Randomisation by cluster and the problem of social class bias

FREDA ALEXANDER,1* M MAUREEN ROBERTS,2 WALTER LUTZ,1 AND WILMA HEPBURN.1

From 1The Medical Statistics Unit, University of Edinburgh; and 2The Edinburgh Breast Screening Clinic.

ABSTRACT For randomised population trials the unit of randomisation is normally the individual person. In some situations, however, investigators take other groups as basic unit and one such design is cluster randomisation. Considerable attention has been given to this design recently in statistical and epidemiological literature. The Edinburgh randomised trial of breast cancer screening is an example which takes general practices as clusters of patients. The experience of this trial is reported here. Mortality from all causes, cardiovascular disease and lung cancer over the first 5 year period of follow up are examined. We found that spurious mortality differences were present in the analyses, which do not allow for socio-economic status. From examination of methods of adjusting for this, we conclude that allowance can be made at the analysis stage, and it is intended that this approach will be adopted when breast cancer mortality is analysed in the Edinburgh trial. Nevertheless, we recommend that for future studies with outcome related to socio-economic status, randomisations which use this design be stratified by socio-economic criteria where this is feasible.

For controlled trials statistical considerations suggest individual randomisation as the method of choice. There are some situations, however, where it is desirable or even essential to take a group or “cluster” as the basic unit. The statistical effect of such sampling is to weaken the efficiency of the randomisation since the effective sample size is considerably smaller. Thus unanticipated differences are more likely to arise between the randomised groups, and these, if undetected, could bias the results.

The use of cluster randomisation as a design for non-therapeutic intervention trials has received considerable attention recently in the epidemiological1 2 and statistical3-7 literature. This work has been concentrated in the USA and relates primarily to cardiovascular disease. The main thrust of the arguments has been that these designs are valid, but that the effective sample size is reduced and that any analysis must take account of between cluster variation, otherwise spurious significance may result. At the outset Cornfield3 stated that “randomisation by cluster accompanied by an analysis appropriate to randomisation by group is an exercise in self deception . . .”. Less attention has been paid to the possibility of bias in the randomisation though Donner and Donald7 discuss a design with matching of background factors before randomising so as to avoid this problem. The aim of this study is to examine the experience of one UK trial which adopted cluster randomisation as its design.

The Edinburgh randomised trial of breast cancer screening8 is based on randomisation of the general practices. Five years of follow up have been completed. Breast cancer mortality will be compared between the screened and control populations for the first time when 7 years of follow up have been completed. Meanwhile it has been shown9 that there is a statistically significant relationship between the socio-economic status of the GP practice and all cause mortality amongst its women. This study begins by examining how unexpected differences in the socio-economic status of the GPs in the screening and control arms of the trial can affect comparisons of mortality from broad disease categories. Methods of adjusting for these effects at the analysis stage are then considered. There has been no examination of breast cancer mortality because it was agreed8 that the first such analysis would be after 7 years of follow up.

Methods

Most women between the ages of 45 and 64 who were resident in Edinburgh between 1978 and 1981 have been enrolled into the study population of the Edinburgh randomised trial, and comprise the “initial
cohort" of the trial. Eighty seven of the city's general
drives are participating in the study and women in
these practices and of the appropriate age were
enrolled, one practice at a time, over the 4 year period.
The total number of women involved is 45 130. Of
these, 23 227 women from 43 general practices are
being offered regular mammographic screening over 7
years. The remaining 21 903 women from 44 general
practices form the control population. The 87 general
practices were randomised to screening and control
status respectively.

The randomised design was the responsibility of one
of us (WL). It was based on a number of assumptions
which we now describe.

