Surrogate Markers: Lessons from the Next Gen?

Brian J. Reid1,2

Abstract

The article by Banerjee and colleagues published in this issue of the journal involving a randomized control prevention trial of ursodeoxycholic acid (UDCA) in Barrett esophagus reported a null outcome despite being well designed and executed. Possible reasons for this null outcome are discussed focusing on use of surrogate endpoints in the trial. The trial is especially topical because it comes at a time when there are calls for a Pre-Cancer Genome Atlas (PCGA) for "understanding the earliest molecular and cellular events associated with cancer initiation..." This commentary discusses current concepts in prevention research including branched evolution that leads to therapeutic resistance. Length bias sampling postulates underdiagnosis is due to rapidly progressing disease that is difficult to detect by screening because it progresses to cancer too rapidly and that overdiagnosis is the result of very slowly or nonprogressing disease that is easy to detect by screening because it persists for a lifetime and the patient dies of unrelated causes. Finally, it also explores study designs, including surrogate endpoints in Barrett esophagus trials, and opportunities and pitfalls for a PCGA in the context of high levels of over and underdiagnosis of Barrett esophagus as well as many other cancers and their precursors. Cancer Prev Res; 9(7): 512–7.

©2016 AACR.

See related article by Banerjee, et al., p. 528

The article "Clinical study of ursodeoxycholic acid (UDCA) in Barrett’s esophagus (BE) patients" by Banerjee and colleagues in this issue of Cancer Prevention Research poses troubling questions for cancer prevention strategies in Barrett esophagus and perhaps more broadly for cancer prevention in other conditions as well. The study is well designed with a compelling biologic hypothesis, an intervention with a biologic rationale that is supported by other studies, and relatively objective surrogate immunohistochemical endpoints. These endpoints are in contrast to the usual dysplasia endpoints, which have been shown repeatedly to not be reproducible in observer variation studies yet are continuing to be used in many clinical Barrett esophagus trials. The UDCA study used a systematic biopsy protocol with unbiased tissue sampling including one biopsy in each of four quadrants every two centimeters in the Barrett esophagus segment. Early-phase randomized prevention trials such as this one are difficult to perform in Barrett esophagus and in many other conditions, and the authors are to be congratulated on successful completion of the study.

Meeting enrollment goals for these prevention trials can be a challenge. It took 3 1/2 years to identify and consent 80 potentially eligible participants, 36 of whom met all eligibility criteria and 29 completed the study. This is a generalized experience in Barrett esophagus trials. In a dietary intervention study performed in Seattle in the 1990s, a total of 1,151 patients were referred to the study and screened for eligibility and 93 randomized patients (1).

Meeting enrollment goals for these prevention trials can be a challenge. It took 3 1/2 years to identify and consent 80 potentially eligible participants, 36 of whom met all eligibility criteria and 29 completed the study. This is a generalized experience in Barrett esophagus trials. In a dietary intervention study performed in Seattle in the 1990s, a total of 1,151 patients were referred to the study and screened for eligibility and 93 randomized patients (1).
The immunohistochemical results with SOHdG in the UDCA study raised concern about nonspecific staining so that only moderately or strongly stained nuclei were used in the final analyses. Prevention trials are very challenging, and ideally we would like to learn more from each trial whether or not it is successful so that future trials can be designed to be more successful. As the authors note, few surrogate markers, including those used in this study and our earlier dietary intervention study, have been validated as predictors of progression to esophageal adenocarcinoma in clinical studies.

The question of the validity of surrogate measures is becoming fundamental to the design of early-stage prevention trials. Rigorous standards for using surrogate biomarker endpoints have been well-established for more than 2 decades (6, 8, 9). Valid surrogate markers should be in key causal pathways to esophageal adenocarcinoma, have substantial ability to distinguish patients who progress to cancer from those who do not, and be objectively measured (7). Multiple studies in Barrett esophagus meet all or most of these requirements, including a cancer incidence endpoint although measures of dysplasia as predictors continue to not be “objectively measured” (10–18). Although it has become increasingly difficult to meet all of these standards, alternative study designs have emerged that preserve much of the original intent of the guidelines and are consistent with the results of other rigorous studies (19). Study designs that “map” the spatial distribution of genomic abnormalities in Barrett esophagus surrounding esophageal adenocarcinomas can provide highly useful information. However, one of the processes by which Barrett esophagus evolves to esophageal adenocarcinoma (20–22). However, these studies must be interpreted cautiously because they lack nonprogressing control populations and can easily lead to overdiagnosis and overtreatment. A further limitation of spatial study designs is that none of them address the critical element of time: how fast do the changes occur that lead to esophageal adenocarcinoma in general and underdiagnosis of esophageal adenocarcinoma specifically?

