Implications of Findings in the Coronary Drug Project for Secondary Prevention Trials in Coronary Heart Disease

THE CORONARY DRUG PROJECT RESEARCH GROUP

SUMMARY The associations observed in the Coronary Drug Project between two baseline factors — serum cholesterol level and number of cigarettes smoked per day — and 5-year mortality in men who had a history of myocardial infarction are used to derive required sample sizes for future risk factor intervention trials in the secondary prevention of coronary heart disease. Consider a trial in which it is anticipated that the baseline cholesterol level will be 250 mg/dl in the control group and that a 20% reduction in this level to 200 mg/dl will be experienced by the treated group. Let it also be assumed that this reduction in cholesterol level will have the effect of immediately reducing the mortality risk to that corresponding to a patient with a baseline level of 200 mg/dl. Given a type I error rate, $\alpha$, of 0.05 for a two-sided test, a type II error rate, $\beta$, of 0.10, and a follow-up period of 5 years, the required size of the trial would be over 10,000 patients. A similar sample size would be required for a single-factor trial focusing on cigarette smoking cessation. The sample size might be reduced to 6000 for a two-factor trial with intervention on cholesterol and cigarette smoking simultaneously. These numbers increase two- to threefold when dropout patients and “dropin” patients (i.e., those in the control group who decide on their own to begin the intervention therapy) are considered and when other assumptions are made concerning the anticipated time required for the treatment to achieve maximum benefit with respect to the mortality or morbidity end point.

MANY FACTORS — demographic, historical, clinical, biochemical, electrocardiographic and hygienic — are associated with the development of first and recurrent events of coronary heart disease (CHD). Factors that can be modified by diet, drugs or other prophylactic or therapeutic measures have been the foci during the past 2 decades for clinical trials in the primary and secondary prevention of CHD. Although the rationale for such trials lies largely in the observed association between a particular risk factor and manifestation of the disease, rather infrequently has information on the strength of such an association been used in deriving the required sample size for the trial. Rather, it has been customary to assume that the intervention will produce a 25%, 35% or even 50% reduction in the rate of CHD events compared with the control group, and then to derive the sample size required to detect such a reduction.

Data from the Framingham study on the relationship of serum cholesterol level to first events of CHD have been used to determine how large the sample sizes would have to be for trials of dietary intervention in the primary prevention of CHD.$^{1,2}$ Data from the same study on the joint association of serum cholesterol, diastolic blood pressure, and number of cigarettes smoked daily with first events of CHD have similarly been used to derive the required sample size for the Multiple Risk Factor Intervention Trial (MRFIT), sponsored by the National Heart, Lung, and Blood Institute, in the primary prevention of CHD.$^{3}$ In the present paper the associations observed in the Coronary Drug Project (CDP) between two baseline factors — serum cholesterol level and number of cigarettes smoked per day — and 5-year mortality in men who had a history of myocardial infarction (MI) are used to derive required sample sizes for future risk factor intervention trials in the secondary prevention of CHD. Blood pressure, another treatable risk factor strongly associated with first events of CHD, is not considered here because of its weak association with 5-year mortality in men after MI.$^{4}$ (Coronary Drug Project Research Group: The natural history of myocardial infarction in the Coronary Drug Project: long-term prognostic importance of blood pressure. Manuscript in preparation). The statistical methods used in this paper are similar to those used in the papers dealing with the question of primary prevention of CHD.$^{1,3}$

Methods

The background, design and organization of the CDP have been described in earlier reports.$^{5,6}$ The primary objective was to test the efficacy and safety of several lipid-influencing drugs in the long-term therapy of CHD in men with proved previous MI. For this purpose, 8341 patients were recruited by the 53 participating clinical centers and were randomly assigned to six groups. Approximately one-third of the patients (2789 men) were allocated to the placebo group.

Patients accepted into the CDP were men ages
30–64 years who had documented evidence of one or more MIs, were considered on entry to be in functional class I or II of the New York Heart Association classification, and were free from a specified list of excluding diseases and conditions. All patients had had their latest MI at least 3 months earlier, and none had evidence of recent worsening of CHD or of other major illnesses. In a series of standardized baseline examinations, extensive data — demographic, historical, clinical, electrocardiographic and biochemical — were collected for all patients.

All men reported to the clinic for an interval visit every 4 months over a period of follow-up ranging from 54–101 months. About 96% of the men were followed for at least 60 months. This report uses baseline observations and deaths reported during 5 years of follow-up for the 2789 men in the placebo group.

