REVIEW ARTICLE: *Research methods in epidemiology, IV*

Primary prevention trials in cardiovascular disease

Stuart J Pocock, Simon G Thompson

In principle, it seems highly desirable to use randomised controlled trials to evaluate the potential benefit of any primary prevention policy for reducing the risk of cardiovascular disease in the community. Observational epidemiology can lead to the development of hypotheses regarding which primary interventions *ought* to work, but one can argue that only by comparative (and usually long term and large scale) experimentation can one achieve scientifically acceptable evidence that intervention *really does* (or does not) work. Clinical trials can also answer questions about the time scale of the effect of an intervention, and can quantify the side effects of specific regimens, both important issues which cannot be answered reliably from observational studies. However, the planning and interpretation of primary prevention trials is often fraught with controversy. Both trials with negative and positive conclusions (eg, the Multiple Risk Factor Intervention Trial [MRFIT][1] and the Lipid Research Clinics Coronary Prevention Trial[2]) provoke strong reactions from protagonists in opposing camps, such that one fears at times that the credibility of randomised trials in primary prevention is being questioned. Given the vast expense that many such trials require, these anxieties are all the more pressing. In an attempt to review the current state of the art in primary prevention trials, this article draws together several key aspects of study design, analysis and interpretation using examples from recent experience to illustrate the various problems. While the focus is on cardiovascular disease, we would hope that many of the principles could also be applied to the primary prevention of other diseases, eg cancer.

Which questions to tackle?

It is of fundamental importance to investigate preventive approaches that are both scientifically plausible and feasible to implement in practice. Since the aetiology of cardiovascular disease is multifactorial, one could argue that the trials of greatest overall relevance to cardiovascular health are those that study the impact of a *multifactorial intervention policy* where the specific intervention in each individual subject is adapted to his particular risk factor profile. Here the treatment policy is a broad package including changes in life style such as advice on dietary change, exercise and stopping smoking. However, such trials can have organisational and intervention problems;

for instance they can be very costly and labour intensive (as in MRFIT[1]), the actual impact on life style in middle age may be limited (as in the UK component of the WHO factories intervention trial[3]), and the uniqueness of their particular treatment package can make it difficult to draw generalisable conclusions.

An alternative approach is to intervene on one particular risk factor. This can be non-pharmaceutical, eg, trials of dietary change, exercise programmes or stopping smoking, but more commonly we are concerned with *drug trials*, eg, for lowering serum cholesterol or blood pressure. In view of the considerable profits to be made from widespread medication, even if confined to high risk subjects, it is understandable why the pharmaceutical industry funds such studies. Also, they tend to be easier to organise and interpret than non-drug trials. The consequence is that market forces may encourage more drug trials, to the detriment of research into broader policies of primary prevention. At present these comments seem particularly applicable to trials of lipid lowering drugs. First they emphasise just one component of coronary risk (blood lipids) and secondly they emphasise drug therapy to an increasingly wide sector of the general population. While such trials have some scientific value, their current dominance may be distorting the real research needs in primary prevention.

Which subjects to study?

At one extreme coronary prevention campaigns can relate to the whole community (as in North Karelia[4]), but it is more common to undertake studies *in middle age*. Also since women are at lower coronary risk (although not necessarily lower stroke risk) most studies recruit *men only*, even though the conclusions may be extrapolated to women. The focus on middle age is essentially to achieve sufficient cardiovascular events for an adequate statistical comparison while also studying subjects who are young enough (and perhaps compliant enough) to achieve substantial long term benefit. Because the rate of cardiovascular events increases exponentially with age, the choice of an *upper age limit*, 55, 60 or 65 years say, may crucially affect the number of events and hence the statistical power of the trial, whereas including subjects at younger ages, eg, under 45 years, may contribute little to the conclusions. Thus, while a study may recruit over

Clinical Trials
Research Group,
Department of
Epidemiology and
Population Sciences,
London School of
Hygiene and Tropical
Medicine, Keppel St,
London WC1E 7HT
S J Pocock
S G Thompson

Correspondence to:
Professor Pocock

a wide age range (eg, 35–64 years in the MRC Mild Hypertension Trial[5]) the statistical findings will be dominated by subjects in the top 10 year age group.