Ki67, which detects actively proliferating cells in G1, S, and G2–M, is one of the oldest measures of proliferation in Barrett esophagus and many other conditions, yet it has rarely been tested in prospective studies to determine the extent to which it actually discriminates between individuals who will and will not progress to cancer. One study used DNA content/Ki67 multiparameter flow cytometry in 853 diploid biopsies from 362 patients with Barrett esophagus (23). Of these patients, 276 were followed prospectively for an average of 6.3 years with 29 developing Barrett adenocarcinoma (23). Of these patients, 276 were followed prospectively for an average of 6.3 years with 29 developing Barrett esophagus (23). These increased 4N fractions and increased G2/tetraploid in the DNA content literature, and DNA content aneuploidy have consistently been reported in esophageal adenocarcinoma and in Barrett esophagus that progresses to esophageal adenocarcinoma using a number of different platforms, including DNA content flow cytometry (10, 11, 26, 27). The historical conflict between flow and image cytometry seems to have abated with recent results showing that both methodologies provide comparable information when the proper prep is used and the analyses are adjusted for clumping of nuclei that could lead to false-positive 4N cells (28). At the present time, DNA content cytometry is the only method that has been shown to reliably detect genome doublings at early stages when they constitute only a small proportion of the cells in the biopsy (~6%–15%) especially in the presence of one or more aneuploid cell populations in the same biopsy (11). Advances in bioinformatics employing collaborations between scientists with genome and cytometric expertise are likely to improve detection of whole-genome doublings as important predictors of progression in Barrett esophagus using DNA sequencing and SNP arrays just as advances in DNA content cytometric sample preparation were made by collaborations between investigators using image and flow cytometric methods (28).

Recently, there has been a call for a “Pre-Cancer Genome Atlas” (PCGA; ref. 29). We have learned much from The Cancer Genome Atlas (TCGA), but it was a case-only study design. This case-only design was appropriate to determine the atlas of genomic alterations in advanced cancers to guide targeted therapy. However, such a case-only study design will not be appropriate for precursors of cancer. Any PCGA will need well-designed studies that include nonprogressing controls to avoid exacerbating overdiagnosis and overtreatment (30, 31). For example, a recent study of healthy, sun-exposed skin reported an unanticipated abundance of somatic mutations, including purported drivers of skin carcinogenesis such as NOTCH1 (32).

Properly designed intervention studies will have to include many of the elements of the Banerjee UDCA study including unbiased protocol biopsy sampling in space and time to determine the evolutionary trajectory of different genomic alterations and robust, reproducible measures of genomic alterations that determine whether the patient will progress to cancer or not. The authors of this UDCA study have wisely avoided dysplasia grade, which meets none of the standards for surrogate markers (7).

A recent meta-analysis reported that 25% of all esophageal adenocarcinomas are missed after initial screen detection of Barrett esophagus, including esophageal adenocarcinomas detected within a year (33). Thus, both clinical and prevention research communities share a need for more robust measures of efficacy for the dietary intervention, and the results were also null (1).
risk for progression to esophageal adenocarcinoma than are currently being used. A combination of recent research advances and rediscovery of an ancient pathology literature is shedding light on why most Barrett esophagus remain stable for a lifetime ("overdiagnosis") while a minority of Barrett esophagus rapidly progress to esophageal adenocarcinoma ("underdiagnosis"). First, several lines of evidence indicate that Barrett esophagus has a large number of physiologic, expression, and proteomic functions that provide epithelial cell protection in the harsh environment of gastroesophageal reflux disease (7, 34–36). Second, recent studies in model organisms and human autopsy studies from the 1950s and earlier indicate that both the embryonic human and mouse esophagus have a columnar lining that is replaced by a squamous lining during fetal development and that these cells can be identified in the adult (37–39). Third, the expression profile of the columnar lining in the embryonic mouse is very similar to that of adult human Barrett esophagus (39). Reappearance of a latent embryonic program that arises as a protective adaptation in the presence of severe reflux disease may help explain why so few cases of Barrett esophagus progress to esophageal adenocarcinoma (7).

