Editorial

Are Clinical Trials the “Gold Standard” for Cancer Prevention Research?

Alan R. Kristal
Fred Hutchinson Cancer Research Center, Seattle, Washington

Most cancer epidemiologists would cite results from a randomized clinical trial as both the ultimate arbiter of controversial findings from observational studies and the basis for cancer control interventions. Indeed, we consider a trial definitive when its intervention demonstrably lowers cancer risk. But most cancer prevention trials have not shown that interventions appropriate for disease prevention in healthy individuals can lower cancer risk, and dishearteningly, several have found that these interventions can increase cancer risk. The interpretation of these “null” studies is far from straightforward. Does it mean that the behavior change or chemopreventive agent being tested does not prevent cancer? Or has some aspect of the design or execution of the trial led to a false null result? Consider some of the many ways in which a cancer prevention trial might fail: the intervention dose was too high or low, the planned trial duration was too short, an unexpected side effect ended the trial early, intervention adherence was poor, too many control participants “dropped in” to the intervention, only subgroups of participants were susceptible to the intervention, or the intervention itself affected end point detection. This leaves cancer prevention researchers with a conundrum: we (and the Food and Drug Administration) accept that prevention trials are necessary to show intervention efficacy, yet experience suggests that even the best designed and executed trials have a good chance of yielding false null results. In this editorial, I use examples from several large trials to illustrate ways in which modifications to our approaches to trial initiation, design, and analysis could make it more likely that cancer prevention trials will identify effective interventions to reduce cancer risk.

One of the most difficult decisions in the design of a clinical trial is to define the intervention and its dose. Chemoprevention trials, especially those using micronutrients or phytochemicals, are especially difficult in this regard. Do we choose a maximum tolerable dose; a dose comparable with what is generally obtained from food; a dose derived from results of in vitro, animal, or human observational studies; or a dose tested in a previous clinical trial? Do we use foods high in the micronutrient, a mix of naturally occurring forms of a micronutrient, or a single compound? These were challenging questions for the design of the Selenium and Vitamin E Cancer Prevention Trial (SELECT), and the decisions about dose and formulation of study agents were made only after careful, thoughtful, and at some times very cantankerous deliberations (1). For vitamin E, SELECT used a dose of 400 IU/d; this dose was believed to have no serious side effects, but it was far above both the 30 IU/d generally available from food and the 50 IU/d that was associated with reduced prostate cancer risk in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (2). For selenium, SELECT used selenomethionine; this was believed to be safer than inorganic selenium compounds but was only one of many selenoproteins found in the selenized yeast that had been associated with reduced prostate cancer risk in the Nutritional Prevention of Cancer Trial (3). The SELECT intervention ended ~5 years before its planned stopping date, after interim analyses found little chance that either compound could prevent prostate cancer. These disappointing SELECT results cannot be attributed to errors in statistical design or study execution; all assumptions for recruitment, adherence, drop-in, dropout, and cancer incidence were either overly conservative or correct. However, the possibility remains that the decisions of SELECT on supplement dose and formulation were wrong. Could this type of uncertainty be minimized in the design of future trials? I propose that two types of studies are needed. First, we must identify likely mechanisms of action and develop measures of these mechanisms that can be used as intermediate end points in human research. Second, we must complete small-scale clinical trials that are designed and powered to test how alternative doses and formulations affect these intermediate end points and this must include careful scrutiny for effects that are potentially harmful. It is helpful but not sufficient to show that an agent affects cells in vitro or, when ingested, can be detected in target tissues. This research is essentially the transition of mechanistic studies from in vitro and animal models to humans; it is admittedly difficult but could help clinical trials choose the interventions most likely to reduce cancer risk.

