



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2013 April ; 22(4): 521–527. doi:10.1158/1055-9965.EPI-13-0084.

Precursors in cancer epidemiology: aligning definition and function

Sholom Wacholder, PhD

Abstract

A precursor of a disease is a definable pathologic state that progresses directly to disease without a known intermediate step, and whose presence substantially increases the likelihood of disease. Precancers, or precursors of cancer, can help provide detail about the dynamic pathogenesis process before clinical disease. Thereby, ascertainment of properly defined precancers can increase precision of estimates and power in epidemiologic and clinical studies. Besides providing targets for direct treatment and improving tools for risk assessment in screening programs, precancers can help establish temporal ordering of cause and effect; can identify relatively homogeneous subsets of cancer that have passed through a given precancer state; and provide a basis for choosing high-risk individuals for detailed longitudinal study. Although the most appropriate definition of the precancer will vary with its function in particular research or clinical applications, the proportion of cancers that progress from the precancer and risk of cancer progressing from the precancer can be important measures of the value of a precancer in translational efforts.

A precursor of a disease is a definable pathologic state that frequently progresses directly to disease. Well-studied precancers, or precursors of cancer, have illuminated several complex aspects of natural history for cervical and colorectal cancer, and have helped to suggest and to evaluate successful intervention programs in prevention and patient management. Figure 1 displays two distinct pathways between an exposure and cancer. In the *indirect* causal pathway, the exposure causes the precancer which may then progress to cancer; the precancer is a causal mediator between exposure and disease, but also a pathogenic state. In the *direct* pathway, no precancer is known. Detection of the precancer between time of exposure and cancer incidence does not, of course, imply that the indirect pathway, rather than another mechanism entirely, led to the cancer.

The requirement that cancer progress from the precancer implies that complete removal of the precancer eliminates risk of malignancy arising from the precancer. Precancers at two sites, cervix and colon, meet this requirement, and, not coincidentally, have had great translational impact. The causal requirement is far stricter than association with cancer; the frequency requirement makes sure that the precursor has a strong effect on subsequent risk of cancer. The definition does not include biomarkers like prostate specific antigen (PSA) which do not cause cancer: an intervention that biochemically eliminated PSA from blood would not reduce prostate cancer risk. Nor does it include all ancestor cells with early somatic changes, like stem cells in Barrett's esophagus or neural crest in melanoma, which rarely lead to clinical disease. The definition includes rare conditions that nearly always progress, but not common conditions that progress to cancer only rarely, such as minor

Corresponding author: Sholom Wacholder, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, 6120 Executive Blvd., Room 5050, National Cancer Institute, Bethesda, MD 20892, WacholdS@MAIL.NIH.GOV, Voice: 301 496 3358, Fax: 301 480 3915.

The author has no conflicts of interest to report.

cervical cytologic or histologic abnormalities. Unlike the definition of Franco and Rohan,^{1, p. 7} this definition does not restrict cancer to be a solid tumor. Although this strict definition does not allow measures of either human papillomavirus (HPV) or *Helicobacter pylori* infection to be precancers, even when the infections may confer high subsequent risk of cancer, some properties of precancers generalize to other states.

Previous work has not emphasized the distinct value of using precursors in etiologic and translational research, nor how to choose the best definition of a precancer for a particular application. For example, the definition of precancer from a 2004 conference focused on 5 criteria to highlight pathologic distinctions;² the more general definition here, aimed at epidemiologic researchers, shares the first two criteria of increased risk of cancer and that the cancer arise from the cells of the precancer. This paper provides a comprehensive catalog of uses of precancers with examples and some guidance on how to define precancer in a given application.

Defining the precancer

Dependence on function of precancer

The precise precursor definition chosen for a particular research application should vary with the question being addressed.. For example, diagnosis of cervical intraepithelial neoplasia grade 3 (CIN3), the proximal precancer to invasive squamous cervical cancer, represents a success in a screening study because it is treatable without hysterectomy, but a rare manifestation of failure to prevent infection in a study of vaccine efficacy. Early trials used CIN2, cervical intraepithelial neoplasia grade 2, attributed to vaccine type as the endpoint; CIN2 is less serious but far more heterogeneous than CIN3. Later trials use persistent type-specific HPV infections as trial endpoint^{3, 4} because of its reproducibility, clear meaning and the ease of assigning HPV type; the assumption that all cervical cancers originate from persistent infection is supported by cross-sectional data.⁵

Functions of precancers in research

As a target for treatment in screening programs

Precancers themselves can be targets for interventions because a precancer allows earlier identification of future cancer before it appears clinically. Interventions to prevent cancer or reduce its consequences in those who previously experienced a rare precancer conferring additional risk of cancer can be more efficient, in the sense of greater benefit per individual receiving the intervention. The basis of successful cervical cancer prevention programs is excision or destruction of the transformation zone of the cervix where almost all epithelial cervical cancers arise after histologic confirmation of an advanced lesion. Colon cancer prevention programs using colonoscopy or virtual colonoscopy eliminate the adenomatous polyps in order to reduce risk of colon cancer and its consequences, including indication for treatment of the cancer and death, for several years by interrupting the natural history of carcinogenesis through elimination of the cells most likely to invade. In principle, detection of a precancer may raise the level of risk of cancer to indicate an intervention, not targeted at the precancer itself, to prevent malignancy.

