Cancer Epidemiology, Biomarkers & Prevention

Hypothesis/Commentary

Cancer Epidemiology in the 21st Century

Cultivating Cohort Studies for Observational Translational Research

David F. Ransohoff

Abstract

Background: "Discovery" research about molecular markers for diagnosis, prognosis, or prediction of response to therapy has frequently produced results that were not reproducible in subsequent studies. What are the reasons, and can observational cohorts be cultivated to provide strong and reliable answers to those questions?

Experimental Methods: Selected examples are used to illustrate: (i) what features of research design provide strength and reliability in observational studies about markers of diagnosis, prognosis, and response to therapy? (ii) How can those design features be cultivated in existing observational cohorts, for example, within randomized controlled clinical trial (RCT), other existing observational research studies, or practice settings like health maintenance organization (HMOs)?

Results: Examples include a study of RNA expression profiles of tumor tissue to predict prognosis of breast cancer, a study of serum proteomics profiles to diagnose ovarian cancer, and a study of stool-based DNA assays to screen for colon cancer. Strengths and weaknesses of observational study design features are discussed, along with lessons about how features that help assure strength might be "cultivated" in the future.

Conclusions and Impact: By considering these examples and others, it may be possible to develop a process of "cultivating cohorts" in ongoing RCTs, observational cohort studies, and practice settings like HMOs that have strong features of study design. Such an effort could produce sources of data and specimens to reliably answer questions about the use of molecular markers in diagnosis, prognosis, and response to therapy. *Cancer Epidemiol Biomarkers Prev; 22(4); 481–4. ©2013 AACR.*

The Challenge of Observational Research

A challenge for observational epidemiology in the 21st century is to advance clinical and public health practice by "bridging an evidence gap" in addressing questions along the translational continuum of T0–T4 phases (1). In the last 10 to 20 years, the focus of much translational research has been on basic scientific discoveries (T0) and early descriptive noninterventional studies (T1), about markers of sensitivity and specificity (to assess molecular markers for diagnosis), and about predictive value (to assess molecular markers for prognosis or prediction). A major problem is that many "discoveries" have turned out to be not reproducible in subsequent studies (2–5). One reason for this lack of reproducibility is inattentiveness to epidemiologically sound study design and conduct, particularly when a study is "observational" and may not readily have the safeguards against bias that a randomized controlled clinical trial (RCT) can have. At one extreme, "convenience samples" might be used for discovery or for validation, where little or no attention has been given to the "design of the study" that collected the specimens that receive extensive biochemical, molecular, or mathematical analysis. At another extreme are prospective carefully designed and conducted observational studies that produce results that are reproducible and change practice or provide a reliable foundation for future work (6). From these experiences, it has become clear that investigators in this translational domain need to understand principles of observational epidemiology study design and conduct and details of what makes some observational studies strong and others weak. Said another way, investigators need to understand that all the activities that occur before blood or tissue collection must be considered to be part of a "research study" whose methods will be described in detail in a research report, so that readers can judge the strength and potential reproducibility of a study result. Those activities include

Author's Affiliation: University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Note: The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Note from the Editor-in-Chief: This is one in a series of commentaries that have arisen from an initiative of the National Cancer Institute to advance epidemiological science in the 21st century.

Corresponding Author: David F. Ransohoff, CB 7080, UNC-CH, Chapel Hill, NC 27599. Phone: 919-966-1256; E-mail: ransohof@med.unc.edu

doi: 10.1158/1055-9965.EPI-13-0140

selecting subjects for study, arranging methods to collect and store specimens, and arranging comparisons (e.g., between people with and without cancer), so that they avoid systematic differences or bias between groups that cause incorrect and misleading results. Serious problems can occur when these details of design and conduct are thought of only after the biochemical, molecular, and mathematical analyses are done. Rather, they need to be considered at the onset of any study before any analysis (6).

With this background, it is useful for investigators working in translational research, including researchers in basic science and technology as well as in clinical research, to consider lessons from recent experiences:

1. What makes observational study design strong or weak?
2. Can we cultivate and leverage existing cohorts and, if necessary, create new ones?