Two potentially treatable baseline risk factors, serum cholesterol level and number of cigarettes smoked per day, have been selected for consideration in this report. The relationship between 5-year total mortality and one of these baseline risk factors can be expressed by the following logistic regression equation:

\[
y(x) = [1 + \exp (-a - bx)]^{-1}
\]

where \(y\) = expected 5-year mortality and \(x\) = value of baseline risk factor. The parameters \(a\) and \(b\) of the regression equation are estimated using the Walker-Duncan method.

Suppose a controlled clinical trial is to be conducted to determine whether intervention on one of these risk factors will lead to a reduction in 5-year mortality in a group of post-MI men. To determine the required sample size for this trial, one must specify the expected 5-year mortality rates, \(\pi_1\) and \(\pi_2\), in the control and intervention groups, respectively, under the alternative hypothesis that there is truly an effect of intervention on mortality. First, the simplifying assumption is made that all patients in the control group have at baseline and maintain throughout the study a value of \(x_1\) of the risk factor, and that all patients in the intervention group achieve and maintain throughout the study a value of \(x_2\) of the factor. From equation 1, given estimates \(\hat{a}\) and \(\hat{b}\) from the CDP placebo group data, the expected 5-year mortality in the control group of the new clinical trial is \(\pi_1 = y(x_1)\). If it were to be assumed that the modification by intervention of the risk factor from a level of \(x_1\) to a level of \(x_2\) were to have the effect of immediately changing the patient’s mortality risk to that corresponding to a patient with a baseline level of \(x_2\) in the intervention group could be obtained from equation 1 as \(\pi_2 = y(x_2)\).

The assumption that all patients in a treatment group will have the same value of \(x\) is obviously unrealistic and in some cases too restrictive (although in other cases, it may make little difference in the final results). Relaxing this assumption, let us suppose that the risk factor can take on as many as \(m\) distinct values. Let \(f_i(x)\) denote the relative frequency of occurrence of the value \(x_i\) among patients in the control group, and let \(f_i(x)\) denote the relative frequency of \(x_i\) in the intervention group, \(i = 1, \ldots, m\), after intervention. Then,

\[
\pi_1 = \sum_{i=1}^{m} f_i(x_i)y(x_i)
\]

and

\[
\pi_2 = \sum_{i=1}^{m} f_i(x_i)y(x_i). \tag{2}
\]

The required sample size for a two-sided test of the null hypothesis, \(\pi_1 = \pi_2\), is given by

\[
n = \frac{(Z_{1-\alpha} \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)} + Z_{1-\beta} \sqrt{2\pi(1-\pi)})}{(\pi_1 - \pi_2)^2}, \tag{3}
\]

where for any number \(\gamma\) between 0 and 1, \(Z\), denotes the 100\(\gamma\)-th percentile point of the standard normal distribution. \(\alpha\) denotes the significance level or probability of a Type I error, i.e., rejecting the null hypothesis in the absence of a true effect, \(\beta\) is the probability of a type II error, i.e., failing to detect a true effect of the treatment on the end point. \(\pi\) is \((\pi_1 + \pi_2)/2\), and \(n\) is the required number of patients in each of the two groups.

The values chosen for the parameters \(\pi_1\) and \(\pi_2\) are obtained from the logistic function 1 using estimates \(\hat{a}\) and \(\hat{b}\) from an analysis of the CDP placebo group data. These estimates, of course, are subject to random sampling error. Taking into account this sampling variability in \(\hat{a}\) and \(\hat{b}\), and hence in \(\pi_1\) and \(\pi_2\), in determining sample size, one can derive confidence interval estimates of the required sample size. The details of this derivation are provided in the appendix.

The application of the methods described above is illustrated in the Results section for the case of serum cholesterol. Required sample sizes are also derived for intervention trials, based on cigarette smoking considered singly and on the two risk factors simultaneously.

**Results**

**Serum Cholesterol**

The relationship between 5-year total mortality and baseline serum cholesterol level in the CDP placebo group is shown in figure 1. On the basis of data from 2789 patients in the placebo group, the parameters \(a\) and \(b\) of the logistic regression equation 1 have been estimated to be \(\hat{a} = -2.13240\) and \(\hat{b} = 3.21927 \times 10^{-4}\). The logistic regression line based on \(\hat{a}\) and \(\hat{b}\) is superimposed on the bar diagram in figure 1. The estimated variances and covariance of these estimates are: \(\text{Var}(\hat{a}) = 0.0653578\), \(\text{Var}(\hat{b}) = 9.62264 \times 10^{-7}\), and \(\text{Cov}(\hat{a},\hat{b}) = -2.44612 \times 10^{-4}\).

