Controls Who Experienced Hypothetical Causal Intermediates Should Not Be Excluded from Case-Control Studies

Charles Poole

It has been suggested that controls with adenomatous polyps of the colon and rectum should be excluded from case-control studies of cigarette smoking and colorectal cancer. A claim has been made that the presence of such controls creates a bias toward the null. The polyps are an intermediate step in a hypothetical causal pathway between the exposure and the disease. Thus, the recommendation logically extends to the exclusion of all controls who experienced hypothetical causal intermediates from all case-control studies. It is shown, in the simple case of an exposure that acts solely through the pathway involving the intermediate, that such exclusions create a bias away from the null. The rationale for recommending the detrimental exclusions appears to stem from a variant of the "trohoc fallacy": the mistaken view of case-control studies as comparisons between diseased and healthy groups and not as comparisons between groups that differ by exposure.


case-control studies; colorectal neoplasms; epidemiologic methods; research design; retrospective studies; smoking

"Ideally, of course, your controls would be the healthiest people in your study population." These words from a professor were chilling. I had learned in the classroom that the controls in a case-control study should represent the study’s population at risk (1). Had I confined my control group to the population’s healthiest members, I might have introduced control-selection bias into my study. Fortunately, we agreed it was too late: for the professor, too late to improve the study; for me, too late to damage it.

The professor was trapped inside the "trohoc fallacy" (2, p. 544): the mistaken view of the case-control study as a comparison between a group of persons who are sick and a group of persons who are healthy. When we conduct case-control studies, we do ascertain cases, select controls, and compute case-to-control ratios of exposure odds. Our “bottom line,” however, is to interpret those exposure odds ratios as estimates of relative risk. If they were not so interpretable, they would not be worth computing and the studies would not be worth doing.

A relative risk is not a comparison between diseased and non-diseased groups; it is a comparison between groups of persons who differ by exposure. It follows that the case-control study is most profitably viewed as an efficient way of conducting a cohort study (1-4). In principle, the only difference is that, instead of including all the persons or person-time composing the denominators of the risks or rates to be compared in the relative risk estimate, we take a sample. That sample is the control group.

The trohoc fallacy manifests itself in two main guises. One is a misplaced concern with comparability between the controls and the cases, when the comparability of real concern is between the controls and the population at risk. Another is evident in a line of reasoning originally advanced by Boutron (5) and elaborated more recently by Terry and Neugut (6) in an attempt to explain why cigarette smoking has shown weak and inconsistent associations with colorectal cancer, despite the presence of strong and consistent associations between smoking and adenomatous col-

Received for publication September 1, 1998, and accepted for publication March 4, 1999.

From the Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC.
orectal polyps and between the polyps and colorectal cancer:

The paradoxical results may stem from the inclusion of subjects with adenomas in the control groups of most colorectal cancer case-control studies... Given the likely high proportion of subjects with adenomas in an unscreened control group and the consistent association found between smoking and adenoma risk, it is possible that studies of smoking and colorectal cancer have been biased toward the null hypothesis of no association. (6, p. 904)

Like the professor from my past, Boutron, Terry, and Neugut are concerned that the control groups are not healthy enough or ""clean"" (6, p. 908) enough because they include persons who experienced a step (polyp formation) hypothesized to be a causal intermediate between the exposure (smoking) and the disease (colorectal cancer). Potter endorses this reasoning and characterizes the alleged problem of controls with polyps as one of "misclassification of outcome" (7, p. 911). This characterization fits with the "the cases are diseased so the controls should be healthy" outlook of the trohoc paradigm, but it does not conform to our ordinary concept of outcome misclassification. The outcome in these studies is colorectal cancer. If there were a problem with polyps in the control groups, that problem would exist even if the sensitivity and specificity of colorectal cancer classification were 100 percent.

Neither Boutron (5), Terry and Neugut (6), nor Potter (7) offer a verbal explanation, an algebraic argument, or a hypothetical example showing that controls with polyps create a bias toward the null. All that is offered is an analysis of results from an actual study, in which the estimated effect of cigarette smoking on colorectal cancer becomes stronger when the controls with polyps are removed. As shown below, this empirical "test" (6, p. 908) is misleading.

