American Journal of Epidemiology
Vol 136, No. 9

Printed in U.S.A.

Sample Size for Studying Intermediate Endpoints within Intervention Trials or Observational Studies

Laurence S. Freedman[1] and Arthur Schatzkin[2]

An intermediate endpoint is a biologic event or marker that is a precursor to a given health outcome. Examples of potential intermediate endpoints include serum cholesterol for coronary heart disease, endogenous steroid hormones for breast cancer, and CD4 count for acquired immunodeficiency syndrome. When one is studying a potential intermediate endpoint in the context of an intervention trial, five types of questions may be investigated: 1) Does the intervention affect the intermediate endpoint? 2) Is the intermediate endpoint associated with prognostic or risk factors? 3) Is the intermediate endpoint associated with the main outcome? 4) Is the intervention effect on the main outcome mediated by the intermediate endpoint? 5) Are the prognostic or risk factor effects mediated by the intermediate endpoint? In this paper, the authors show that each of these questions has different sample size requirements, and they illustrate their point with a discussion of an ancillary study of large bowel epithelial proliferation in the National Cancer Institute's Polyp Prevention Trial. The same methods may be used in an observational study, in which case questions 2, 3, and 5 are relevant. However, much larger numbers than those used in the Polyp Prevention Trial example will be required when the main outcome is rare. *Am J Epidemiol* 1992;136:1148–59.

adenoma; biological markers; colonic neoplasms; colonic polyps; diet therapy; risk factors

An intermediate endpoint is a biologic event or marker which acts as a precursor to an event of primary interest, such as the development of a certain chronic disease. Epidemiologists are increasingly interested in studying potential intermediate endpoints, since identifying a valid intermediate endpoint can represent a major advance in understanding the etiology of a disease and ultimately lead to its prevention. Examples which have received much attention in recent years have been the role of serum cholesterol levels in the development of coronary heart disease and the role of endogenous steroid hormones in the etiology of breast cancer. Clinical investigators have also been interested in intermediate endpoints as possible surrogates for treatment success or failure. Here the event of primary interest may be death or the recurrence of disease. The serum CD4 count, which appears to decline in rough parallel with the worsening course of acquired immunodeficiency syndrome, is one example of a potential surrogate endpoint.

Criteria for the validation of a potential intermediate endpoint in intervention trials or observational studies have been discussed by Susser (1), Prentice (2), Schatzkin et al. (3), and Freedman et al. (4). Machado et al.

Received for publication July 29, 1991, and in final form April 13, 1992.

[1] Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.

[2] Cancer Prevention Studies Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.

Reprint requests to Laurence S. Freedman, Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Executive Plaza North, Suite 344, Bethesda, MD 20892.

The authors thank Dr. Barry Graubard for his valuable comments and Drs. Mark Shiffman and Sandy Dawsey, who helped to formulate the basic concepts for evaluating intermediate endpoints. Dorothy Dowty prepared the manuscript.

(5) have investigated the use of an intermediate endpoint as a surrogate endpoint for death in the context of a clinical trial of treatment for acquired immunodeficiency syndrome. Generally, authors have been cautious about the use of intermediate endpoints as surrogate endpoints, mainly because the effect of an intervention upon an intermediate endpoint may not always reflect its effect on the main outcome. Nevertheless, incorporation of an intermediate endpoint determination within the framework of a randomized trial or observational study can provide useful information regarding the relation between the intermediate endpoint, the intervention or exposure, and the main outcome.

In this paper, we identify the questions which may be addressed in such an exercise, and we show how the ability to answer these questions depends on the number of subjects in the study and the proportion of subjects in whom the intermediate endpoint is determined. These topics are discussed in the context of a particular trial, the National Cancer Institute's Polyp Prevention Trial.

POLYP PREVENTION TRIAL

The Polyp Prevention Trial is a randomized trial designed to evaluate the effect of dietary modification on the recurrence of adenomatous polyps in patients who have recently had removed an adenomatous polyp of the colon or rectum. The aim is to enter 2,000 patients from 10 clinical centers in the United States and to randomly allocate half of the patients to intensive dietary counseling and half to their usual diet. Dietary counseling is to be provided on an individual basis. Each patient is set the goal of reducing fat intake to 20 percent of total caloric intake and increasing fiber intake to 18 g/kcal and fruit and vegetable intake to 5–8 servings per day. Recruitment began in the spring of 1991. All patients will be followed for 4 years and will receive repeat colonoscopies at the end of the first and fourth years.

The underlying motivation for the trial is to further clarify the hypothesis linking diet to colorectal cancer. There is considerable epidemiologic evidence supporting a direct association of colorectal cancer incidence with fat consumption (6, 7) and supporting inverse associations with intake of fruit, vegetables, and fiber (8–10). Testing this hypothesis directly in an intervention trial would require studying several tens of thousands of patients followed for at least a decade, since the incidence of colorectal cancer in the population is relatively low.