It was not practical to use individual randomisation
because (a) there was no adequate list of women
to use as a sampling frame either at the
practices or at the local health board; (b) doctors
would not accept a scheme whereby some of their
patients were offered screening and others not;
(c) contamination of the two arms could arise if
control women became aware of their acquaintances
having been screened and insisted on screening for
themselves. Sampling by practice was favoured
because it was simple and overcame these difficulties.
However stratification by number of partners in the
general practices was introduced for two reasons.
(1) There would be greater statistical efficiency if the
two arms of the trial contained approximately the
same number of women; the number of patients was
assumed to correlate well with the number of partners.
(2) There was concern that the larger group practices
might be associated with a different type of patient.
Stratification would avoid any consequent bias.

Socio-economic factors were known to be relevant
but there seemed to be no practical way of taking this
into account other than by relying on the
randomisation. The possibility of using some form of
"matching" for social area or other simple criteria was
considered but rejected as it was believed that practices
tended to have widespread catchment areas. The
practice records and the local health board provided
no data on socio-economic status. Neither small area
statistics nor the post code directory were available.
Even if they had been they could not have been used
without a listing of individual women, and so it would
have been necessary to delay the start of screening
until all the women in all participating practices had
been listed.

Initially the trial was planned for South Edinburgh.
The GP randomisation was carried out by one of us
(WH). After excluding practices which had already
participated in an earlier breast screening project and
two practices whose lists were restricted, and including
as a screening practice one which had been used to
pilot the intervention, the remaining 45 practices were
listed in strata according to the number of partners.
Then, using a table of random numbers, approximately a third of each stratum was selected as
follows: one doctor practices, 7/20; two doctor
practices, 3/11; three doctor practices, 2/6; four doctor
practices, 1/3; and five–six doctor practices, 2/5. These
were allocated the status of screening practices.

At the time of randomisation two adjustments were
made: practices subsequently found to be operating
from the same building were allocated the same status;
thus in at least two cases re-allocation of status was
made with a screening practice and a control practice
from the same size stratum exchanging status. The
practices were all visited (MMR) to discuss their
status, and control practices in two buildings were
inadvertently told that they had been selected for
screening. The new status was allowed to stand. One
practice selected for screening refused to participate in
the trial. In all these cases the changes occurred before
any women entered the study.

In order to enlarge the study a similar procedure was
used to select 16/44 practices in North Edinburgh for
screening in 1980. The practices were entered into the study one at a
time during the period 1979–81. As the practice was
entered, all women on the list of the appropriate age
were entered. Thereafter, each practice was reviewed
annually and women newly on its list or who had
recently attained the age of 45 were entered. In all cases,
each woman took her survey entry date or "date of
randomisation" from the date when her practice
entered the study or was reviewed.

A difficulty we encountered was that GP practices are
not fixed entities. They split and amalgamate as doctors
retire. There was a time lag between the decision to enter
a practice and the completion of its register. In one case a
control practice ceased to exist during this period. It had
been amalgamated with a practice already entered as a
screening practice and so the women on its list entered
the screening arm of the trial.

Amalgamation and splitting of practices which occurred
after they had entered the trial is ignored for the
purpose of the trial. "Practices" as they existed at the
relevant time in the period 1979–81 are allocated
study numbers and each woman belongs to one such
"practice". For the most part these are of course
actual general practices.

At the beginning of 1981, after all practices in North
and South Edinburgh had been allocated their status,
it was decided that the screening arm of the trial was
too small. The control practices were stratified as
before and three from the largest stratum randomly
selected. For them their entire trial history to that date
and that of the women on their list was quashed. They
entered the study in the normal way as screening
practices in late 1981.
Randomisation by cluster and the problem of social class bias

Mortality data of women in the study population are being collected in two ways: the Vital Statistics Branch of the General Register Office for Scotland provide weekly and quarterly lists of all deaths in the Lothian Health Board region and these are matched against the study population register. Also all women have been flagged with the Scottish General Registry Office NHS Central Register for Scotland. Both these sources provide cause of death information obtained from death certificates.