For trials in mild hypertension one factor in recruitment has been the undue focus on the diastolic pressure with scarcely any reference to the systolic pressure, even though the latter is more predictive of cardiovascular events.[6]

Another major decision concerns whether to recruit *a representative group* (say of middle aged men) *or* to select only subjects at *high risk*. The recent aspirin trials[7 8] have opportunistically recruited an unrepresentative low risk group (physicians). This does not deny the validity of the randomised comparison, but raises some doubts about generalisability to other higher risk subjects. Also, the absolute benefits in terms of reduction in events per person-year becomes artificially low. The British factorial design trial of low dose warfarin and aspirin[9] is recruiting middle aged men who are classified as high risk in terms of being in the top fifth of a multifactorial risk score. This seems desirable in increasing statistical power, focusing attention on the subjects in greatest need and perhaps avoiding unnecessary anxiety by not intervening in subjects at lower risk.

Trials of antihypertensive or lipid lowering treatment (usually drugs) are inevitably aimed at recruiting subjects with higher risk factor levels. However, they often mistakenly ignore other risk factors by selecting all subjects above a given cholesterol or blood pressure level. This means subjects in such a trial may have widely varying risks and some may not even be at lower than average risk if they are free of all other risk factors barring (say) a raised cholesterol level. Another problem in these trials is the *choice of a cut off level* of, say, total cholesterol or low density lipoprotein cholesterol. Set it too low and the trial recruits a high proportion of all screened subjects, which if extrapolated to subsequent practice would imply an unrealistic scale of mass medication with lipid lowering drugs. Set it too high, as perhaps occurred in the Lipid Research Clinics trial,[2] and the findings then relate to an unduly selected group with the consequent danger of clinical extrapolation beyond the study population.

A run in period prior to randomisation is important for several reasons. First, it allows removal of subjects not likely to comply with the intervention, and increasing the compliance rate can greatly increase the statistical power of the trial.[10] Second, it allows time for regression to the mean to exclude falsely high initial readings. Third, such a run in period should include appropriate non-pharmacological intervention (eg, dietary advice) for all subjects to see if risk factor reduction can first be achieved without drugs.

Randomisation and control groups

The inclusion of a randomised control group not receiving intervention is important to any primary prevention trial. Non-randomised comparative community intervention studies such as the North Karelia project[4] provide some valuable information on practical problems of implementation but one is never able to infer confidently how much any changes in the intervention community are directly attributable to the intervention package.

The conventional approach is to *randomise individual subjects* to intervention or control group, and in double blind placebo controlled drug trials this usually presents no problem. However in trials involving some change in life style as part or all of the intervention package, it can be difficult to prevent individuals allocated to control being aware of and affected by the proposed intervention. Indeed it is thought that the MRFIT trial[1] may have failed to show a relative benefit of special intervention because of the effective management of risk factors in the "usual care" control group.

An alternative is to undertake *group randomisation* as carried out in the WHO factories study,[11] where factories in several European countries were randomly assigned to a coronary risk intervention programme or control. Here the group (or factory) becomes the unit for statistical analysis, so the main problem is to achieve a sufficient number of groups so that the consequent loss of statistical power is kept small.[12] One British study of coronary risk reduction in general practice[13] is currently being planned with both individual and group randomisation. The individual approach has the risk of carry over effect of the intervention into the controls in the same general practice, while the group approach has to get sufficient practices to agree to randomisation.