On the basis of the logistic equation 1 with estimates \(\hat{a}\) and \(\hat{b}\), table 1 gives the estimated 5-year mortality for different levels of baseline cholesterol, \(x\). The estimated 5-year mortality for a group of patients all with a baseline cholesterol level of 250 mg/dl is 0.20956. If the patients with a baseline cholesterol of 250 mg/dl were to have this level reduced by diet or...
drug by 10% (to 225 mg/dl), and if this reduction were to have the effect of immediately changing those patients’ mortality risk to that corresponding to patients with a baseline cholesterol level of 225 mg/dl, the expected reduction in 5-year mortality would be from 0.20956 down to 0.19654, or a decrease of 6.2% (table 1). For a 20% reduction in cholesterol, from 250 to 200 mg/dl, the corresponding mortality reduction is 12.1%. For 10% and 20% reductions in cholesterol from a baseline of 350 mg/dl, the corresponding mortality reductions are 8.0% and 15.6%, respectively.

Instead of assuming that all patients in a treatment group have the same cholesterol level (e.g., 250 mg/dl for the control group and 225 mg/dl for the intervention group), assume that the patients follow a distribution somewhat similar to that indicated in figure 1 for CDP patients. Suppose, for simplicity, that five distinct values of cholesterol in the control group occur in the following proportions: 200 mg/dl, 0.18; 225 mg/dl, 0.20; 250 mg/dl, 0.24; 275 mg/dl, 0.20; and 300 mg/dl, 0.18. Suppose also that the cholesterol values achieved in the intervention group follow the same distribution except that each value is 10% lower in the control group, i.e., 175.5, 204.3, 223.2, 244.8 and 288.0. The estimated 5-year mortality for this situation is 0.21013 for the control group (mean cholesterol of 250 mg/dl) and 0.19700 for the intervention group (mean cholesterol of 225 mg/dl) (table 1, part B). These values are very close to those based on the assumption that all patients in a treatment group have the same cholesterol value.

Table 2 is a list of required sample sizes using formula 3 for comparing a cholesterol-lowering therapy to a control therapy, given \( \alpha = 0.05 \) (two-sided) and \( \beta = 0.10 \) (thus, \( Z_{1-\alpha/2} = 1.96 \) and \( Z_{1-\beta} = 1.28 \) from a table of the standard normal distribution), two baseline cholesterol levels (250 and 350 mg/dl), and two levels of cholesterol reduction (10% and 20%). The results in table 2, part A are based on the assumption that all patients in a treatment group have the same cholesterol value. Table 2, part B indicates that when a distribution is assumed for the cholesterol values, the required sample sizes differ only 1.5% from those in part A. For the case of patients with a baseline cholesterol level of 250 mg/dl in the control group and a 10% reduction in cholesterol effected in the inter-

**TABLE 2. Required Sample Sizes Per Treatment Group for Comparing a Cholesterol-lowering Regimen to a Control Group Regimen, \( \alpha = 0.05 \), \( \beta = 0.10 \)**

<table>
<thead>
<tr>
<th>Baseline cholesterol (mg/dl)</th>
<th>% cholesterol reduction</th>
<th>Required sample size per treatment group</th>
<th>95% confidence limits Lower</th>
<th>95% confidence limits Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Patients all with same value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>20,065</td>
<td>7,962</td>
<td>117,685</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>5,142</td>
<td>2,057</td>
<td>29,075</td>
</tr>
<tr>
<td>350</td>
<td>10</td>
<td>8,676</td>
<td>3,250</td>
<td>64,475</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2,233</td>
<td>841</td>
<td>15,965</td>
</tr>
<tr>
<td>B. Patients with distribution of values (see text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>19,767</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>5,067</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
vention group, the required sample size is 20,065 for each of the two treatment groups. Taking into account the uncertainty as to the true values of a and b, as indicated by a 95% confidence interval estimate, the required sample size falls in a range of 7962–117,685 per treatment group. For a 20% cholesterol reduction from a baseline of 250 mg/dl, the required sample size is 5142 per group, with a range of 2057–29,075. For a group with a baseline cholesterol of 350 mg/dl, the required sample sizes are substantially smaller than for a baseline cholesterol of 250 mg/dl. However, the group of patients available with such high cholesterol levels is very limited, so recruitment is more difficult; e.g., in the CDP only 3.2% of the men in the study had baseline cholesterol levels of 350 mg/dl or higher.6