Most colorectal adenocarcinomas are thought to develop from adenomatous polyps (11, 12). Thus, prevention of the formation of polyps should lead to prevention of colon cancer. The Polyp Prevention Trial was designed to test whether manipulating the dietary factors associated with colorectal cancer can reduce the recurrence rate of colonic polyps. If this can be shown, it will greatly strengthen the hypothesis linking diet to colorectal cancer incidence.

The main outcome for the trial is therefore the recurrence of one or more adenomatous polyps. Because the proportion of patients who have a recurrence over the 4-year follow-up period is expected to be quite high (approximately 27 percent in the control group), the number of patients required in the trial is much below the several tens of thousands that would be required for a trial using cancer as the outcome. The total of 2,000 patients gives a statistical power of 90 percent for detecting (using a one-tailed statistical test) a reduction of approximately 20 percent in the annual recurrence rate among the patients given dietary counseling. Such a reduction would lead to approximately 21 percent of the patients given dietary counseling having a recurrence, compared with the 27 percent in the control group.

Several biologic mechanisms for the formation of polyps and subsequent development of cancer have been proposed (13–15). The initiation of a major trial such as the Polyp Prevention Trial affords a special opportunity to study some of these proposed mechanisms. Cell proliferation (16), the mucin characteristics (17), and ornithine decarboxylase levels (18) are examples of biologic markers that may all be determined from

normal colorectal tissue, and these may be studied as intermediate endpoints for the development of recurrent colorectal polyps. As the illustrative example for our paper, we consider the cell proliferation rate (measured by the labeling index (19)) as the intermediate endpoint; the main outcome is recurrence of an adenomatous polyp, and the intervention is dietary counseling.

TYPES OF QUESTIONS THAT MAY BE ADDRESSED

By incorporating an intermediate endpoint into a randomized or observational study, we can address five types of questions relating to the intermediate endpoint and its putative role in the disease process. We list these questions in the first column of table 1 and discuss them below. Questions 1 and 4 are specific to intervention trials; the others are common to both intervention and observational studies.

Does the intervention affect the intermediate endpoint?

For the Polyp Prevention Trial, the question is "Does dietary modification affect a patient's cell proliferation rate?" The randomized design allows a controlled assessment of this question. Suppose cell proliferation rates were assessed 1 year after entry into the trial. Then rates in the dietary counseling group could be compared with rates in the control group. Randomization would allow us to ascribe a large significant difference in the mean proliferation rates as an effect of the dietary modification. If baseline measurements of cell proliferation rates were also made, then *changes* in the cell proliferation rates over the 1-year period could be compared between the two groups. Examining changes in rates usually provides greater statistical power to detect any intervention effect, so the inclusion of baseline measures would give a more precise comparison. The extent of the improvement increases with the strength of the correlation between baseline and 1-year rates.

Addressing this question is only indirectly relevant to the relation of the intermediate endpoint to the primary endpoint. If we already knew that a given diet reduced polyp recurrence and we were then able to show that the same diet modified cell proliferation rates, this would point toward a causal relation between cell proliferation rates and polyp recurrence. However, such knowledge of the diet-polyp relation is the main topic of the trial, so the interpretation of any relation between diet and cell proliferation is dependent on the main result of the trial.

Is the intermediate endpoint associated with prognostic or risk factors?

Several factors, assessed at baseline in the Polyp Prevention Trial, have been established to be correlated with polyp recurrence. These include number of previous

TABLE 1. Questions that may be posed regarding the intermediate endpoint, cell proliferation, in the Polyp Prevention Trial, and the measurements, patient groups, and sample sizes required for each question*

Question	Cell proliferation measured at:	Patient group	Sample size required (*N*)
1. Diet → Cell proliferation	1 year	All	64†
2. Polyp multiplicity ↔ Cell proliferation	Baseline	All	67
3. Cell proliferation → Recurrence	1 year	All	
	or		222
	Baseline	Control	
4. Diet → Cell proliferation → Recurrence	1 year	All	2,140‡,§
5. Polyp multiplicity ↔ Cell proliferation → Recurrence	Baseline	Control	143§

* Based on a one-sided significance test at the 5% level and 90% statistical power.
† *N* could be smaller if the baseline intermediate endpoint was also measured
‡ With *N* = 2,000, power = 88%.
§ Numbers required to show that the intermediate endpoint explains at least one half of the effect.

polyps and polyp size (20). Finding the cell proliferation rate to be associated with some or all of these prognostic factors would add to the notion that cell proliferation is involved in the process of polyp recurrence.

However, as with the previous question, the evidence is somewhat indirect. It would be possible for the cell proliferation rate to be associated with a prognostic or risk factor but nevertheless to be unrelated to polyp recurrence itself.

An important point is that this question is best posed only in the control group of patients using 1-year determinations of cell proliferation rate, or for all patients but using baseline determinations of the cell proliferation rate. The dietary intervention may affect the 1-year proliferation rate so that determinations of the rate after the start of the intervention may not be suitable for relating to the baseline prognostic factors.