A factorial design is an efficient way of attempting to answer two primary prevention hypotheses within the same trial. For example, the US physicians study[7] has both an aspirin component aimed at reducing cardiovascular disease, and a β carotene component aimed at decreasing the incidence of cancer. This seems an ideal circumstance for a factorial design, since the postulated mechanisms and principal outcome measures for the two interventions are distinct. The use of such designs for related interventions and outcomes is more problematic. For example, a British factorial trial of low dose warfarin and aspirin[9] may have a reduced power to detect reductions in cardiovascular disease rates if the combined benefit of both regimens turns out to be less than the sum of the individual benefits of each regimen. Allowance for the possibility of such a moderate interaction has been made by increasing the sample size.

Which outcome measures?

The motivation for most primary prevention policies is to reduce the occurrence of major cardiovascular events (eg, myocardial infarction and stroke) and also to reduce mortality (both from cardiovascular disease and all causes). While not wishing to disagree with such broad goals, we see a need for each primary prevention trial to specify *realistic priorities regarding which should be the primary outcome measure(s)*.

For instance, total mortality may seem the most important overall measure. However, studies lack power to show any reduction in total deaths because of the diluting effect of non-

cardiovascular deaths, which one might anticipate would be unrelated to intervention. While the WHO clofibrate trial[14] with its excess of deaths on active treatment is an exceptional salutary experience, it would be detrimental to primary prevention research if it were used as an argument for emphasising total mortality.

There still remains a substantial selection of cardiovascular end points: fatal and/or non-fatal events (both major and minor), coronary and/or cerebrovascular events. The priorities for any given trial will depend on the mode of intervention and type of subject recruited. For aspirin trials the most appropriate summary end point may be occurrence of any major cardiovascular event (eg, myocardial infarction or stroke) both fatal and non-fatal, whereas major coronary events alone for lipid lowering trials and stroke alone for mild hypertension trials would be more directly related to intervention. While deaths are more important than non-fatal events the latter are more common so that greater power can be achieved in combining the two. In some fields, such as lipid lowering trials, it appears that treatment may have a greater proportional effect on non-fatal events. However, it would seem peculiar to exclude fatal events from the primary end point since they clearly matter in the overall picture.

It is desirable to have *a pre-specified primary end point*, since post hoc emphasis on the more significant outcome measures can be misleading. For instance, the US Physicians Study report[7] emphasised the 44% reduction in myocardial infarction on aspirin, whereas the original primary end point, cardiovascular death, showed no significant difference between aspirin and placebo (but also had insufficient deaths for a reliable comparison). In retrospect, all major cardiovascular events might have been a more judicious choice of primary end point, showing an 18% risk reduction on aspirin (p = 0·01).

Lastly, some primary prevention trials, especially of life style intervention, may have the more limited goal of studying the actual impact on risk factors rather than cardiovascular events. For instance, in Britain it is important to encourage coronary risk reduction programmes in a general practice setting, in the overall spirit that primary care facilities should be concerned with prevention as well as treatment of disease. Hence there need to be trials which evaluate the effectiveness of realistic and well defined coronary intervention programmes to see how individual middle aged men (and their families) respond in terms of risk factor reduction over a time period of, say, one year. Here it is relevant that particular controversies, such as the benefit (if any) of cholesterol screening, be properly assessed using randomised controlled trials rather than being routinely introduced as the latest fashion.

Data monitoring and stopping rules

The public health implications for extrapolating the findings of primary prevention trials into routine practice are usually considerable so that great care should be taken to reach reliable conclusions. Accordingly, while data monitoring of interim results is desirable one should be wary of early trial termination and publication in favour of intervention unless the evidence is overwhelming. Since such trials relate to intervention (often mass medication) in the general population the ethical need to stop early if "treatment" appears effective is less strong than in trials of sick patients. However, it remains just as pressing to stop early if "treatment" appears harmful. Thus, there needs to be an asymmetry in stopping rules, with more frequent looks at the data and less stringent nominal p values for early termination if the intervention is doing worse than control, either in terms of efficacy or adverse events.

The US physicians study[7] is an interesting example, since the trial stopped early because of a substantial reduction in myocardial infarctions on aspirin while the primary end point, cardiovascular deaths, showed no benefit of aspirin. The wisdom of this decision is open to debate, especially since the British physicians' trial[8] did not show such a difference.