The most challenging aspect of executing cancer prevention trials is their long duration, which requires extraordinarily long-term commitments from investigators and participants. During the many years between trial design and end, much may happen to challenge the implementation of the study: well-publicized observational studies may report no effect or even harm from the intervention being tested, adherence to the intervention
could deteriorate, costs increase and funding may become inadequate, or secular trends may affect both disease incidence and control-arm lifestyle characteristics. These all pose difficult problems for investigators, who must assure that the trial design variables are met through effective retention and compliance programs and justify to participants, data safety monitoring committees, and institutional review boards that the intervention remains safe, likely to prevent disease and unproven in effectiveness. Many of these problems surfaced during the 25+ years that elapsed between the original proposal to test whether a low-fat diet could prevent breast cancer and the 2006 publication of results from the Women’s Health Initiative (WHI; ref. 4). Between approximately 1980 and 2000, dietary fat intake (as a percentage of total energy) among U.S. women decreased from 36.1% to 32.8% (5), making it difficult to recruit participants with high fat intakes. Few women randomized to the intervention arm reached the goal of 20% energy from fat, despite the use of a well-tested dietary intervention and the Herculean efforts of study staff and investigators. And in 1996, a much-publicized, pooled cohort analysis based on 4,980 cases and 337,819 women reported no association of dietary fat intake with breast cancer risk (6), which could have affected participant (and perhaps staff and investigator) commitment to the trial. Any or all of these could have contributed to the smaller than expected effects of the intervention on both fat intake and breast cancer risk: the difference in fat intake between women in the intervention and comparison arms was ~70% of the design goal, and instead of the expected 14% reduction in breast cancer risk in the low-fat arm, risk was reduced by only 9% (95% confidence interval, −1% to 17%). Notwithstanding both this nearly significant result and the widening separation of the cumulative incidence curves for the low-fat and control arms, the trial was terminated as planned. Some of the challenges in the dietary intervention arm of the WHI, as well as in other trials of testing effects of complex, long-term behavior change, may be insurmountable. Investigators can do nothing about secular trends or well-publicized scientific findings that do not support the trial hypothesis. But two issues related to long study duration can be addressed. First, long-term intervention adherence is a well-known problem that is often studied but not solved; trials should not begin until their intervention effectiveness can be assured. Second, the decision of when to end a trial is enormously complex, but we should be open to extending a trial beyond its planned end date, especially when original design assumptions are not met. Using the WHI as an example, whether an additional study year with continued dietary intervention would have resulted in a statistically significant outcome cannot be known.

The last issue I will raise is the use of secondary analyses, specifically subgroup analyses, analyses that do not conform to intent-to-treat criteria and analyses that combine information from observational and experimental findings. These analyses can be used to better understand trial outcomes (e.g., by correcting for errors in design or adjusting for problems in execution). Three examples illustrate the power of these secondary analyses to understand study outcomes. Based on very small changes in serum carotenoids in the intervention arm, it is clear that long-term adherence to at least the fruit and vegetable component of the low-fat, high-fruit, and high-vegetable diet in the Polyp Prevention Trial was modest (7); when analyses excluded noncompliant intervention-arm participants, the findings strongly confirmed study hypotheses (8). Long-term adherence to the low-fat diet never met design assumptions in the WHI; when analyses were restricted to women with high fat intake at baseline, the results confirmed study hypotheses (4), and among women in the comparison arm, dietary fat intake at baseline strongly predicted breast cancer risk (9). Finally, treatment with finasteride significantly reduced prostate size (10) and increased the sensitivity and sensitivity of prostate-specific antigen screening for detecting high-grade prostate cancer (11), thereby increasing the probability of detecting high-grade cancer in the treated arm of the Prostate Cancer Prevention Trial; when analyses were adjusted for the probability of cancer detection, there was no longer evidence that finasteride increased the risk of high-grade disease (12). Admittedly, these post hoc analyses are all subject to either chance or bias, but they may better represent the true test of the original trial hypothesis. To some extent, these analyses can be anticipated, especially those that examine highest-risk subgroups or adjust for intervention adherence. More complex problems, such as treatment effects on outcome detection, are very challenging to incorporate into an analysis; yet, when there is evidence that it may occur, an approach to evaluate its effect, based on specifically designed ancillary studies or other sources of data, could be developed and specified a priori.

Perhaps clinical trials cannot be the only “gold standard” for cancer prevention research. Their size and duration, along with their inherent problems in long-term adherence, make them unfeasible for addressing many important questions, especially those related to behavior change. Admittedly, the effect of some cancer prevention trials has been profound (e.g., the rapid drop in breast cancer incidence in 2003 that can be attributed to the WHI trial of hormone replacement therapy; refs. 13, 14), but in general, we have learned more from trials about cancer biology and epidemiology than about effective policies for cancer control. The most vexing problem is a trial with null results because we are left wondering whether the underlying hypothesis of the trial was adequately tested. I believe we can do better. We should raise the bar on the amount of information needed before initiating a trial, and in particular, we should have a better understanding of the biology of the proposed intervention in humans. This would allow us to more rationally choose agents for new trials from among the many compounds with beneficial effects in in vivo, animal, and human observational studies (e.g., sulforaphane, green tea catechines, and vitamin D). We should be realistic in our expectations for long-term adherence to behavioral interventions. There is substantial motivation for a trial of weight loss to prevent prostate cancer recurrence after initial treatment, but it would be unwise to mount such a trial without an intervention that could maintain weight loss for at least 10 years. Finally, we should more seriously consider findings from secondary, a priori, and non–intent-to-treat analyses. Although these analyses do not meet the
Food and Drug Administration standards for proven efficacy, they should be communicated to the public and incorporated into evidence used to make public health recommendations to reduce cancer risk.

References

Are Clinical Trials the "Gold Standard" for Cancer Prevention Research?

Alan R. Kristal

Cancer Epidemiol Biomarkers Prev 2008;17:3289-3291.

Updated version Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/17/12/3289

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.