The concept of leverage is appealing because it may be economical to use already existing infrastructure (see recommendations in this issue; ref. 7). Examples of leveraging provide lessons about how to, in the future, further "cultivate" this kind of effort. The goal of this commentary is to begin a larger discussion about how to cultivate strong observational research in a deliberate, economical, and productive manner.

Examples of Strong Observational Research About Molecular Markers of Diagnosis and Prognosis

One example of successful leveraging of an observational study for prognosis assessed whether an RNA expression profile could predict prognosis of women who had had surgical treatment of breast cancer and were considering adjuvant therapy. The question was whether risk of cancer recurrence could be predicted; if the rate of recurrence were low enough, a woman might decide to forego the expense and side effects of adjuvant therapy. The study of the OncoTypeDx test, that assessed RNA expression profiles of tumor tissue taken at the time of surgery, was conducted "retrospectively" within an already existing cohort (8). The study used banked formalin-fixed paraffin-embedded (FFPE) tissue specimens, along with data from the already completed National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 RCT. In that study, the placebo group, having received no adjuvant therapy, could be assessed to see in whom breast cancer recurred and whether RNA expression of multiple genes in tumor tissue at initial surgery could predict risk of recurrence. To measure RNA levels in FFPE specimens, study investigators had to develop new technical methods to measure RNA in tissue that had not been quick-frozen but rather had been routinely collected as FFPE specimens. Development of that translational technology in effect unlocked a new universe of "already completed" studies, like RCTs, as a source of data and specimens for molecular and mathematical analyses that could be used to study markers of prognosis and prediction.

An example of a strong observational study that was leveraged to answer a question about diagnosis addressed whether a blood-based proteomics test could be used for screening to find asymptomatic early ovarian cancer. Before this study, several strong claims had been made about new blood-based proteomics markers for ovarian cancer screening that were much better and nearly 100% sensitive and specific than the CA125 serum marker (9, 10). Those claims, however, had been strongly disputed on the basis of methodologic concerns (4, 11, 12), but the question of "how sensitive and how specific" remained unanswered. To test the initial assays as well as others, a high-quality already completed RCT was leveraged to provide blinded assessment of 5 independent panels of markers. Blood specimens were selected from the biospecimen repository that had been already been created as a part of the National Cancer Institute's (NCI) Prostate, Lung, Colon, Ovary (PLCO) RCT (13). In that RCT, serial blood samples had been drawn among people who were being followed for the development of cancer, including ovarian cancer, so that a blood specimen obtained shortly before ovarian cancer diagnosis could be identified and used to assess whether a blood test could detect asymptomatic cancer. Each proteomic panel was "tested" on exactly the same serum specimens. This study found that none of the new assays did better than CA125. Although the PLCO study was a RCT, only its "observational" features were used in this assessment; randomization to one or the other intervention group and assessment of ovarian cancer mortality were not relevant in this added-on study about whether markers could detect cancer.

A lesson from this study is how a new research question and study can be piggybacked on to an RCT or well-designed cohort study, in which repeated collection of blood or other specimens is done, to assess a diagnostic test. Such cultivation in the future might involve adding collection of serial blood samples (using appropriate volumes and collection and storage practices) to ongoing research cohorts like the Framingham Study, the Nurses' Health Study, Women's Health Initiative, or within practice cohorts like in an HMO like Kaiser-Permanente.

For both examples above, major challenges in "cultivation" involved arranging access to data and specimens, assuring that "enough" specimen was available, and using, and sometimes developing, novel methodologies to assess available material (i.e., measuring RNA in FFPE tissue or measuring protein in small quantities of serum or plasma) to investigate questions that had not been asked or planned for in the original study.

A third example is a study not "piggybacked" onto a RCT or other existing infrastructure but that was prospectively planned for the specific purpose of assessing a stool DNA test to detect colon cancer (14). The study involved conducting screening colonoscopy (to determine the presence or absence of colon cancer) preceded by the collection of stool samples that would then be examined for fecal DNA–containing cancer cell DNA. The true state remained blinded to study investigators. Although this study had a single dedicated purpose with focus on stool as the source of diagnostic information and on one panel of markers, it is easy to imagine cultivation of this type of resource for larger uses. For example, more stool specimen might be collected for other investigators to use in later discovery, or validation of other stool-based assays or adding blood samples collected before the true state examination might be done for other investigators to look at blood-based CRC markers. In a parallel manner, entirely different sources of specimens might be cultivated for studying a blood-based test for CRC by piggybacking specimen collection onto an ongoing RCT in which colonoscopy screening is already being done (15). Such kinds of efforts might leverage, through a small amount of additional funds or infrastructure, studies whose main expense (learning the true state by colonoscopy) was already in place or such an infrastructure might be appropriately added onto a large HMO or other medical practice.

