear to be 17% higher – a substantial amount of the apparently high death rate has been 'accounted for' by understanding the class and employment characteristics of Sedgefield's men. The first rate (17% higher than the national average) is not excused or justified – the results mean most of that excess is due to the inequalities which persist between differing social classes and between those in and out of work.

17. So What? How Are These Figures Helpful?

When these figures are calculated at two points in time (for the early 1980s and early 1990s), the relationship between them becomes interesting because it tells us about how the power of information about social class and employment status to account for differences between the observed and expected numbers of deaths, has changed. Here is a table continuing the example for Sedgefield, presenting figures for the 1980s as well as those already explained for the 1990s

Time Period	1980s	1990s	Change
SMR based on age and sex	120	117	-3
SMR based on age, sex & social class	117	114	-3
SMR based on age, sex, social class and employment status	113	108	-5
Amount accounted for by social class	3	3	0
Amount accounted for by social class and employment status	7	9	+2

As can be seen from these figures, there was a slight fall in the relative (to the national average) mortality rates for men in Sedgefield aged less than 65, between the early 1980s and the early 1990s. When social class and employment status are taken into account, the SMR has fallen more substantially. This tells us that socio-economic position has become *more* important in determining Sedgefield's mortality rates, during the decade in question.

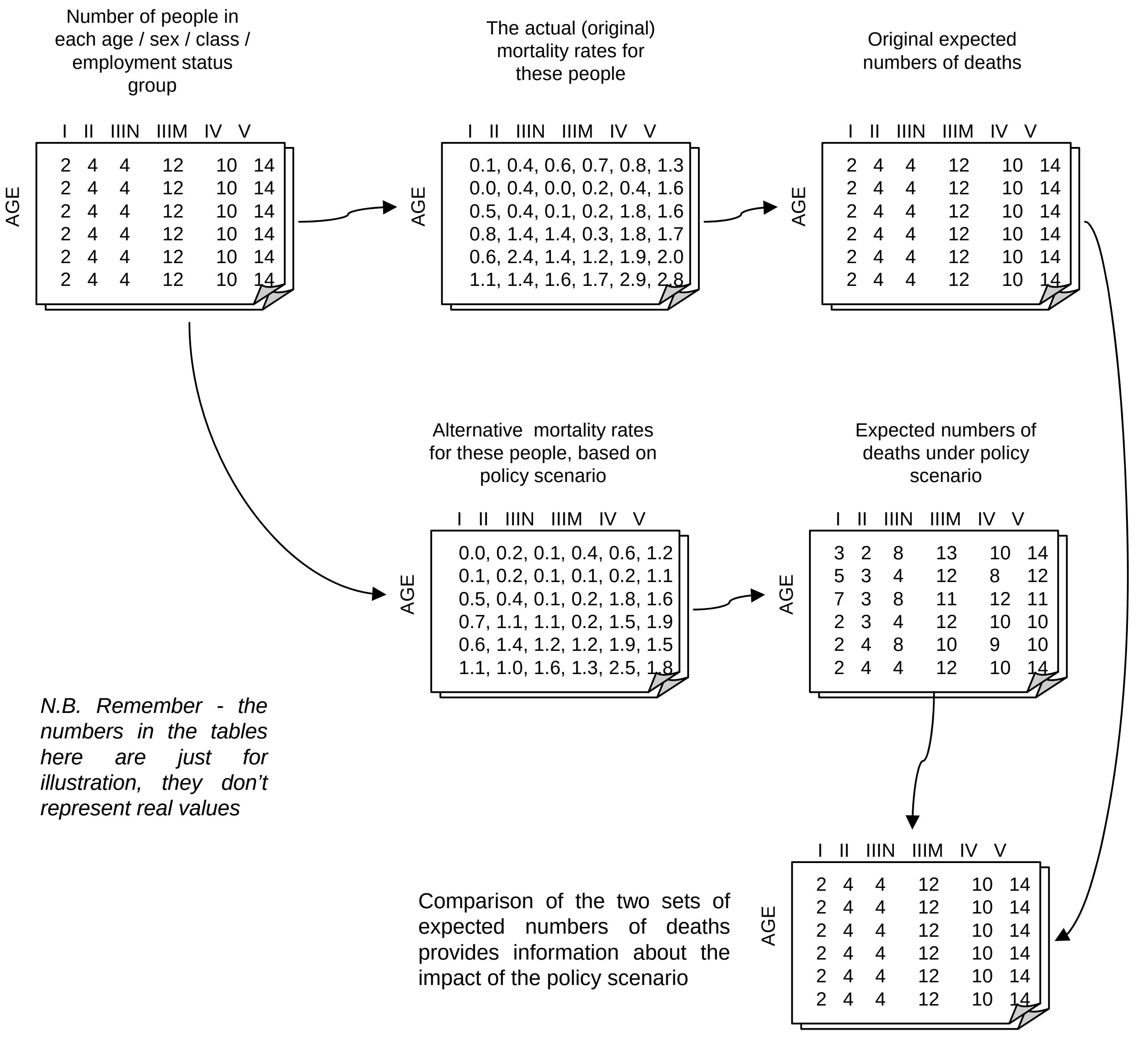
The first aim of the project was to find out how much of the geographical variation in mortality rates across **Britain** could be explained through understanding how the employment status and social class of the population varied. The simple steps outlined in sections 12 - 16 have shown how this was achieved for one constituency. Achieving it for the whole country was simply a matter of identifying all the deaths which either took place or did not occur when they were expected to do so, using models which take account of age and sex, then age, sex and social class and finally, age, sex, social class and employment status. The relationships between these sets of figures give rise to the analyses presented in Chapter 3 of the main report.