The aim of the trial is to determine whether screening reduces breast cancer mortality. To have adequate statistical power it was decided\(^8\) to make the first comparison of breast cancer mortality when all women in the initial cohort of the study had had 7 years of follow up. The present study specifically excludes any examination of breast cancer mortality. It examines mortality from other causes during the first five years of follow up which were completed on 31 December 1986. We have considered all cause mortality and mortality from two broad disease categories: cardiovascular disease (ICD 390–458), and lung cancer (ICD 160–163). These were selected because we suspected a socio-economic status disparity between the two arms of the trial; there is continuing evidence (eg \(^9\)) of differences in mortality from these causes between various socio-economic groups. The cause of death was taken to be the primary cause recorded on the death certificate. For each cause, mortality was expressed as rates/1000 women years at risk. All analyses were stratified into four age groups: 45–49, 50–54, 55–59 and 60–64 at randomisation.

Small area statistics based on 1981 census data first became available in 1983 and these have been used\(^9\) to assign socio-economic indices to 78 of the 87 participating practices. The remaining nine practices were excluded because they had less than 1000 women years of follow up. Of the 78 practices, 37 were screening and 41 control practices. The list of a general practitioner is of course not a precise geographical or census area, so the indices are essentially weighted averages of proportions available from census data. The factors which were included are shown in the first column of table 1. Principal component analysis\(^11\) was used to derive the linear combination of these indices which best summarises the data. The coefficients are shown in the second column of table 1. These enable a single number to be attached to each general practice which is a measure of the socio-economic status of the people and especially of the women on its list. We shall refer to this as its socio-economic status. The 78 practices were divided into three groups of 26: those with the highest, medium and lowest socio-economic status. These are called socio-economic categories. Our analyses are restricted to these 78 practices.

### Table 1 The socio-economic indices

<table>
<thead>
<tr>
<th>Index(^1)</th>
<th>Coefficient(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social class I,II (percentage of households)</td>
<td>-0.95865</td>
</tr>
<tr>
<td>Social class IV, V (percentage of households)</td>
<td>0.97132</td>
</tr>
<tr>
<td>Social class I,II (percentage of economically active married females)</td>
<td>-0.89006</td>
</tr>
<tr>
<td>Social class IV, V (percentage of economically active married females)</td>
<td>0.93045</td>
</tr>
<tr>
<td>Social class I,II (percentage of economically active SWD females)</td>
<td>-0.88606</td>
</tr>
<tr>
<td>Social class IV, V (percentage of economically active SWD females)</td>
<td>0.90087</td>
</tr>
<tr>
<td>No car in household (percentage of households)</td>
<td>0.91061</td>
</tr>
<tr>
<td>&gt;1.5 persons/room (percentage of households)</td>
<td>0.97045</td>
</tr>
<tr>
<td>&lt;1.0 persons/room (percentage of households)</td>
<td>-0.97806</td>
</tr>
<tr>
<td>Total unemployment (percentage of economically active persons)</td>
<td>0.84836</td>
</tr>
</tbody>
</table>

SWD = single, widowed or divorced

\(^1\) Note that these are not exact proportions for each GP practice but are weighted averages of proportions for census enumeration districts represented in the practices.

\(^2\) The coefficients are used to derive a linear combination of the different parameter values for each general practice. This is its “socio-economic status”.

Mortality is analysed by logistic regression, modified by the method of Williams\(^12\) to permit extra binomial variation. The hope was that although the randomisation was by general practice the result would be essentially equivalent to individual randomisation and could be analysed as such. Then mortality rates in the screened population as a whole are compared with those in the control population. An alternative, and for this design technically more correct, approach is to derive mortality rates in each of the general practices and then to compare these. These two are reported here and referred to as individual and cluster randomisation. Each is then repeated with allowance made for social class: more precisely, the personal randomisation method of analysis is stratified by socio-economic category and the cluster randomisation method uses socio-economic status as a continuous covariate. Mortality differences are expressed as a ratio. This is the ratio of mortality rates in the control population to those in the screened population calculated under the assumption that this ratio is not dependent on the covariates (see \(^13\)).