These examples illustrate how one study may be leveraged to evaluate multiple markers and multiple different questions (5). The examples illustrate the kinds of challenges that must be addressed in an effort to "cultivate": Is "study" design and conduct (including the activities that happen before specimens are collected) strong enough, overall, to produce reliable unbiased comparisons and results? Is "enough" specimen obtained, using the right collection and storage methods, for multiple users? These examples begin to provide lessons about how cohort studies, done for some other purpose, might be "added to" or cultivated in ways that let them be sources of specimens and data for "strongly designed and conducted" studies of diagnosis, prognosis, and prediction in observational translational research about T0 and T1 questions.

Next Steps

It will be useful to consider additional examples and lessons about how cohorts may be cultivated. What makes a cohort strong or weak about challenges of bias and generalizability? What are potential sources of data and specimens and their strengths and weaknesses, including RCTs, observational cohort research studies, and practice cohorts like in HMOs? (5). Will these resources be willing to openly share data and specimens for research questions that the original study did not set out to address? What features might need to be added to "cultivate" existing infrastructure and to have the strong design and conduct that ultimately will be reported in a Methods section? Can such approaches lead to strong and reliable comparisons as would be arranged in a dedicated, prospectively designed, and conducted formal study? The challenge is not just to ensure sharing of data or specimens that have already been collected. Indeed, such sharing is already being done or planned among major cohort resources like NCI's Cohort Consortium or NCI's Cancer Research Network. A more fundamental challenge is to make sure that the data and specimens collected in a cohort can be part of a "study" with features of "study design" that provide unbiased comparisons and generalizable results (5, 6). In cultivating those features, for example, do new types of outcome data or exposure data (or specimens) need to be collected? If so, when and how? Does blinding need to be arranged to avoid serious biases, if so, then in what places in "design"? Identifying and addressing these kinds of challenges may help to grow new sources of material that can be used to strengthen T0 and T1 translational research to advance clinical and public health practice.

Disclosure of Potential Conflicts of Interest

D.F. Ransohoff is an unpaid consultant/advisory board member of Exact Sciences and Epigenomics.

Grant Support

This work was supported by NCI. The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH, through grant Award Number UL1TR000083.

Received February 4, 2013; revised February 15, 2013; accepted February 17, 2013; published OnlineFirst March 5, 2013.

References

1. Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. Am J Epidemiol 2010;172:517–24.
2. Micheel C, Nass SJ, Omenn GS, editors. Evolution of translational omics: lessons learned and the path forward. Washington, DC: Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine; 2012.
3. Ioannidis JP, Khoury MJ. Improving validation practices in "omics" research. Science 2011;334:1230–2.
4. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst 2005;97:307–9.
5. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. J Clin Epidemiol 2007;60:1205–19.
6. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. J Clin Oncol 2010;28:698–704.
7. Khoury MJ, et al. TBA. Transforming Epidemiology for 21st Century Medicine and Public Health. Cancer Epidemiol Biomarkers Prev 2013;22:508–17.
8. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.

9. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.
10. Visintin I, Feng Z, Longton G, Ward DC, Alvero AB, Lai Y, et al. Diagnostic markers for early detection of ovarian cancer. Clin Cancer Res 2008;14:1065–72.
11. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst 2005;97:315–9.
12. McIntosh M, Anderson G, Drescher C, Hanash S, Urban N, Brown P, et al. Ovarian cancer early detection claims are biased. Clin Cancer Res 2008;14:7574.
13. Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RM, et al. A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res 2011;4:375–83.
14. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. N Engl J Med 2004; 351:2704–14.
15. Bretthauer M. The NordICC clinical trial; 2010 [Accessed 2010April 12]. Available from: "http://www.clinicaltrials.gov/ct2/show/NCT00883792?term=bretthauer&rank=2.