Association between intermediate endpoints and risk factors may also be investigated in observational studies. For both randomized and observational studies, we stress the use of the word "association," since we often have no idea of the chronologic order of events or the underlying mechanisms that might be responsible for the relation.

Is the intermediate endpoint associated with the main outcome?

An important and more direct question than those above is whether the cell proliferation rate is associated with polyp recurrence. Do patients with high proliferation rates have higher rates of recurrent polyps? In the Polyp Prevention Trial, this question may be addressed principally by relating the cell proliferation rate at 1 year to the development of recurrence between years 1 and 4. This analysis can include patients in the intervention group as well as in the control group, assuming that dietary intervention will have caused by 1 year any change in the cell proliferation rate that was going to occur. As a precaution, in estimating the relation between proliferation rate and recurrence, we might control for the allocated treatment group (intervention or control). An alternative analysis using baseline proliferation rates rather than 1-year rates could be conducted within the control group; however, analysis using baseline rates within the intervention group could be invalid if the intervention affected cell proliferation rates. In an observational cohort study, baseline proliferation rates are assessed in relation to subsequent polyp recurrence (M. Wargovitch, University of Texas M. D. Anderson Cancer Center (Houston, Texas), personal communication, 1991).

Some investigators have advocated relating the change in the intermediate endpoint to the main outcome. We have less preference for this analysis, unless there is a clear biologic reason for expecting the main outcome to be affected more by the change in intermediate endpoint then by the actual level of the intermediate endpoint.

Generally, in observational studies, the timing of the intermediate endpoint assessment is important. When the intermediate endpoint assessment precedes the main outcome event (cohort study), the hypothesis that the intermediate endpoint is a precursor of the event can clearly be tested. When the intermediate endpoint assessment is concurrent with the event or after the event (case-control studies), then the possibility of reverse causation (i.e., the disease's affecting the intermediate endpoint) has to be considered.

Does the intermediate endpoint mediate the intervention effect?

Should the dietary intervention reduce the rate of polyp recurrence, it will be of considerable interest to examine whether this effect can be explained by concomitant changes in the cell proliferation rate. This question may be addressed by examining the intervention effect after adjustment for cell proliferation rate (3). The intervention effect will tend to disappear (become zero or be reduced) after such adjustment, if the cell proliferation rate does mediate the effect of dietary counseling.

Ability to obtain reliable information from such an analysis depends on the strength of the dietary intervention effect.

Clearly, if there were zero effect on the recurrence rate, then there could be no mediating variable. Likewise, if the effect were not statistically significant, little extra information would be likely to accrue from the adjusted analysis. The most satisfactory situation for addressing this question is when there is a highly significant intervention effect (4).

The analysis requires 1-year cell proliferation rates as the intermediate endpoint and recurrence between 1 and 4 years as the main outcome. Baseline rates, since they are not affected by the dietary intervention, would not be useful here.

Does the intermediate endpoint mediate the prognostic or risk factor effects?

As with the previous question, this can be addressed by considering prognostic or risk factor effects adjusted for cell proliferation rate. Unlike the dietary intervention effect, one can feel confident that prognostic factor effects will be found in the Polyp Prevention Trial, and it is possible that some of these effects will be highly statistically significant.

It is necessary to restrict analysis to the control group when addressing this question, since dietary intervention could affect both proliferation rates and recurrence rates and so interfere with the three-way relation between prognostic factors, cell proliferation, and polyp recurrence. Either baseline or 1-year determinations may be used.

SAMPLE SIZE CALCULATIONS

To perform sample size calculations for each of the five questions considered above, we need to make certain assumptions regarding the distribution of cell proliferation rates, the sizes of the various effects that we wish to test, the statistical significance level, and the statistical power required.

Lipkin et al. (19) published data on the cell proliferation labeling index of 53 subjects. Nine subjects had adenomatous polyps, and the mean labeling index was 9.5 percent with a standard deviation of 2.7 percent. Other groups studied by these authors had higher standard deviations (low risk, 4.4 percent; cancer, 5.8 percent), so we conservatively assume the true standard deviation for this group of patients to be equal to 5.0 percent. Generally, the distributions of labeling index among populations appear skewed to the right, so assuming a lognormal distribution is reasonable. This implies that the log labeling index is normal. The mean m and variance v of the log labeling index may be obtained by solving the simultaneous equations (21):

$$9.5 = \exp(m + v/2)$$

$$5.0^2 = (\exp(v) - 1)\exp(2m + v)$$

with solution $m = 2.13$ and $v = 0.2445$. The median labeling index is given by $\exp(m) = \exp(2.13) = 8.4$ percent.