Cigarette Smoking

The 5-year mortality in the CDP by number of cigarettes smoked per day at baseline is shown in figure 2. The parameters a and b of the logistic regression equation 1 have been estimated to be \( \hat{a} = -1.43466 \) and \( \hat{b} = 0.303003 \), where the regressor variable, \( x \), is defined as 0 for nonsmokers and 1 for smokers at baseline. One might conclude from figure 2 that if intervention on cigarette smoking is to have any chance of improving long-term survival after acute MI, it will have to be by getting the patients to stop completely, and not just to cut down. Consider an intervention trial that includes only post-MI men who smoke cigarettes. Without intervention, the 5-year mortality in these patients is estimated from the CDP data to be 0.24386 (i.e., 258 of 1058 from figure 2, or \( y(1) \) from equation 1 using the estimates of \( a \) and \( b \) given above). Suppose that the goal of the intervention program is for 25% of the men to stop smoking. If one assumes that those who stopped smoking stopped very soon after the start of the trial, and that the mortality risk of those who stopped immediately reverted to that risk for the nonsmokers at baseline, the 5-year mortality might be expected to be reduced to 0.23016 (obtained by substituting \( x = 0.75 \), the proportion of smokers after intervention, into equation 1, and using the estimates of \( a \) and \( b \) given above). The sample size required to detect such a mortality reduction is 20,252 in each treatment group (table 3). For a more optimistic intervention goal, e.g., 50% of the men stopping smoking, 5169 patients would be required in each of the two groups (table 3).

Both Risk Factors

In view of the rather large sample sizes required for univariate risk factor intervention trials in the secondary prevention of CHD, the next step is to consider trials that include intervention strategies for several risk factors simultaneously. The MRFIT study4,5 is an example of such a trial for prevention of first coronary events. The risk factors of interest in the MRFIT are serum cholesterol, diastolic blood pressure and cigarette smoking. Consider now, as a further illustration of the procedure described above, a multiple risk factor intervention trial in secondary prevention of CHD involving serum cholesterol and cigarette smoking. It is again assumed, for simplicity, that all of the patients in the trial have a baseline cholesterol level of 250 mg/dl and also assumed that 37.935% (i.e., 1058 of 2789 from figure 2) of the patients are cigarette smokers at baseline. A logistic regression analysis has been carried out to obtain the following relationship of 5-year mortality with baseline serum cholesterol and cigarette smoking status jointly:

\[
y(x_1,x_2) = [1 + \exp(2.22380 - 0.0031181x_1 - 0.292910x_2)]^{-1}
\]

(4)

where \( x_1 \) denotes baseline serum cholesterol level and \( x_2 \) is 0 for nonsmokers and 1 for smokers at baseline. Consider first a trial with a rather moderate intervention goal, e.g., reduction of cholesterol from 250 to 225 mg/dl and cessation of smoking by 25% of the smokers. Without intervention, the 5-year mortality in these patients is estimated from the CDP data to be

| Table 3. Required Sample Sizes Per Treatment Group for Various Intervention Trials, \( \alpha = 0.05, \beta = 0.10 \) |
|---|---|
| Intervention goal | Required sample size per treatment group |
| 1. Cigarette smoking stopped in 25% of smokers | 20,252 |
| 2. Cigarette smoking stopped in 50% of smokers | 5,169 |
| 3. Cholesterol reduced from 250 to 225 mg/dl in all patients and cigarette smoking stopped in 25% of smokers | 11,755 |
| 4. Cholesterol reduced from 250 to 200 mg/dl in all patients and cigarette smoking stopped in 50% of smokers | 3,086 |
0.20864. This figure is obtained by substituting $x_1 = 250$ and $x_2 = 0.37935$ into equation 4. With intervention, the 5-year mortality might be expected to be reduced to 0.19172. This figure is obtained by substituting $x_1 = 225$ and $x_2 = 0.28451$ into equation 4; the figure for $x_2$ represents the proportion of smokers remaining in the intervention group after 25% of the smokers have quit smoking. The sample size required to detect such a mortality reduction is 11,755 for each treatment group (table 3). This sample size is about half of those required for achieving the same goals one factor at a time (i.e., 20,065 for cholesterol and 20,252 for cigarette smoking). For a more optimistic intervention goal, e.g., reduction of cholesterol from 250 to 200 mg/dl and cessation of smoking by 50% of the smokers, the required sample size is reduced to 3036 per group (compared with 5142 and 5169 when intervening singly on cholesterol and cigarette smoking, respectively).