Statistical analysis and reporting

The basic principles of statistical reporting in medical research also apply to primary prevention trials. While significance testing is a valuable tool, especially if one reports *exact p values* rather than just "p < 0·05" or "not significant", primary prevention trials should also pay more attention to the estimated magnitude of effect, especially with the *use of confidence intervals*.

For instance, the summary of the Oslo study of diet and smoking intervention in healthy men[15] reported a 47% lower incidence of myocardial infarction and sudden death in the intervention group than in the controls (p = 0·028). The statistically uninformed reader may be inclined to believe that such intervention really does halve the risk of coronary disease, by failing to consider the extent of random variation. In fact, the 47% reduction is based on 19/604 v 36/628 coronary events in intervention and control groups respectively. The 95% confidence interval, from 5% to 68% reduction in risk, is wide and hence provides the necessary caution in the interpretation of the data.

In primary prevention trials it has become usual to report the estimated proportionate risk reduction attributed to intervention. While a proportionate risk reduction may be the most reliable statistic for extrapolating from the trial to the general population, consideration of *absolute treatment differences* is appropriate for assessing the balance of risks and benefits. For instance, the US physicians trial[7] reported "139 myocardial infarctions amongst those assigned to aspirin and 239 amongst those assigned to placebo (relative risk, 0·56; 95 percent confidence interval, 0·45 to 0·70; p < 0·00001)." They go on to point out that each group had around 55 000 person-years of follow up, so that the absolute treatment difference is 1·8 myocardial infarctions per 1000 person-years with 95% confidence interval from 1·1 to 2·5 per 1000 person-years. For the more global combined end point of non-fatal myocardial infarction, non-fatal stroke and cardiovascular death the absolute treatment difference is less, ie, 1·1 events per 1000 person-years. A simple summary of the possible benefit

(−) and harm (+) of aspirin treatment is achieved by the following point estimates of differences (aspirin-placebo) in numbers of events: −100 myocardial infarctions, −63 cardiovascular events in total, −2 cardiovascular deaths, −10 deaths from all causes, +21 strokes, +31 ulcers, +731 bleeding problems of which +20 required blood transfusion. This use of absolute differences provides an important perspective on the effects of intervention, the inclusion of which journal editors could demand.

Like other clinical trials, primary prevention studies can generate a multiplicity of data, so that the usual reservations about data dredging, eg, across multiple end points and subgroup analyses, need to be borne in mind so that post hoc inferences from unanticipated secondary analyses are not given undue weight.

Size of trials and overviews

Coronary heart disease is the most common cause of death and major illness in British middle aged men, so that in terms of the ability to evaluate primary prevention strategies the randomised controlled trial remains more feasible than in other diseases such as cancer.[10] Nevertheless *the required size of trial*, as regards number of subjects and length of follow up, to achieve adequate statistical power *tends to be considerably greater than in secondary prevention trials.*

For instance, let us consider trials of men with high serum cholesterol levels. The Oslo study mentioned earlier[15] screened 16 202 men and randomised 1232 of those with high serum cholesterol to dietary and smoking intervention or to control. In five years follow up there were 55 coronary events. This number of events is the main determinant of statistical power in such trials, and clearly a total of 55 events is not really large enough to estimate reliably the magnitude of the intervention's effect. Although the trial did achieve statistical significance this was only because the point estimate, an observed 47% risk reduction, was considerably larger than in other related trials.

The Lipid Research Clinical Coronary Prevention Trial[2] was a larger scale study. Out of 480 000 men screened, 3806 with high cholesterol levels were randomised to cholestyramine or placebo. In an average seven years follow up 342 coronary events were observed. However, even with this larger scale of investigation the estimated 19% reduction in coronary events associated with cholestyramine was still only of borderline statistical significance. Indeed, the questionable use of a one sided p value did provoke some controversy.