For the purposes of this paper, we shall assume that statistical tests are conducted with a one-sided significance level α of 5 percent and that a power $1 - \beta$ of 90 percent is required to detect specified effects. The corresponding normal deviates are $Z_\alpha = 1.64$ and $Z_\beta = 1.28$.

Does the intervention affect the intermediate endpoint?

Alberts et al. (22) have reported that labeling index was reduced by 22 percent after an 8-week wheat bran diet. The intervention in the Polyp Prevention Trial is over a prolonged period and involves fat reduction as well as fiber increase. We therefore assume that the intervention reduces 1-year labeling index levels in all subjects by 30 percent, i.e., to 70 percent of their pre-intervention mean levels. This would lead to a 1-year mean labeling index of 6.65 percent (9.5×0.7) and standard deviation of 3.5 percent (5.0×0.7), with new solutions for m and v: $m = 1.77$ and $v = 0.2445$. Notice that the value of v does not change from that above, but m is reduced.

When there are no baseline measurements of labeling index, the effect of intervention on labeling index may be tested using a two-sample t test of the 1-year log labeling index in the control and intervention groups. Required total sample size N is given by

$$N = \frac{4(Z_\alpha + Z_\beta)^2 v}{(m_1 - m_2)^2}, \qquad (1)$$

where v is the common variance of 0.2445 and $m_1 - m_2$ is the difference between the control and intervention means on the log scale. Here $m_1 = 2.13$ and $m_2 = 1.77$. This formula gives $N = 64$ (see table 1).

When baseline measurements are available, one may compare the change in log labeling index from baseline to 1 year in the intervention group to that in the control group. If there is considerable interperson variation in the labeling index, then examining changes rather than absolute levels may reduce the variance considerably. For example, if the variance of the change in log labeling index were reduced to $0.5v$, then the number required would be only one half that of the N calculated above (see equation 1). Including baseline measurements does often lead to increased precision. However, without information on the variance of changes in the labeling index, the exact savings in sample size cannot be reliably predicted, and then it is best to stay with the more conservative sample size based on the test of absolute levels of labeling index.

Is the intermediate endpoint associated with prognostic factors or risk factors?

Several investigators have reported that patients with multiple polyps have higher recurrence rates than those with single polyps (20, 23, 24). Let us assume, similarly to the report of Neugut et al. (20), that 60 percent of patients have single polyps and 40 percent have multiple polyps. To test whether baseline labeling index is higher in the group with multiple polyps, we may compare the baseline labeling indexes in the two groups using a two-sample t test.

The sample size formula given for question 1 needs to be modified to account for the fact that the two prognostic groups are not the same size. Let f be the proportion in the single polyp group and let N be the total sample size required. Then

$$N = \frac{(Z_\alpha + Z_\beta)^2 v\left(\frac{1}{f} + \frac{1}{1-f}\right)}{(m_2 - m_1)^2} \qquad (2)$$

Here m_1 and m_2 are the mean log labeling indexes in the single and multiple polyp groups, respectively, and as in the previous question, one group has labeling indexes which are 30 percent lower than those of the other group. Then $m_2 - m_1 = 0.36$ and $v = 0.2445$ as above. Application of the formula leads to $N = 67$, a very similar result to that obtained for question 1 (see table 1).

Is the intermediate endpoint associated with the main outcome?

Since the labeling index is a continuous variable, the usual statistical method of relating labeling index to the dichotomous variable polyp recurrence is through the logistic regression model. Methods of calculating sample size for logistic regression models have been developed (25) and could be used for our problem. However, a simpler approach is to consider the difference between the mean log labeling indexes in the patients with polyp recurrence and those without. We may then use equation 2 for the sample size N with f equal to the fraction of patients with recurrence, m_1 and m_2 equal to the mean log labeling index among the patients with and without recurrence, respectively, and v equal to the variance of the mean log labeling index within these groups of patients.

Values of m_1, m_2, and v must be chosen to be consistent with the assumptions that the recurrence rates in the control and intervention groups are 27 and 21 percent, respectively, and that the intervention reduces the labeling index by 30 percent. The method of finding these values of m_1, m_2, and v is described in the first part of the Appendix.

Within the control group, the method described in the Appendix leads to $m_1 = 2.29$, $m_2 = 2.07$, and $v = 0.24$, with $f = 0.27$. Within the intervention group, the method determines that $m_1 = 1.95$, $m_2 = 1.72$, and $v = 0.24$, with $f = 0.21$. Normally, for an analysis stratified by treatment group, we could not use equation 2 directly to calculate sample size. However, since $m_2 - m_1$ (0.22 and 0.23) and f (0.21 and 0.27) are so similar

for each group and since v is the same, we may use equation 2 with $m_2 - m_1 = 0.225$ (the average over the two groups), $v = 0.24$, and $f = 0.24$ (the average over the two groups) with very little loss of accuracy. This formula yields a sample size estimate of $N = 222$ (see table 1).

Is the intervention effect mediated by the intermediate endpoint?