**Discussion**

During the past 2 decades a number of randomized controlled trials of lipid-lowering drugs and diets in the secondary prevention of CHD have been reported. These include the Edinburgh, Los Angeles, and Chicago trials of estrogen, the Newcastle and Scottish trials of clofibrate, the CDP trial of estrogen, clofibrate, dextrothyroxine and nicotinic acid, the London trial of a lowfat diet, and the Oslo and London trials of a diet low in saturated fats and cholesterol and rich in highly unsaturated fats. Fewer intervention trials have been done for cigarette smoking. Intervention trials of cigarette smoking have not been done in persons who already have a history of CHD; a randomized controlled primary prevention trial of cigarette smoking cessation has been carried out in London and two primary prevention trials are in progress in which cigarette-smoking intervention is included in a multifactorial attack on CHD risk factors.

The sample sizes of the secondary prevention trials cited above range from 100–8341; the largest, the CDP, owes its size in part to the fact that it included six treatment groups. The required sample sizes for the two-group single-factor trials considered in this paper are larger than that for the CDP; even for the more optimistic intervention goals, the sample sizes exceed 10,000. By considering multifactor intervention trials the sample size requirement can be reduced to about 6000.

The discrepancy in range of sample sizes between the secondary prevention trials actually carried out and those considered in this paper is due in part to lack of adequate funding to carry out large trials and in part to inability, because of lack of available data, to consider, in sample size calculations for the actual trial, the relationship between the risk factor targeted for intervention and subsequent mortality and morbidity. For all clinical trials, regardless of the specific disease entity or the particular risk factors being treated, the specification of an alternative hypothesis for the purpose of sample size determinations is extremely important. One cannot simply specify an idealized figure, i.e., that representing a true treatment effect that is deemed to be of clinical importance. It is desirable to bring as much information as possible, particularly with respect to the relationship between the risk factors targeted for intervention and the end point, on the question of what size treatment effect one could reasonably expect to achieve with the proposed therapy.

Other factors may have an impact on the required sample size of a clinical trial, and will be considered as we review the assumptions that went into the computation of the sample sizes in the preceding section. This review focuses primarily on the risk factor serum cholesterol; the principles involved are applicable to cigarette smoking.

(1) Reductions in serum cholesterol of 10% and 20% were assumed. None of the studies cited report a cholesterol reduction of more than 20%. However, a surgical procedure has been reported to reduce serum cholesterol by 35–50%, and a controlled clinical trial of this procedure is in progress to determine whether this degree of cholesterol-lowering translates into improved survival in persons with a history of myocardial infarction. (Long JM, Matts JP, Windler CM, Bearman JE: Preliminary assessment of lipid-lowering effects of partial ileal bypass surgery in a controlled study. Presented at the 18th Annual Conference of Cardiovascular Disease Epidemiology, Orlando, Florida, March 1978). The sample size requirements might be expected to be reduced if more effective intervention approaches are developed.

(2) The computations were based upon the best estimate (using the maximum likelihood method) of the relationship between baseline cholesterol level and 5-year mortality as obtained from the CDP placebo group. Random variation is associated with the estimates of the parameters of the mathematical model expressing this relationship, so 95% confidence interval estimates of the required sample sizes were obtained. For the case of a 20% cholesterol reduction from a baseline of 250 mg/dl, the lower limit of the total required sample size is 4114 (table 2). Thus, as information accrues from additional studies, and as the true values of $a$ and $b$ become more precisely estimated, the best estimate of the cholesterol-mortality relationship may prove to be stronger than that determined from the CDP data and lead to a reduction in required sample size. However, one additional factor must be considered: In a subgroup of 129 male survivors of acute MI in the Framingham Heart Study, ages 40–69 years, the estimated parameters of a logistic function relating baseline cholesterol with 5-year total mortality were $a = -2.03$ and $b = 2.43 \times 10^{-3}$ (Halperin M, McGee D: personal communication). This indicates a relationship somewhat weaker than that determined from the CDP data.

(3) The end point of total mortality has been used in the foregoing computations. An end point that includes both fatal and nonfatal events would yield a
higher incidence of events, and thus might require a smaller sample size. However, for this presumption to be valid, the association between the new combination end point and the risk factor must be at least as strong as that between total mortality and the risk factor. For the end point coronary death or definite nonfatal MI, which had an incidence in the CDP 30% greater than that of total mortality, the association with baseline cholesterol is weaker than that between total mortality and cholesterol. The result is, in fact, larger required sample sizes for the combination end point than for total mortality.

(4) All men in the CDP placebo group, ages 30–64 years at time of entry into the trial, were included in the logistic regression analysis. The association between total mortality and cholesterol could be stronger in a subgroup of men with a more restricted age range. However, in logistic regression analyses of total mortality and serum cholesterol for men with age ranges of 40–59, 45–59, 30–54, and 55–64 years, none of the estimates of the parameter \( b \) (essentially the slope of the regression curve) were substantially different from one another or from that for the entire group of men (ages 30–64 years). Thus, for secondary prevention trials involving intervention on serum cholesterol, nothing is to be gained by narrowing the age range further than 30–64 years.