Any difference in all cause mortality or mortality from the three broad categories must be a chance result or the consequence of differences between the two arms of the trial other than the intervention. Our aim has been to see whether this does occur and if so, whether the methods described above correct the results.

### Results

The numbers of women years in the 5 year period for each arm of the trial and each socio-economic category are shown in table 2. They confirm our impression that there are substantial differences between the two, with the screening population...
containing many women in the higher socio-economic status. On the other hand table 3 shows that the age distribution in the two populations is almost identical.

The numbers of deaths over the same 5 year period are given in table 4. The numbers are higher in the control population for all cause mortality and for each of the disease categories; after allowing for the difference in population size the total deaths in the control population exceed those in the screening population by 20%.

**Table 2** Women-years over the 5 year period (%)

<table>
<thead>
<tr>
<th>Population</th>
<th>Highest</th>
<th>Medium</th>
<th>Lowest</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>58,444</td>
<td>23,725</td>
<td>29,177</td>
<td>111,346</td>
</tr>
<tr>
<td></td>
<td>(52.5)</td>
<td>(21.3)</td>
<td>(26.2)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>15,409</td>
<td>41,398</td>
<td>44,038</td>
<td>100,845</td>
</tr>
<tr>
<td></td>
<td>(5.3)</td>
<td>(41)</td>
<td>(44.7)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3** Age at survey entry date (%)

<table>
<thead>
<tr>
<th></th>
<th>45-49</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>25.5</td>
<td>25.8</td>
<td>26.4</td>
<td>22.3</td>
</tr>
<tr>
<td>Control</td>
<td>26.5</td>
<td>24.4</td>
<td>26.7</td>
<td>22.4</td>
</tr>
</tbody>
</table>

**Table 4** Number of deaths over the 5 year period

<table>
<thead>
<tr>
<th>Cause of death</th>
<th>Population</th>
<th>All cause</th>
<th>Cardiovascular disease</th>
<th>Lung cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>814</td>
<td>304</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>921</td>
<td>370</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>Difference (%)</td>
<td>20</td>
<td>25</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>

The difference is calculated after allowing for difference in population size.

**Table 5** Analyses of population differences for all cause mortality and mortality from cardiovascular disease and lung cancer.

<table>
<thead>
<tr>
<th>Cause of death</th>
<th>Method of analysis</th>
<th>Estimate of rate ratio (95% confidence interval)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All causes</td>
<td>Individual randomisation</td>
<td>1.26 (1.14-1.38)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.26 (1.11-1.38)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Individual randomisation</td>
<td>1.35 (1.16-1.57)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.35 (1.16-1.57)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Individual randomisation</td>
<td>1.29 (1.02-1.39)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.19 (1.01-1.39)</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

The mortality data were formally analysed with stratification only by age. This is how we would have anticipated analysing breast cancer mortality. The results are shown in table 5. The differences between the two populations are statistically highly significant for all cause mortality and for cardiovascular disease mortality. Those for lung cancer just attain statistical significance. The two methods of analysis yield almost identical results.

Table 6 gives the mortality rates for each socio-economic category, and the figure shows the rates for each practice plotted against socio-economic status. These show a trend within each population towards higher mortality rates for lower socio-economic status. The formal analysis was therefore repeated with allowance for socio-economic category and status. The results are shown in table 7. For the individual randomisation method statistically significant differences between the populations remain, though the estimate of the rate ratios are considerably reduced. With cluster randomisation method the differences are no longer statistically significant.
Randomisation by cluster and the problem of social class bias

All-cause mortality

Cardiovascular disease

Lung cancer

Figure  Mortality rates/10^5 women years over 5 years by GP practice against socio-economic status

Table 7 Analyses of population differences with adjustment for socio-economic factors.