Freedman et al. (4) discuss a statistical criterion proposed by Prentice (2) for the validation of intermediate endpoints. Suppose there is some intervention that affects an event such as (occurrence or) recurrence of disease. In a statistical analysis relating intervention to recurrence of disease, one would see a significant intervention effect, assuming sufficient numbers of patients were studied. If the intermediate endpoint were to mediate the effect of the intervention, then the intervention effect should disappear when adjusted for the intermediate endpoint. Hence, the statistical validation procedure involves investigation of the intermediate endpoint-adjusted intervention effect. In reality, because of interpatient variation, one cannot usually establish that the intermediate endpoint mediates totally the intervention effect. Freedman et al. (4) propose a quantified measure of how much of the intervention effect is mediated by the intermediate endpoint, given by

$$\hat{M} = 1 - (\hat{\tau}_{1a}/\hat{\tau}_1)$$

where $\hat{\tau}_1$ is the unadjusted estimate of the intervention effect and $\hat{\tau}_{1a}$ is the estimate of the intervention effect adjusted for the intermediate endpoint. The lower confidence limit $\hat{M}_L$ of $\hat{M}$ places a plausible lower bound on how much of the intervention effect is predicted by the intermediate endpoint, and one may wish to show that $\hat{M}_L$ is greater than, say, 0.5 or 0.75. By calculating the probability that $\hat{M}_L$ is greater than some chosen fraction h, Freedman et al. (4) lay the basis for sample-size calculations for this analysis. In fact, it may be shown that if $\hat{M}_L$ is the lower $100(1 - 2\alpha)$ percent confidence limit, h is the quantity that we require $\hat{M}_L$ to exceed ($h = \frac{1}{2}$ or $\frac{3}{4}$, usually), and $1 - \beta$ is the required power for rejecting the hypothesis $M = h$, then

$$Z_\alpha + Z_\beta = (1 - h)(\hat{\tau}_1/SE(\hat{\tau}_1))/ (2(1 - \rho)(1 - h) + h^2)^{1/2},$$

where ρ is the correlation between $\hat{\tau}_{1a}$ and $\hat{\tau}_1$. Since $SE(\hat{\tau}_1)$ depends on sample size N, we may use the above relation to calculate N. The value of ρ may be estimated using a method given in the second part of the Appendix. Typically, ρ will be positive and large.

Assume as we did in planning the trial that the recurrence proportion among control patients is 27 percent and that the proportion among intervention patients is 21 percent. There are equal numbers of patients in each group. Then the expected value of $\hat{\tau}_1 = 0.27 - 0.21 = 0.06$ and

$$SE(\hat{\tau}_1) = ((0.27)(0.73)/0.5N + (0.21)(0.79)/0.5N)^{1/2} = 0.85/\sqrt{N}.$$

In addition, ρ is estimated to be 0.94 (see part 2 of the Appendix).

Hence, using the above equation,

$$N = \left(\frac{0.85}{0.06}\right)^2 \frac{(Z_\alpha + Z_\beta)^2}{(1 - h)^2}(0.12(1 - h) + h^2)$$

with $Z_\alpha = 1.64$ and $Z_\beta = 1.28$.

For $h = \frac{1}{2}$, $N = 2{,}140$ patients are required, slightly above the planned sample size (table 1). We may ask with what statistical power can the question (with $h = \frac{1}{2}$) be addressed when $N = 2{,}000$, the planned sample size. To answer this, we substitute $N = 2{,}000$ in the above equation and solve for Z_β. This results in the solution $Z_\beta = 1.18$, and the statistical power is 88 percent. Hence, there is good power to show that cell proliferation rates explain at least one half of the intervention effect, in the event that the intervention effect is the size anticipated and the effect is entirely mediated by cell proliferation. However, similar calculations show that the power to show that cell proliferation rates explain at least three fourths of

the intervention effect ($h = 0.75$) is only 0.27.

Is the prognostic or risk factor mediated by the intermediate endpoint?

The same analysis as that applied above may be used to determine whether the effect of a prognostic factor or risk factor is mediated by the cell proliferation rate. In this case, the intervention and control groups are replaced in the analysis by two prognostic factor groups.

Suppose our prognostic factor, as before, is the dichotomous variable representing single or multiple polyps and that 60 percent of patients present with single polyps. We assume a relative risk for recurrence of 2.5 for patients with multiple polyps versus single polyps. This determines that the recurrence proportion is 17 percent for single polyp patients but 42 percent for multiple polyp patients. Thus, the expected value of $\hat{\tau}_1 = 0.42 - 0.17 = 0.25$, and

$$\mathrm{SE}(\hat{\tau}_1) = ((0.17)(0.83)/0.6N + (0.42)(0.58)/0.4N)^{1/2} = 0.92/\sqrt{N}.$$

The estimated value of ρ is 0.94 (see part 2 of the Appendix).