(5) Treatment was assumed to exert no beneficial effect upon survival apart from the intermediate effect of lowered levels of serum cholesterol. This may not always be a safe assumption, as any given treatment may exert several favorable biochemical and physiologic effects on the patient. Of course, it is possible that certain deleterious effects of a treatment may partially or wholly cancel any beneficial effects that result from the cholesterol lowering.

(6) An \( \alpha \) of 0.05 (two-sided test) and a \( \beta \) of 0.10 were assumed. Smaller sample size requirements are possible if larger type I and type II error rates can be tolerated.

(7) No account was taken of the possibility of patient dropouts, which dilute any real treatment effects and thus require a compensatory increase in sample size.

(8) Nor was any account taken of the possibility that patients in the control group might "drop in" to the intervention group. In other words, some of the patients in the control group may, as a result of local or national public health campaigns or out of concern over their declining health, voluntarily stop smoking or embark on a cholesterol-lowering diet. Thus, if the intervention were effective, the dropins would cause a reduction in mortality in the control group as well, making a true effect of the intervention more difficult to detect. This, too, would require a compensatory increase in sample size.

(9) We assumed that the reduction in serum cholesterol level from \( x_1 \) to \( x_2 \) would result in an immediate reduction in mortality risk to the level of a patient having a cholesterol level of \( x_2 \) at baseline. This may be an optimistic assumption. For certain intermediate treatment effects, such as reduced platelet aggregation or control of cardiac arrhythmias, one might reasonably expect a nearly immediate response in survivorship. However, considering the length of time it takes for atheromatous deposits to develop (or regress) in arteries, any effect on mortality might not be fully realized until after a prolonged period of sustained cholesterol-lowering, if at all. Thus, the sample sizes in table 2 may be substantially understated. This assumption may also be rather optimistic with regard to intervention on cigarette smoking. There are data suggesting that as long as 15–20 years may be required, depending on smoking intensity and duration, before those who quit smoking achieve the risk of a comparable group that has never smoked.

(10) As noted earlier in the discussion, the trial design demands that we estimate as well as we can what risk reductions one could reasonably expect to achieve as the result of a specified intervention. The strict applicability of the risk function approach to answer this question rests directly on two assumptions: First, the risk function, a descriptive summary of the observed relationship between incidence and risk factors (in the absence of intervention) in a particular population is relevant to a different population (the trial population). The second assumption, which is probably more debatable, is that the same risk function (i.e., with the same variables and coefficients) will remain valid in the presence of intervention on risk factors. That is, the only effect of the intervention is to change levels of risk factors; the relationship between incidence and risk factors is unchanged.* It should thus be kept in mind that the basic assumption that the risk function coefficients for control and treatment groups would be the same could be misleading. Nevertheless, this assumption is made here because of lack of data that allow a more justifiable approach to estimating anticipated risk reduction. It is also made because of the view that the answer to such a question is a more suitable basis for trial design than the arbitrary specification of a level of risk reduction, which may be, in fact, totally unattainable by the proposed intervention. A more defensible approach to estimating potential risk reduction would be to use the results, if available, from similar trials to the one being contemplated.

The impact of assumptions 7–9, listed above, on sample size requirements is illustrated in table 4. For a single-factor trial of cholesterol lowering with the goal of a 20% reduction from a mean baseline level of 250 mg/dl, the required sample size of 5142 per treatment group (table 2) is increased to 8341 when the follow-

---

*This assumption is similar to that sometimes made in the analysis of natural history data relating incidence of some event to risk factors, i.e., that cases and non-cases arise from multivariate normal distributions with the same covariance matrix (same relationship among risk factors) but different mean values of risk factors. Quite apart from the normality assumption, it is generally clear from the data that the assumption of variance-matrix equality is unreasonable. Whether or not this is important will depend on the question being asked.
ing are assumed: a 5-year incidence of dropouts of 10%, a 5-year incidence of dropins of 10% and a 15-month period (one-fourth of the follow-up period) before the intervention achieves maximum therapeutic benefit. By further changing these assumptions (and for many situations making them even more realistic) to 20% dropout and dropin incidences, and a 30-month period to achieve maximum benefit, the required sample size is increased to 14,462 per treatment group. Similarly, for a two-factor trial with the intervention goal described in table 3, line 4, the required sample size of 3036 per treatment group is increased to 4935 and to 8571 under these two sets of more stringent and realistic assumptions. (The methods for taking into account dropouts, dropins, and time to achieve maximum benefit are described elsewhere.3, 12, 13, 20)

The sample size requirement may be nearly halved by considering an intervention trial on two risk factors simultaneously (table 3). Further reductions in sample size might be effected with a trial in which other CHD risk factors in addition to cholesterol and cigarette smoking are simultaneously targeted for intervention. However, for this to be accomplished, further research is needed to identify such additional factors that are definitely associated with increased risk of mortality in survivors of MI and have the potential of attenuation through appropriate intervention.