How These Data and Techniques Were Used To Simulate The Effects of Different Social Policies

The main report explains the rational for doing this kind of analysis, and that will not be repeated here. The rest of this report explains how the analyses were achieved, from a technical point of view. Three policies were tested; the attainment of full employment, a modest redistribution of income and the erradication of child poverty. Each policy was only tested for its impact on the 1990s data, and only for impact on the population aged under 65.

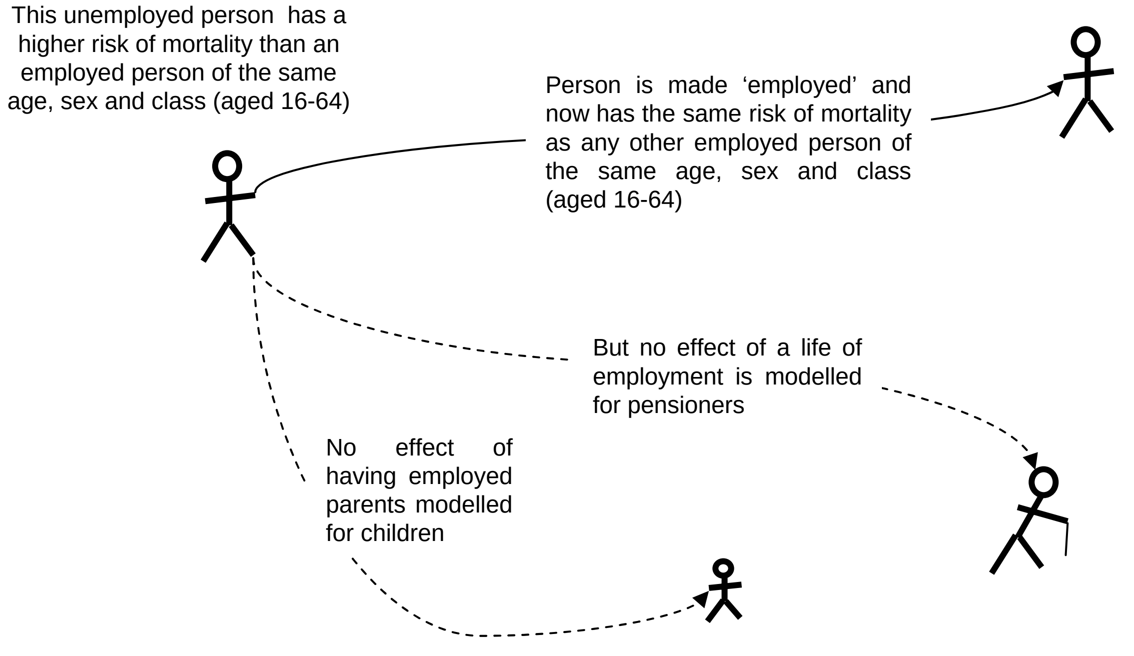
18. How, In Principle, Were The Policies Tested?

The principle behind the policy testing was very simple. Each policy scenario refers to changes in the mortality rates to which some, or all of the population are exposed. In the first stages of the work, for example, we were careful to apply appropriately higher mortality rates to those people who were unemployed (See section 9, part q). In the proposed scenario where full employment is attained, we modelled and measured the effect on mortality by adjusting the death rates to which the unemploye people were exposed - in simple terms, everyone who was unemployed was given a job (or in fact, the chances of dying equivalent to an employed person of the same age, sex and social class).