Hence, using the above equation, we obtain

$$N = \left(\frac{0.92}{0.25}\right)^2 \frac{(Z_\alpha + Z_\beta)^2}{(1-h)^2}(0.12(1-h) + h^2).$$

For $h = 1/2$, $N = 143$ patients, all of whom must be in the control group (table 1). For $h = 3/4$, $N = 1{,}095$ control patients, slightly greater than the number planned for the trial. Hence, including all of the control subjects in the intermediate endpoint study would achieve just under 90 percent power for $h = 3/4$.

DISCUSSION

Measurement of intermediate endpoints within the framework of a randomized trial or observational study can yield valuable additional information regarding the disease process and the mechanism of action of the intervention. Such measurements can also be uncomfortable to the participants and expensive and may increase the organizational complexity of the trial. Therefore, they are not to be undertaken lightly. When such measurements are possible, it may be feasible to make them only on a subset of patients, and the number of determinations per patient may also be limited to one or two. However, with a smaller number of patients and fewer occasions on which intermediate endpoint measurements are made, we have less power with which to answer the questions of interest. It is therefore important to set out different questions related to intermediate endpoint measurements at the design stage and to consider the sample sizes and the number of assessments which may be required to answer them. Table 1 summarizes the questions, measurements, and sample sizes needed for studying cell proliferation rates within the Polyp Prevention Trial, as described in this paper.

Measuring the cell proliferation rate requires taking biopsies of normal rectal mucosa at an examination in addition to the regular colonoscopy. The preparation of the biopsy tissue for the assay requires specially trained personnel. These practical considerations may make it difficult to include more than 50 percent of the trial patients in the intermediate endpoint study, i.e., more than 1,000 patients.

From table 1, we see that questions 1–3 and 5 may be reliably addressed with less than 300 patients. Thus, we should be able to discover the effect of diet on cell proliferation rates, the relation between cell proliferation and established prognostic factors, and the relation between cell proliferation and subsequent recurrence. In addition, for strongly prognostic factors, we may be able to assess whether cell proliferation mediates their effect. However, the question of whether cell proliferation mediates the effect of diet requires larger numbers of patients.

Generally, in an intervention trial, the questions regarding mediation of prognostic

factor effects will be more readily addressed than the question of mediation of the intervention effect. This is because, in most disease settings, established prognostic factors or risk factors have a greater influence on outcome than does intervention. In the Polyp Prevention Trial, the questions regarding mediation of prognostic factor effects could be addressed quite reliably (90 percent power for $h = \frac{1}{2}$) by including approximately 14 percent of the trial's control patients in the intermediate endpoint study (i.e., 140 patients). Note that doing so would not provide any information on the effect of diet on cell proliferation, so a further group of patients in the intervention group would be required to study question 1. However, study of the same control patients would answer question 2 as well as question 5, and using a further 80 control patients would also enable one to address question 3.

Table 1 indicates that all of the patients included in the trial would need to be studied to attain reliability in answering whether cell proliferation mediates the intervention effect (question 4). Were all 2,000 trial patients included in the intermediate endpoint study, then the estimated power to detect such mediation is 88 percent; by "detecting such mediation," we mean showing that the intermediate endpoint explains at least half of the intervention effect ($h = \frac{1}{2}$). This power of 88 percent is calculated assuming that the diet reduces the polyp recurrence proportion from 27 percent to 21 percent. If the effect were larger, then the power to detect mediation by the cell proliferation rate would also be greater. Whether or not it is worthwhile including all trial patients in the intermediate endpoint study will depend ultimately upon the level of discomfort, expense, and inconvenience caused by the intermediate endpoint measurement.

Our analysis for checking whether the intervention effect is mediated by the intermediate endpoint relies on a comparison of the adjusted estimate of the intervention effect with the unadjusted estimate. If mediation of the intervention effect by the intermediate endpoint occurs, then the adjusted estimate has the expected value of zero. Thus, mediation implies that the expected value of $\hat{M}$ is 1. Finding that $\hat{M}$ is equal to 1, with a narrow confidence interval, would strongly support the concept of mediation, but such a demonstration cannot be regarded as *proof* that the intermediate endpoint indeed lies on a causal pathway between the intervention and the disease event. Evidence of this type must be interpreted alongside the results of other epidemiologic and laboratory studies in deciding whether such a causal pathway does operate. However, we believe that in circumstances where the weight of evidence extraneous to the trial already indicated the existence of such a causal pathway, a clear positive result from such a mediation analysis would be very convincing.