No large-scale intervention trials of cholesterol and/or cigarette smoking in post-MI men are being planned, to our knowledge. The analysis and data presented in this paper suggest that the sample size requirements to insure the success of an intervention trial on cholesterol and/or cigarette smoking could be quite prohibitive. The success of trials with affordable and attainable numbers of patients depends in large measure on the development of intervention strategies that are more effective than those discussed in this report. The greatest hope for feasibility clearly lies in the direction of intervention trials focused on several CHD risk factors simultaneously. In the absence of definitive data from such randomized controlled trials, physicians will have to continue to advise their patients with CHD on the basis of their best judgment of the available evidence. The CDP data presented here and elsewhere support the conclusion that measures that can safely reduce elevated serum cholesterol levels and interventions that lead to cessation of cigarette smoking may be useful as part of the total therapeutic regimen to improve long-term prognosis after myocardial infarction.4, 31, 32

References


Table 4. Required Sample Sizes for Coronary Heart Disease Risk Factor Intervention Trials, 5-year Follow-up Period, $\alpha = 0.05$, $\beta = 0.10$, Various Assumptions

<table>
<thead>
<tr>
<th>% dropout incidence</th>
<th>% dropin incidence</th>
<th>Time (months) to achieve max. benefit</th>
<th>Estimated 5-year mortality</th>
<th>Required sample size per treatment group</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2056</td>
<td>0.18414</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.2056</td>
<td>0.18541</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0.20829</td>
<td>0.18414</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0.2056</td>
<td>0.18732</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>15</td>
<td>0.20845</td>
<td>0.1843</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>30</td>
<td>0.20765</td>
<td>0.19240</td>
</tr>
<tr>
<td>B.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.20864</td>
<td>0.17586</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.20864</td>
<td>0.17750</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0.20700</td>
<td>0.17586</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0.20864</td>
<td>0.17906</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>15</td>
<td>0.20721</td>
<td>0.18139</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>30</td>
<td>0.20618</td>
<td>0.18651</td>
</tr>
</tbody>
</table>


30. Schork MA, Remington RD: The determination of sample size in treatment-control comparisons for chronic disease studies in which dropout or non-adherence is a problem. J Chronic Dis 20: 233, 1967


Appendix

Derivation of Confidence Interval Estimate of Required Sample Size

Let $x_i$ denote the value of the risk factor for all patients in the control group and $x_j$ the value for all patients in the intervention group. Once again, define $\pi_i = [1 + \exp(-a - bx_i)]^{-1}$, $i = 1, 2$, and $\bar{x} = (\pi_x + \pi_0)/2$, and also define $w = (\pi_x - \pi_0)[4\pi(1 - \bar{x})]^{-1}/2$. By simple algebraic manipulation, formula 3 in the Methods section of this report can be written as

$$n = \left[ Z_{1-a} (1-w)^{1/2} + Z_{1-a/2} \right]^{2/\Delta^2}$$

The first objective is to obtain the variance of $w$ and hence 95% confidence limits for $w$. These limits will then be substituted in equation 5 to obtain 95% confidence limits for $n$. The expression for $w$ can be simplified by taking first order Taylor series expansions of $\pi_i$ about $x = \bar{x}$ and of $\pi_0$ about $x = \bar{x}$, where $\bar{x} = (x_i + x_j)/2$. Thus is obtained

$$\pi_i = \frac{1}{1 + Q} + \frac{bQ}{(1 + Q)^2}(x_i - \bar{x})$$

where $Q = \exp(-a - bx)$. From equation 6, $\pi_i - \pi_0 \approx bQ(x_i - x_j)/(1 - Q)^2$ and $\tilde{w} = (1 + Q)^{-1}$. Thus,

$$w = \frac{bQ^2}{2(1 + Q)}(x_i - x_j)$$

For $(x_i - x_j)$ reasonably small, the approximation (formula 7) for $w$ is very good. It can be shown that for $b(x_i - x_j) \leq 0.351$ this approximation formula for $w$ always yields values which are within 1% of the correct value, and for $b(x_i - x_j) \leq 0.108$ the approximation is always within 0.1% of the correct value. For the example given later in this appendix, $b(x_i - x_j) = 0.080$.