In contrast, esophageal adenocarcinoma have evidence of massive genomic alterations. For example, the transition from Barrett esophagus to esophageal adenocarcinoma in many patients occurs rapidly through chromosome instability (gain or loss of whole chromosomes or large regions of chromosomes) and whole-genome doubling with a 2- to 4-year window of opportunity for early detection of esophageal adenocarcinoma using a single endoscopic biopsy every 2 cm in the Barrett esophagus segment (17). Whole-genome sequencing studies have reported a median of 26,161 mutations (range, 18,881–66,225) per esophageal adenocarcinoma (40). It may be that the light microscope is not the optimal method for detecting the complex processes that lead to evolution of Barrett esophagus to esophageal adenocarcinoma.

Reliance on selective sampling based on nonreproducible, nonrobust clinical assessments such as histopathologic interpretation of dysplasia in Barrett esophagus will likely undermine the goals of any PCGA. Temporal assessment of the trajectory will be essential because length bias sampling is believed to underlie observer variation (45-47). Lack of reproducibility of dysplasia interpretation leads to the perception that the genomic/genetic results are not reproducible because two or more pathologists interpret the dysplasia status of biopsies differently.

These comments should not be considered a criticism of the pathology community that has worked very hard to improve the current dysplasia classification system. In many centers, Barrett esophagus cases may receive input from two or more specialty pathologists and difficult cases are frequently referred to centers with special expertise in Barrett esophagus, where multiple gastrointestinal pathologists participate in the final review. Yet, the problems persist despite the best efforts of both gastrointestinal and pathology communities, who in general establish close working relationships to try to provide the best possible care. This seems to be the best we can do with the current tools, and these tools and the second opinion options are far better than they were 30 years ago. However, there is still room for improvement as witnessed by the recent report that 25% of esophageal...

TP53 appears to be the only gene mutated at sufficiently high frequency (~70%) in esophageal adenocarcinoma and Barrett esophagus that progresses to esophageal adenocarcinoma to be highly informative for targeted early detection and prevention, although assessment of chromosome instability and genome doublings are likely to be useful in early detection and prevention as well (17–19, 40). These results raise the question of whether we should consider genomic alterations such as TP53 mutations, chromosome instability, and whole-genome doublings as markers of risk in prevention and early detection studies. A risk prediction model based on SNP array results in a single biopsy every 2 cm in the Barrett esophagus segment was highly predictive of progression to esophageal adenocarcinoma (area under the curve or AUC = 0.94) and outperformed histopathology on the basis of four biopsies every 2 cm (AUC = 0.81 for high-grade dysplasia and 0.78 for high- and low-grade dysplasia combined; ref. 18). Interestingly, the AUC for DNA content flow cytometry (0.79) was essentially equivalent to dysplasia classification even though only one biopsy was evaluated at each 2 cm interval compared to 4 for histopathology. The window of opportunity afforded by measuring these genomic alterations may allow more effective design of prevention studies in Barrett esophagus, potentially using interventions such as resveratrol and aspirin or other NSAIDs that have already shown promise in controlling some of these processes including whole-genome doublings and aneuploidy as endpoints (16, 24, 25).

Biopsies should not be selected for analyses in genomic studies on the basis of dysplasia classification, which is subject to observer variation (45–47). Lack of reproducibility of dysplasia interpretations can undermine progress to developing more robust indices of risk. This can become especially problematic if the nonreproducibility of dysplasia interpretation leads to the perception that the genomic/genetic results are not reproducible because two or more pathologists interpret the dysplasia status of biopsies differently.

These comments should not be considered a criticism of the pathology community that has worked very hard to improve the current dysplasia classification system. In many centers, Barrett esophagus cases may receive input from two or more specialty pathologists and difficult cases are frequently referred to centers with special expertise in Barrett esophagus, where multiple gastrointestinal pathologists participate in the final review. Yet, the problems persist despite the best efforts of both gastrointestinal and pathology communities, who in general establish close working relationships to try to provide the best possible care. This seems to be the best we can do with the current tools, and these tools and the second opinion options are far better than they were 30 years ago. However, there is still room for improvement as witnessed by the recent report that 25% of esophageal...
Adenocarcinomas are missed by current screening and surveillance practices (33). Two recent genomic studies have provided data indicating that whole-genome doublings may evolve so rapidly that they are missed by current screening and surveillance strategies (17, 22).