The example we have chosen to present in this paper is that of a randomized intervention trial. However, many of the same considerations apply to examining intermediate endpoints within an observational study, and questions 2, 3, and 5 are all relevant. There are also some differences between our Polyp Prevention Trial example and the examination of intermediate endpoints in a typical observational study. Firstly, if the study is a prospective follow-up of a cohort and the disease is rare, then the numbers of subjects required for all but question 1 or 2 will be much larger than for the Polyp Prevention Trial. This would also be true for an intervention trial with a rare event, such as disease incidence, for the main outcome. Secondly, observed mediation by an intermediate endpoint of an intervention effect is somewhat easier to interpret than observed mediation of a risk factor effect, since, in a randomized study, it is known that the observed intervention effect is not a result of confounding, whereas in an observational study, the possibility of confounding must always be borne in mind. Thirdly, if the observational study has a case-control design, reverse causation is often a real possibility that needs to be excluded to infer true mediation of a risk factor effect.

Investigating intermediate endpoints in subsets of patients participating in randomized trials or observational studies can an-

swer many questions of interest. Investigators should be aware of the potential of nesting an intermediate endpoint study within such investigations.

REFERENCES

1. Susser M. Causal thinking in the health sciences. New York: Oxford University Press, 1973.
2. Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med 1989; 8:431–40.
3. Schatzkin A, Freedman LS, Schiffman MH, et al. Validation of intermediate endpoints in cancer research. J Natl Cancer Inst 1990;82:1746–52.
4. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 1992;11:167–78.
5. Machado SG, Gail MH, Ellenberg SS. On the use of laboratory markers as surrogates for clinical endpoints in the evaluation of treatment for HIV infection. J Acquir Immune Defic Syndr 1990;3: 1065–73.
6. Drasar BS, Irving D. Environmental factors and cancer of the colon and breast. Br J Cancer 1973; 27:167–72.
7. Kolonel L. Fat and colon cancer: How firm is the epidemiologic evidence? Am J Clin Nutr 1987;45: 336–41.
8. McKeown GE, Bright-See E. Dietary factors in colon cancer: international relationships. Nutr Cancer 1984;6:160–70.
9. Jensen OM, MacLennan R, Wahrendorf J. Diet, bowel function, fecal characteristics and large bowel cancer in Denmark and Finland. Nutr Cancer 1982;4:5–19.
10. Greenwald P, Lanza E, Eddy G. Dietary fiber in the reduction of colon cancer risk. J Am Diet Assoc 1987;87:1178–88.
11. Berg JW. Epidemiology, pathology and the importance of adenomas. In: Steele G Jr, Burt RW, Winawer SJ, et al, eds. Basic and clinical perspectives of colorectal polyps and cancer. New York: Alan R Liss, Inc, 1988;13–21.
12. Muto T, Bussey JHR, Morson BC. The evolution of cancer of the colon and rectum. Cancer 1975; 36:2251–70.
13. Reddy BS. Diet and excretion of bile acids. Cancer Res 1981;41:3766–68.
14. Kritchevsky D. Dietary fiber and cancer. Nutr Cancer 1985;6:213–19.
15. Yang CS, Newmark HL. The role of micronutrient deficiency in carcinogenesis. Crit Rev Oncol Hematol 1987;7:267–87.
16. Lipkin M. Biomarkers of increased susceptibility to gastrointestinal cancer: new application to studies of cancer prevention in human subjects. Cancer Res 1988;48:235–45.
17. Boland CR. Mucin histochemistry in colonic polyps and cancer. Semin Surg Oncol 1987;3:183–9.
18. Luk GD, Moshier JA, Ehrinpreis MN. Ornithine decarboxylase as a marker for colorectal polyps and cancer. In: Steele GR Jr, Burt RW, Winawer SJ, et al, eds. Basic and clinical perspectives of colorectal polyps and cancer. New York: Alan R Liss, Inc, 1988:227–39.
19. Lipkin M, Enker WE, Winawer SJ. Tritiated-thymidine labelling of rectal epithelial cells in "non-prep" biopsies of individuals at increased risk for colonic neoplasia. Cancer Lett 1987;37:153–61.
20. Neugut AI, Johnsen CM, Forde KA, et al. Recurrence rates for colorectal polyps. Cancer 1985;55: 1586–9.
21. Lindgren BW. Statistical theory. 1st ed. New York: The McMillan Company, 1960:89.
22. Alberts DS, Einspahr J, Rees-McGee S, et al. Effects of dietary wheat bran fiber on rectal epithelial cell proliferation in patients with resection for colorectal cancers. J Natl Cancer Inst 1990;82:1280–5.
23. Wegener M, Börsch G, Schmidt G. Colorectal adenomas: distribution, incidence of malignant transformation and rate of recurrence. Dis Colon Rectum 1986;29:383–7.
24. Olsen HW, Lawrence WA, Snook CW, et al. Review of recurrent polyps and cancer in 500 patients with initial colonoscopy for polyps. Dis Colon Rectum 1988;31:222–7.
25. Hsieh FY. Sample size tables for logistic regression. Stat Med 1989;8:795–802.

APPENDIX

Calculating the Distribution of Labeling Index among Patients with Recurrence and Patients without Recurrence

Let X denote the log labeling index of a patient. We assume, as in the main text, that for patients in the control group, $X \sim N(2.13, 0.2445)$, and for patients in the intervention group, $X \sim N(1.77, 0.2445)$.