<table>
<thead>
<tr>
<th>Cause of death</th>
<th>Method of analysis</th>
<th>Estimate of rate ratio (95% confidence interval)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All causes</td>
<td>Individual randomisation</td>
<td>1.14 (1.01-1.27)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.11 (1.00-1.22)</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Individual randomisation</td>
<td>1.21 (1.02-1.42)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.14 (0.95-1.36)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Individual randomisation</td>
<td>1.11 (0.82-1.50)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td></td>
<td>Cluster randomisation</td>
<td>1.07 (0.79-1.44)</td>
<td>&gt;0.1</td>
</tr>
</tbody>
</table>

Discussion

The study population of the Edinburgh trial contains over 40,000 individual women and at this analysis over 200,000 women years. Nevertheless the randomisation based on general practices has relied on a small number of randomisation units. The screening population has been selected by cluster sampling with each GP practice a single cluster. To be efficient, cluster sampling requires that each cluster should be
practices. Though general practice lists are certainly heterogeneous to some degree, it has been shown that socio-economic differences between them can be identified using census data. We have shown here that there are substantial socio-economic status differences between the two arms of the Edinburgh trial. It may have been peculiarly unfortunate or alternatively there may be aspects of its design which are partly responsible; for this reason we have provided full details of the randomisation. The inference is that an apparently large trial can, if randomised in this way, have important differences between the randomised groups.

Randomisation of general practices is the method of choice for studies where the subjects are general practices. This is not our present concern. We have endeavoured to compile a list of all UK studies with general practices as sampling units, where the target population is all persons (satisfying certain criteria) on the GP lists. Despite having assistance from the Association of University Teachers of General Practice we have found only a small number in addition to our own. These are a survey of cardiovascular disease, and studies of the care of pre-terminally ill patients with solid cancers, of promoting heart disease prevention in primary care, and of the efficacy of common drugs in primary care. One reason for this design is that GPs and their lists provide an appropriate and readily available sampling frame. Random selection of practices can enable a broad sweep of a national population to be made at relatively small cost. Nevertheless we found that the majority of studies, even of community based intervention, which used this sampling frame took the lists of all participating practices and applied methods of individual sampling to persons (eg or families (eg or households (eg). The main reason for this design (or one using the doctor as randomisation unit) is to study the effect or the implementation of a preventive procedure or a modification of primary care management. If this involves a change in behaviour of a doctor to all patients then individual randomisation is not possible. If other staff in the practice are involved or if there is concern about contamination between the lists of different partners in a practice then the entire list from each practice must be assigned to the same study group. A study of this type tends to use the randomised design we have been discussing though in at least one instance the practices were matched; it is of interest that though socio-economic factors featured in the matching criteria subsequent correspondence recommended randomisation rather than matching.

Because general practice lists are not always up to date any use of them to provide population denominators would be suspect. None of the studies we have cited use these. In our own study considerable effort has been put into correcting the lists before they were used as a register of the study population. This involves removing from the denominator duplicate names and those of women who had left Edinburgh or who had died before their entry to the trial. This will be described elsewhere; for the present we would point out that the only difference between the treatment of the denominators for the two populations is that the screening invitation system provides one extra method of pruning the denominator for the screening population. This could bias our observed mortality rates but in the opposite direction to the differences that we have found.

The reasons that the Edinburgh trial adopted this design have been described. An important reason was the assumption that individual randomisation would be unacceptable to general practitioners and their patients. It is interesting that the organisers of studies of other kinds of screening have not had this opinion. Examples of group randomisation abroad include those involving screening for hypertension using general practices as cluster (eg), dietary intervention using hospitals as clusters (eg) and water fluoridation using large geographical communities as clusters. This design is undoubtedly not common; our experiences indicate that individual randomisation should not be abandoned without good reason. Nevertheless, there will continue to be situations where other considerations are of overriding importance. When this design is adopted then we would recommend that the randomisation be stratified by socio-economic criteria. Census data of this type were of course not available to us at that time. Since post codes were still not nested within enumeration districts in England and Wales in 1981 such census data are still not available to our English colleagues.