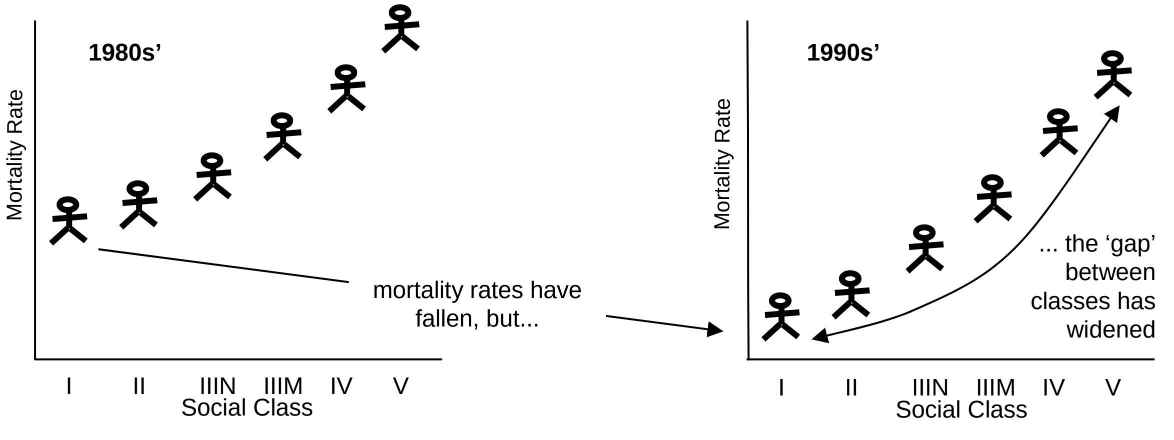


19. Where, in practice, did the alternative mortality rates come from?

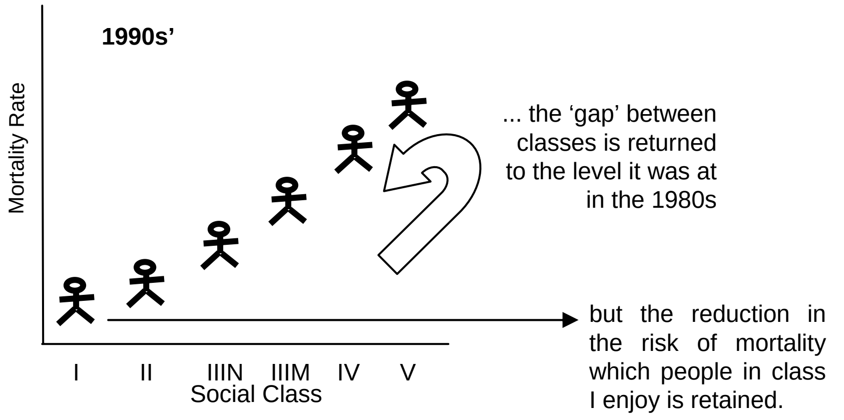
a) Section 18 described briefly the simple adjustment needed to model the effects of full employment. Although this may seem rather crude it does represent an accurate depiction of the policy scenario. Much of the higher risk of mortality associated with unemployment stems from being out of work in the long-term. Full employment is generally agreed to be a situation in which there are jobs available for all, and whilst some people might be between jobs (and technically unemployed), no one should be out of work in the long term. Also, note that the adjustment was carried out only for men and women between the ages of 16 and 64 since one cannot be unemployed before 16, and it is unusual following retirement (for convenience, and due to likely harmonisation in the future, women's age of retirement has been raised to mirror men's). Thus, the models do not include any impact of full employment on the mortality rates of children or the retired - even though in the longer term there would almost certainly be a corresponding improvement in the life chances of children supported by employed parents (where before they were supported by unemployed parents and the state) and in the life chances of the retired now able to benefit from extra savings, pensions and having acquired a better health trajectory in earlier (employed) life.



b) To model the effects of a modest redistribution of income an assumption is made that the redistribution could reduce inequality in mortality rates between the social classes to those levels experienced in the early 1980s. The picture shows how the mortality 'gap' betwen the classes got wider between the early 1980s and early 1990s, even though the chances of dying fell overall.