An estimate of $w$, $\tilde{w}$, is obtained by substituting the estimates $\bar{a}$ and $\bar{b}$ in equation 7. Taking a first order Taylor series expansion of $w$ about $\bar{a}$ and $\bar{b}$, the variance of $w$ can be expressed approximately by

$$\text{Var}\tilde{w} = \frac{\partial^2 w}{\partial a^2} \text{Var}\bar{a} + 2 \frac{\partial^2 w}{\partial a \partial b} \text{Cov}(\bar{a}, \bar{b}) + \frac{\partial^2 w}{\partial b^2} \text{Var}\bar{b}$$

It can be shown by carrying this expansion out to the second order that this is an accurate approximation for the problem. Because $\partial Q/\partial a = -Q$ and $\partial Q/\partial b = -3Q$, it can easily be shown that

$$\frac{\partial w}{\partial a} = \frac{b(x_i - x_j)}{4} \left[ \frac{Q^1 - Q^2}{(1 + Q)^2} \right]$$

and

$$\frac{\partial w}{\partial b} = \frac{b(x_i - x_j)}{2} \left[ \frac{Q^1 - Q^2}{1 + Q} \right] + \frac{\partial b(x_i - x_j)}{4} \left[ \frac{Q^1 - Q^2}{(1 + Q)^2} \right]$$

Consider the risk factor serum cholesterol. From the logistic regression analysis of CDP data, $\bar{a} = -2.13240$ and $\bar{b} = 3.21927 \times 10^5$. Taking $x_i = 250$ and $x_j = 225$, we obtain $Q = 3.9268$. Also, for these values of $\bar{a}$, $\bar{b}$, and $w$, $w = 0.016185$. $\partial w/\partial a = 4.8075 \times 10^5$ and $\partial w/\partial b = 6.1694$. Taking the values of $\text{Var}(\bar{a})$, $\text{Var}(\bar{b})$, and $\text{Cov}(\bar{a}, \bar{b})$ found in the Results section, we obtain from equation 8, $\text{Var}w = 2.35073 \times 10^5$. Thus, 95% confidence limits on $w$ are $w \pm 1.96(\text{Var}w)^{1/2}$ or $(0.006682, 0.025688)$. Finally, substituting these limits of $w$ in equation 5, we have $(7962, 117,685)$ as 95% confidence limits for $n$.

Participants

Policy Board

Robert W. Wilkins, M.D. (Chairman); Jacob E. Bearman, Ph.D.; William M. Smith, M.D., M.P.H.; Christian R. Klintm, M.D., Dr. P. H. (ex officio); Jeremiah Stamler, M.D. (ex officio); Max Halperin, Ph.D. (ex officio); William Zukel, M.D. (ex officio).

Past members: Edwin Boyle, M.D. (deceased); Louis Lasagna, M.D.
Steering Committee

Coordinating Center
Paul L. Canner, Ph.D. (Principal Investigator); Christian R. Klimt, M.D., Dr. P. H. (Co-Investigator); Sandra Forman, M.S.; Elizabeth C. Heinz; Genell L. Knatterud, Ph.D.; William F. Krol, Ph.D., Frances LoPresti, Curtis L. Meintert, Ph.D. Past members: Robert S. Gordon, Jr., M.D.; Yih-Min Bill Huang, Ph.D.; David R. Jacobs, Jr., Ph.D.; Gerard Prud'homme, M.A.; Suketami Tominaga, M.D.

Central Laboratory
Joseph Boutwell, M.D. (Medical Director); Dayton Miller, Ph.D. (Chief); John Donahue, M.S.; James Gill, Jr., M.S.; Sara Gill, Past members: Gerald R. Cooper, M.D.; Eloise Eavenson, Ph.D.; Adrian Hairline, M.D.; Alan Mather, Ph.D.; Margie Sailors.

ECG Center
Henry Blackburn, M.D. (Director of the Laboratory of Physiological Hygiene); Ronald J. Prineas, M.B., B.S., Ph.D.; (Director of the ECG Center); David R. Jacobs, Jr., Ph.D.; Gretchen Newman. Past Member: Robin MacGregor.

National Heart, Lung, and Blood Institute Staff

Drug Procurement and Distribution Center
Salvatore Gasdia (Officer in Charge).

Editorial Review Committee

Principal Investigators, Clinic Research Centers
Implications of findings in the coronary drug project for secondary prevention trials in coronary heart disease. The coronary drug project research group.

P L Canner and M Halperin

_Circulation_. 1981;63:1342-1350
doi: 10.1161/01.CIR.63.6.1342

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1981 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/63/6/1342.citation

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/