It is very difficult to achieve sufficiently precise and generalisable conclusions for an intervention policy on the basis of any single trial, even if such a trial does conform to acceptable standards of design and statistical power. Hence, the ability to combine evidence from a collection of related trials has been formalised by *the use of overview (or meta-analysis) techniques.* For instance, a recent overview of all the available antihypertensive trials[16] has concluded, with fairly narrow confidence limits, that for people with moderately raised diastolic blood pressure, one can achieve an average 40% reduction in strokes with standard antihypertensive drug treatment. Of course, such overviews make substantial assumptions in combining evidence from trials of diverse treatments with differing subject selection and evaluation standards, so that the numerical precision of any overview estimates should be viewed with considerable caution. Nevertheless, if such overviews endeavour to include *all* such trial evidence then they are capable of defining at least the order of magnitude of effect that may be realistically achieved by a given type of prevention policy.

Applicability of trial findings

While all the above methodological issues are important in both designing and interpreting primary prevention trials, perhaps the key issue is whether such trials can provide reliable information on which approaches to improving cardiovascular health in the general population are desirable and cost effective.

Across the spectrum of potential intervention trials, those with the greatest "internal validity" (eg, double blind trials of drug treatment in high risk compliers) are often the ones which are most difficult to extrapolate to future subjects and thus lack "external validity". Thus, we need to be wary of investing too much trial effort into questions that are scientifically appealing, amenable to controlled trials but of limited value to public health.

1 Multiple Risk Factor Intervention Trial Research Group. Multiple Risk Factor Intervention Trial: risk factor changes and mortality results. *JAMA* 1982; 248: 1465–77.
2 Lipid Research Clinics Program. The Lipid Research Clinics Coronary Prevention Trial results. *JAMA* 1984; 251: 351–74.
3 Rose G, Tunstall-Pedoe HD, Heller RF. UK Heart Disease Prevention Project: incidence and mortality results. *Lancet* 1983; i: 1062–5.
4 Puska P, Salonen JT, Nissinen A, *et al.* Change in risk factors for coronary heart disease during 10 years of a community programme (North Karelia project). *Br Med J* 1983; 287: 1840–4.
5 Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *Br Med J* 1985; 291: 97–104.
6 Rutan GH, McDonald RH, Kuller LH. A historical perspective of elevated systolic *vs.* diastolic blood pressure from an epidemiological and clinical trial viewpoint. *J Clin Epidemiol* 1989; 42: 663–73.
7 Steering Committee of the Physicians' Health Study Research Group. Final report of the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989; 321: 129–35.
8 Peto R, Gray R, Collins R *et al.* Randomized trial of prophylactic daily aspirin in British male doctors. *Br Med J* 1988; 296: 13–16.
9 Meade TW, Wilkes HC, Stirling Y, Brennan PJ, Kelleher C, Browne W. Randomized controlled trial of low dose warfarin in the primary prevention of ischaemic heart disease in men at high risk: design and pilot study. *Eur Heart J* 1988; 9: 836–43.
10 Zelen M. Are primary cancer prevention trials feasible? *J Natl Cancer Inst* 1988; 80: 1442–4.
11 World Health Organization European Collaborative Group. Multifactorial trial in the prevention of coronary heart disease: 3. Incidence and mortality results. *Eur Heart J* 1983; 4: 141–7.
12 Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978; 108: 100–2.
13 Anonymous. National Family Heart Study (notes and news). *Lancet* 1989; i: 456.
14 Committee of Principal Investigators. WHO cooperative trial on primary prevention of ischaemic heart disease using clofibrate to lower serum cholesterol: mortality follow-up. *Lancet* 1980; ii: 379–85.
15 Hjermann I, Velve Byre K, Holme I, Leren P. Effect of diet and smoking intervention on the incidence of coronary heart disease. *Lancet* 1981; ii: 1303–10.
16 Herbert PR, Fiebach NH, Eberlein KA, Taylor JO, Hennekens CH. The community-based randomized trials of pharmacologic treatment of mild-to-moderate hypertension. *Am J Epidemiol* 1988; 127: 581–90.