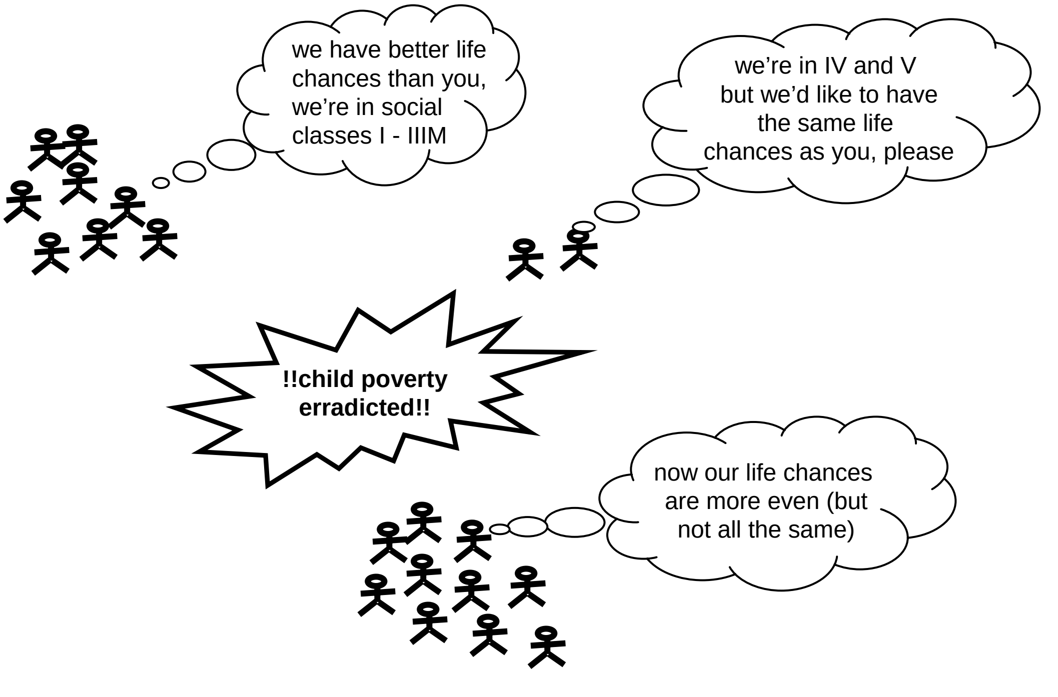


To operationalise this idea, the mortality rates of people in social class I were left unchanged at their 1990s level but the differences in mortality rates between those enjoyed by class I and those in other classes were then fixed at the level they were in the early 1980s. This approach combines the overall rise in standards of living between the 1980s and 1990s, with a reduction in the differential impact of that rise on different classes.



In more technical terms, the mortality rates for people in classes II - V were calculated by applying the 1980s ratio between class I and each other class to the 1990s mortality rate for class I. For example, if in the1980s the ratio between class I and V was 1:3 (i.e. people in class V were 3 times as likely to die in a given year), then the 1990s rate for class V was set at 3 times that of the 1990s rate for class I. This change applied to everyone in the population, apart from those in class I. Distinctions between those in different categories of employment status were made by fixing the ratio's between unemployed, employed and economically inactive mortality rates. Thus, the relative differences between mortality rates for people in different employment circumstances were the same, even though the actual mortality rates were lower.

c. The effect of ending childhood poverty was modelled by giving children in social classes IV and V the same good chances of health which those children in 'higher' classes. To do this, the poorest 20% of children were reassigned to higher social classes on a pro-rata basis and thus given a lower risk of death. In this instance, pro-rata means that if 15% of (class I to IIIM) children in a constituency were in class II, and 100 children were in classes IV and V, 15 of them were 'elevated' to enjoy the risk of mortality associated with being in class II (N.B. this report actually uses a smaller group of children in poverty than the government's own estimates). Although this was a somewhat crude procedure it did realistically reproduce the aims of eradicating childhood poverty: to bring the life chances of the poorest in line with those of the better off.



It is important to understand that, because the model moves only children in classes IV and V, on a pro-rata basis, it does not give *all* children better life chances, or all of them the *same* risk of dying, it simply moves those children at the lowest end of the social class scale (and thus those most likely to be in poverty) into classes where their risk of death maybe lower. For some children that move will have made only a small change to their risk of mortality. Only children aged 0-14 were affected by this change and as before, only the effect on children that age was measured.

When modelling each of the three policy scenarios, mortality rates appropriate to the age, sex, and where appropriate class and employment status of each person were adjusted, therefore retaining as much accuracy as possible. The figures in the main report refer to the absolute number of and proportion of deaths prevented by each policy. Figures are also provided for the whole of Britain and just those constituencyies in which mortality rates are above the national average.

That concludes the technical report. If there are any questions left unaswered by either this report, or the main report, please email us inequalities@social-medicine.com for clarification