**HYPOTHETICAL EXAMPLE 1**

In the hypothetical study population described in figure 1, smoking trebles the risk of polyps and polyps treble the risk of colorectal cancer. For simplification, smoking causes colorectal cancer by no mechanism other than by causing polyps, and the polyps caused by smoking create the same increased risk of colorectal cancer as other polyps. The values in the example are fictional but are roughly realistic. That 30 percent of

---

**FIGURE 1.** A hypothetical population at risk in a study of cigarette smoking, adenomatous colorectal polyps, and colorectal cancer, in which colorectal cancer can occur in the presence or absence of polyps (or of detected polyps).
the population develops polyps compares well with the prevalence of 28 percent in Terry and Neugut’s (6) control group. Given that the lifetime risk of colorectal cancer is about 6 percent (8), the overall risk of 3 percent in the example may not be too inaccurate for less than lifetime follow-up.

In the example, 380 incident cases of colorectal cancer occur in the smoking group and 260 cases occur in the abstention group. Given the equal sizes of the groups, the unbiased relative risk is $380/260 = 1.5$. Whether expressed in relative or absolute terms, the effect of smoking on colorectal cancer is much weaker than either the effect of smoking on polyps or the effect of polyps on colorectal cancer. There is no “paradox” (6, p. 903; 7, p. 911), however. This dilution of effect across steps in a causal pathway was demonstrated by Morrison (9) two decades ago; it is not a bias, but a mathematical consequence of the cascade of effects across two or more independent steps in a multi-step process.

If a case-control study were to be done with a control group consisting of a perfectly representative sample of the entire population at risk, the control group would include a small number of incident colorectal cancer cases and would contain exactly half smokers. The exposure odds ratio would equal 1.5, the unbiased value. A control group perfectly representative of the 19,360 cohort members who did not develop colorectal cancer would contain 49.7 percent smokers. Because of the rarity of the disease, the upward bias in the estimated relative risk would be miniscule (10).

Now consider the recommendation (5, 6) to select a control group from the cohort members who remained free not only of colorectal cancer, but of polyps as well. The prevalence of smoking in this control group would be $5,390/(5,390 + 8,330) = 39$ percent. The exposure odds ratio would equal 2.3, substantially higher than the unbiased value of 1.5. In an actual study, in which the details of figure 1 would not be observable, this creation of a bias away from the null could be mistaken for the removal of a bias toward the null.

Deleting the cases with polyps as well, for the sake of case-control comparability, would have an even worse effect on validity, but in the opposite direction. The smoking prevalence among the cases would be $110/(110 + 170) = 39$ percent and the relative risk estimate would be 1.0. The argument here against removing the cases and the controls with polyps is equivalent to Robins and Greenland’s (11) argument against stratifying by causal intermediates, because a study restricted to polyp-free persons is a study confined to one stratum of a causal intermediate. What Boutron (5) and Terry and Neugut (6) propose, and Potter (7) endorses, is to confine the controls but not the cases to one stratum of a causal intermediate.

HYPOTHETICAL EXAMPLE 2

One objection to the first example might be that only $360/640 = 56$ percent of the colorectal cancer cases have a history of polyps. The example might be more realistic with reference only to detected polyps, especially in a study population that is only partially screened. In the second example (figure 2), colorectal cancer occurs exclusively among persons who develop polyps. The risk of colorectal cancer among persons with polyps has been raised to 0.11 to keep the total number of cases ($n = 660$) and the overall colorectal cancer risk (3 percent) approximately the same as in the first example. Here we see no dilution of effect across steps of the etiologic sequence: Smoking exerts the same threefold increase on the risk of colorectal cancer ($495/165 = 3.0$) that it exerts on the risk of colorectal polyps. A valid case-control study, with controls with polyps included, would produce an unbiased estimate of this relative risk. If controls with polyps were excluded, however, the expected estimate of relative risk from the case-control study would be biased upward to $(495/165)/(5,500/8,500) = 4.6$

DISCUSSION

In both hypothetical examples, the studies are assumed to be perfect in every way except for the ill-advised exclusion of persons with polyps from the control groups of the case-control studies. In reality, of course, epidemiologic studies are never perfect. It is quite possible that the effects of smoking on the risk of colorectal cancer have been underestimated in some studies. For example, it has been suggested that the induction period from smoking to colorectal cancer is very long, on the order of several decades (12, 13). If so, colorectal cancer studies in which smoking in the distant past is not ascertained carefully and analyzed separately from more recent smoking could produce attenuated estimates of effect.