To some extent, the challenges the field faces with overdiagnosis of benign Barrett esophagus and underdiagnosis of early, curable esophageal adenocarcinoma is shared by many other cancers including those of the breast, prostate, lung, and many other organs (30). It is of interest that the sequence of chromosome instability followed by whole-genome doubling observed in Barrett esophagus has been observed in many cancer types in addition to esophageal adenocarcinoma, including breast, lung, colon, and ovarian (26). Extension of this type of research will be best accomplished using biorepositories that have established standardized biopsy protocols that obtain biospecimens at defined points in space and time such as described in Banerjee and colleagues. Thus, the advances of TCGA have provided potential clues for early diagnosis and prevention of multiple different cancer types in different organ systems that might be pursued in well-designed PCGA studies.

Recent DNA sequencing studies of esophageal adenocarcinoma have also reported breakage-fusion-bridge cycles in advanced esophageal adenocarcinoma that result from telomere attrition (44). Short telomeres have been reported in the blood of patients who progress from Barrett esophagus to esophageal adenocarcinoma (48), in Barrett esophagus tissue with evidence of chromosome instability (49) and in esophageal adenocarcinoma (50). What would be the implications of developing sequence-based biomarkers for telomere length for novel surrogates in cancer prevention and early detection research?

Other recent studies have reported results that may open new avenues for future studies, or potentially even for studies already completed provided sufficient biospecimens have been preserved for statistical power. For example, TCGA and other sequencing studies have reported a novel mutation signature that is found in gastric and esophageal adenocarcinoma (40, 44). The mutagen that causes this mutation signature is currently unknown, but its identification could lead to prevention strategies targeting the mutagen or reanalysis of biospecimens from previous trials. Recent studies, including esophageal adenocarcinoma sequencing studies, have also opened the door to reconsidering the etiology of Barrett esophagus and esophageal adenocarcinoma. Genome sequencing studies of esophageal adenocarcinoma are notable for lack of the tobacco signature, which has not been reported in esophageal adenocarcinoma even though tobacco smoking has been reported to be a risk factor in esophageal adenocarcinoma in association studies (40, 44). This finding in combination with population trends showing that lung cancer mortality declined following the decrease in tobacco smoking in the United States, whereas the rates and mortality of esophageal adenocarcinoma increased rapidly suggest that our understanding of the environmental insults that are etiologic for esophageal adenocarcinoma are incomplete. What role does tobacco play in the increased incidence of esophageal adenocarcinoma?
other as yet unknown mutagens slipped into the environment? Can we use DNA mutation signatures to track these carcinogens? Once they are identified, can we target the carcinogens for prevention of esophageal adenocarcinoma?

The vast majority of the above genomic results were reported after the Banerjee study was designed, and these possibilities were largely unknown four or five years ago. The Banerjee trial was well-designed and conducted yet the difficulty in interpreting the surrogate immunohistochemical biomarkers limits our understanding of the null results. Are we using measures that make biological sense (increased proliferation, decreased apoptosis, DNA damage) but haven't yet been validated as surrogate markers? Are current assays not sufficiently sensitive and specific to detect differences associated with progression? This raises questions for the field to answer, not a single individual. However, these and other questions should be discussed before undertaking a massive PCGA study. Now may be the time to take stock of surrogate markers being used in early-phase cancer prevention studies and to address the implications of such surrogate markers on the interpretation of the results, possibly through a conference or multidisciplinary think tank.

**Grant Support**

The study was supported by grants from National Cancer Institute (NCI) P01CA019555 and NCI RCI CA 146973.

Received February 25, 2016; revised April 15, 2016; accepted April 22, 2016; published OnlineFirst May 2, 2016.

**References**


Surrogate Markers: Lessons from the Next Gen?

Brian J. Reid


Updated version: Access the most recent version of this article at: doi:10.1158/1940-6207.CAPR-16-0059

Cited articles: This article cites 50 articles, 12 of which you can access for free at: http://cancerpreventionresearch.aacrjournals.org/content/9/7/512.full#ref-list-1

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.