Since it is well known that socio-economic factors are associated with mortality and morbidity differences it is likely that the socio-economic differences could lead to biased results. The two randomised trials of breast screening which have already reported were both able to demonstrate almost identical all cause mortality in the two arms of their studies. It has already been shown that the socio-economic indices which we have attached to individual GP practices are significantly associated with mortality rates. Now in this study it has been demonstrated that if the data were analysed without any adjustment for socio-economic factors spurious conclusions would follow. The intervention which is offered here can have no effect on mortality from all causes, cardiovascular disease or lung cancer and yet substantial differences have been found; these are statistically significant. Had this been a trial of, for
example, screening for lung cancer there would have been a distinct possibility of a false result.

Thus the second conclusion of this study is that in a study of this type differences in socio-economic status must be regarded as a real possibility; we have demonstrated a viable method of checking whether this has occurred. At least one of the other studies is currently being checked using census data, and another is collecting personal socio-economic data so that comparisons can be made.

Our methods of adjusting for the differences have proved satisfactory, though only when using the approach which respects the randomisation units. Thus we may conclude firstly that in this sort of situation it is necessary, as Cornfield indicated, to analyse with the GP practice as basic unit, and secondly that with this approach it is possible to adjust successfully for large differences between the populations. Though this is the case we would nevertheless recommend stratified randomisation. It is interesting that the first published account using randomisation between rather than within groups appears to be the “factories” trial of CHD intervention. This used a stratified randomisation, since factories were first arranged in pairs, matched according to size, area and nature of industry; within each pair one was randomly allocated to intervention. This is the design discussed by Donner and Donald.

Finally there is the question of which outcomes will require this adjustment. The Black report suggests that almost any morbidity or mortality outcome might do so. In a recent Edinburgh study it has been confirmed that for breast cancer there is no evidence of an association of higher mortality with lower socio-economic status. Indeed there is some evidence of an association in the opposite direction. Nevertheless we would emphasise that these results indicate that an analysis of the Edinburgh randomised trial of breast cancer screening which does not adjust for socio-economic status is likely to be biased. The above considerations suggest that the bias will be small and conservative. The intention is that the methods presented here will be adopted in that analysis. We believe that it is important to place this decision on record before the breast cancer mortality data are examined. With hindsight we believe it is to be regretted that this trial could not use individual randomisation or group randomisation stratified by socio-economic status. However neither was practicable in 1978 and the former was also believed to be unacceptable. Therefore all the authors, whether involved at the design stage or not, agree that the design decision was inevitable at the time.

The Edinburgh breast screening project is funded by the Scottish Home and Health Department and the Cancer Research Campaign. We wish to acknowledge the support of the project committee whose chairman is Professor Sir Patrick Forrest. We also wish to thank Dr Robin Prescott, Mrs Margaret Miller, Miss Sheila Bonner and Mrs Alice Smith of the Medical Statistics Unit: the former for his advice and encouragement and the others for much laborious data processing. We should also like to thank Dr George Freeman of the Association of University Teachers of General Practice for mailing his members, and the many individual members who responded to our request for information about other studies with a similar design.

Address for correspondence and reprints: Dr Freda Alexander, Leukaemia Research Fund Centre, Department of Pathology, University of Leeds, 17 Springfield Mount, Leeds, LS2 9NG, England.

References
7 Donner A, Donald A. Analysis of data arising from a stratified design with the cluster as unit of randomisation. Stat Med 1987; 6: 43–53.
16 Freeling, Paul. Personal Communication.  
32 Smith WCS. Personal communication.  
33 Bland JM. Personal Communication.  

Accepted for publication October 1988