The simplified hypothetical examples also fail to reflect other complexities, such as the possibility of a mixture of direct and indirect smoking effects, some mediated through polyp formation and others through other pathways. Under no circumstances that have been postulated thus far, however, is there a bias that would be lessened by removing controls with polyps. The most plausible result to expect from such deletions is the creation of an upward bias.

Sometimes the trohoc fallacy manifests itself as an obsession with comparability between controls and cases and, at other times, as discussed here, as an
FIGURE 2. A hypothetical population at risk in a study of cigarette smoking, adenomatous colorectal polyps, and colorectal cancer, in which colorectal cancer can occur only in the presence of polyps (or of detected polyps).

equally mistaken view that controls should be healthy because cases are ill. In both instances, the intuitive basis is an inapt analogy with cohort studies, where comparability between exposure groups is a valid concern and where it is reasonable to remove exposed persons from the unexposed group. In either guise, the trohoc fallacy is a deleterious distraction from the fundamental goal that case-control studies and cohort studies share: to produce an estimate of effect that is as valid and precise as possible.

The key to validity in selecting a control group is neither to seek comparability between the controls and the cases nor to seek a control group that is as healthy or as clean as possible. It is to seek comparability between the controls and the study’s population at risk. A control group is valid if and only if it adequately represents the exposure distribution in the persons or person-time that form the denominators of the rates or risks that will be compared when the case-to-control ratio of exposure odds is computed as an estimate of the rate ratio or risk ratio. How healthy the controls are and whether or not they have experienced hypothetical causal intermediates between the exposure and the disease are irrelevant to the validity of the control group.

If the exposure causes or prevents the disease by causing or preventing an intermediate condition or event, the population at risk includes persons who experience that intermediate step. Consequently, a valid control group will include them as well. The “plague” (7, p. 912) is not controls with polyps, but the trohoc fallacy, which makes controls with polyps seem problematic.

Most disappointing is that a bias toward the null seems to have been so intuitively obvious that the step of constructing an algebraic argument or a simple hypothetical example appears not to have been taken. Instead, investigators “tested the hypothesis” empirically by removing controls with polyps from an actual study and interpreted the resulting increase in the estimated effect of smoking on colorectal cancer as “empirical support” for the hypothesis because this result was “not inconsistent with such an explanation” (6, p. 908). Unfortunately, an empirical test of a hypothesis is of little help if it fails to pit that hypothesis against a competing alternative. We might for example drain a swamp next to a village with a high malaria rate and observe that the rate goes down. Of course, the problem with interpreting this result as empirical support for the hypothesis that malaria is caused by mias-
matic emanations from swamps is that the test does not pit the miasma hypothesis against an alternative involving a microbe and a swamp-loving insect. Similarly, the empirical test in the present case did not pit the hypothesis that controls with polyps create a bias toward the null against the alternative hypothesis that excluding those controls would create a bias away from the null. As it turns out, a dispositive test was not an empirical one, but a rather simple theoretical one (figure 1).

Intuition is a marvelous source of inspiration, but reliance on it for conclusions is one of the worst enemies of progress in our field. Intuition is unreliable in part because it is conditioned by paradigms. By the time some paradigms such as the trohoc fallacy are found to be mistaken, they have become so deeply entrenched that we may have great difficulty seeing how strongly and pervasively they have conditioned our intuitions. When this happens, we must struggle to recondition those intuitions and to excise the faulty paradigms from teaching and practice as quickly and as completely as possible, before they do any more harm.

ACKNOWLEDGMENTS

The author thanks Jay S. Kaufman for comments that led to the second example.

REFERENCES