Let Y denote the absence ($Y = 0$) or presence ($Y = 1$) of polyp recurrence in a patient. We assume that

$$P(Y = 1 | X = x) = \exp(a + bx)/(1 + \exp(a + bx)),$$

i.e., that the probability of recurrence is linked to the log labeling index by a linear logistic model.

Our first task is to find the values of a and b that will correspond to an overall probability of recurrence equal to 0.27 in the control group and equal to 0.21 in the intervention group. Let $f_C(x)$ denote the normal probability density function of X in the control group, and $f_I(x)$ the same in the intervention group. Then we may write the equations:

$$0.27 = \int_{-\infty}^{\infty} \frac{\exp(a + bx)}{1 + \exp(a + bx)} \times f_C(x)dx$$

and

$$0.21 = \int_{-\infty}^{\infty} \frac{\exp(a + bx)}{1 + \exp(a + bx)} \times f_I(x)dx.$$

There are two unknown quantities, a and b, in these equations. The integrals may be evaluated for known values of a and b by numerical methods. We used a grid search procedure to find the values of a and b that simultaneously satisfy the equations, leading to the solution $a = -3.07$ and $b = 0.951$.

Having determined the values of a and b, we may now calculate the mean and variance of X (i.e., log labeling index) for those patients with recurrence ($Y = 1$) and for those patients without recurrence ($Y = 0$). We do this separately for patients in the control group and patients in the intervention group.

The probability density function of X conditional on $Y = 1$ for control patients ($g_{1C}(x)$) is given by

$$g_{1C}(x) = \frac{f_C(x) \exp(a + bx)}{1 + \exp(a + bx)} \times \frac{1}{0.27}.$$

Similarly, conditional on $Y = 0$, the probability density function of X ($g_{0C}(x)$) is

$$g_{0C}(x) = \frac{f_C(x)}{1 + \exp(a + bx)} \times \frac{1}{0.73}.$$

For the intervention group, we may write analogously

$$g_{1I}(x) = \frac{f_I(x) \exp(a + bx)}{1 + \exp(a + bx)} \times \frac{1}{0.21}$$

and

$$g_{0I}(x) = \frac{f_I(x)}{1 + \exp(a + bx)} \times \frac{1}{0.79}.$$

Having obtained the probability distributions, we can now calculate the means and variances of the distributions using numerical methods. For example, for control patients with recurrence,

$$m_1 = \text{mean log labeling index} = \int_{-\infty}^{\infty} x \times g_{1C}(x)dx = 2.29$$

for the values of a and b given above;

$$v_1 = \text{variance log labeling index} = \int_{-\infty}^{\infty} (x - m_1)^2 g_{1C}(x)dx = 0.2373.$$

Similarly, for control patients without recurrence, m_2 = mean log labeling index = 2.07 and v_2 = var(log labeling index) = 0.2340. Hence, for control patients, $m_2 - m_1 = 0.22$ and v is approximated by $(v_1 + v_2)/2 = 0.2357$. Similar calculations for intervention patients give $m_2 - m_1 = 0.23$ and $v = 0.2358$.

Calculating the Correlation between Adjusted and Unadjusted Intervention or Prognostic Factor Effects

In a linear model relating independent variable y to two explanatory variables x and z, it can be shown that the correlation ρ between the estimate of the regression coefficient for x unadjusted for z, and the estimate of the coefficient for x adjusted for z, is $\sqrt{1 - r^2}$, where r is the correlation coefficient between x and z. We apply this result to our sample size calculations, recognizing that it is only an approximation for our case, where y is an indicator variable of polyp recurrence and is not continous. In our problem, x (0 or 1) is the group indicator (for intervention or prognostic factor groups) and z is log labeling index.

The correlation r is given by

$$r = \frac{\text{cov}(x, z)}{\sqrt{\text{var } x \times \text{var } z}}.$$

If p is the proportion in the group with $x = 1$ and $\bar{z}_0$, $\bar{z}_1$ are the mean values of the log labeling index in the two groups, then

$$\text{cov}(x, z) = p(1 - p)(\bar{z}_1 - \bar{z}_0)$$

$$\text{var } x = p(1 - p)$$

$$\text{var } z = v + p(1 - p)(\bar{z}_1 - \bar{z}_0)^2,$$

where v is the variance of log labeling index within each group.

For the question regarding mediation of the intervention (question 5), $p = 0.5$, $\bar{z}_1 - \bar{z}_0 = 0.36$ and $v = 0.2445$, leading to $r = 0.3421$ and $\rho = 0.9397$. For the question regarding mediation of the prognostic factor, number of polyps (question 4), $p = 0.4$, $\bar{z}_1 - \bar{z}_0 = 0.36$, and $v = 0.2445$, leading to $r = 0.3359$ and $\rho = 0.9419$.