To identify homogeneous subsets of cancer

Using the precancer to define a subset of complex disease might allow some simplification and increased precision of estimates and power of test of hypothesis not possible when considering a more broadly defined, inherently more complex disease. Head and neck squamous cell carcinomas (HNSCC) may be the best current example where including presence of a precancer in the definition of a cancer subset can be useful. Appropriately

defined precancers might make prevention and treatment of HNSCC more effective. Figure 2 depicts two possible subsets of (HNSCC), distinguished by different hypothetical precancers. In one simplified pathway, smoking and alcohol act together to cause leukoplakia or erythroplakia; in the other pathway, HPV, perhaps with smoking and other cofactors, causes another as yet unidentified precancer that confers high risk of cancer.⁶ Etiology, prognosis, effective screening, molecular features and optimal treatment of the two kinds of oral cancer, distinguished by precancer 1 and precancer 2, may be distinct.^{6, 7}

For cervical cancer, CIN2 caused by HPV16 or HPV18 may be a more useful precancer than all CIN2 when investigating efficacy of a vaccine with only HPV16 and HPV18 antigens. On the other hand, a trial comparing efficacy of a vaccine with two oncogenic antigens against one with seven might use all CIN2 to include the effect of crossprotection against oncogenic types not included in the respective vaccines. In studies of colon polyps, an endpoint definition with restriction to adenomas or high-risk adenomas might increase the homogeneity of risk of malignancy following the endpoint.

More generally, identifying somatic changes, methylation, or pathologic characteristics that signify that a lung cancer was caused at least in part by radon⁸ would allow much easier disentanglement of the joint effects of multiple risk factors.

In analysis of studies

As a surrogate endpoint in clinical trials of prevention programs and epidemiologic studies—Practical considerations for use of surrogate endpoints in non-randomized epidemiologic studies and randomized trials are similar.

A precancer can serve as a surrogate endpoint when the endpoint of true interest is too rare or takes too long to occur, as in trial of an HPV vaccine to prevent cervical cancer. Prentice's criterion⁹ for using a surrogate endpoint in place of the most appropriate endpoint for a randomized trial of an intervention, like disease free survival as a surrogate of death in a trial of a cancer treatment, requires that the effect of the intervention on the endpoint not be modified by whether the precancer was observed. For example, the criterion requires that if the intervention prevents 50% of the precancers, the intervention will prevent 50% of cancer; thus an intervention that is effective when the precursor is present but has no effect on the direct pathway will not meet the criterion. Collection of evidence to show that the precancer surrogate "captures" the full relationship between the exposure or intervention and the cancer⁹ is unrealistically difficult in cervical cancer because the cancer is only observed in the absence of effective cervical cancer screening.

To understand temporal sequence and establish causality

Special epidemiologic and clinical studies of natural history that include precancer endpoints, or, when possible, with both precancer and cancer endpoints, can capture some of the dynamic elements of the disease process and identify determinants of disease that affect the precancer or the transition between the precancer and cancer. A precancer can help establish causality by ruling out reverse causation, or the possibility that the putative cause of disease is an effect of the disease process. The attribution of all grades of CIN to oncogenic HPV infection¹⁰ was a key step in establishing the causal relationship between HPV and cancer¹¹ and in alleviating concerns that HPV was an opportunistic infection that invaded existing cervical lesions.

In study design

Randomized trial with precancer endpoint or *observational* study with cancer endpoint?—The argument over using recurrence of polyps as the endpoint of a trial to

learn about interventions to prevent colon cancer highlights design tradeoffs.^{12, 13} Observational studies with colon cancer endpoints may not be helpful for an intervention that would not be common in an unselected population, like a special diet, and lack randomization. A randomized trial with a colon cancer endpoint would need to be of long enough duration to accrue cancer cases, even in a high risk setting, like those with previously diagnosed polyps. A randomized trial with a polyp endpoint would underestimate the benefit of an intervention that prevents only the polyps most likely to progress to colon cancer; the trial would underestimate the benefit of an intervention that preferentially prevents the polyps that are least likely to progress.

Selection of individuals for study—When possible, followup of patients with a precancer, at high risk for a precancer or using precancer as an endpoint can clarify the natural history of pathogenesis. Repeated collection of biospecimens at short intervals for longitudinal study of markers of exposure and disease process in a natural history study in Guanacaste, Costa Rica¹⁴ allowed important information about acquisition,¹⁵ duration,¹⁶ rate of clearance¹⁷ and serology¹⁸ of HPV infections, from about 3,000 women followed every 6 to 12 months for several years. Safety is critical in these kinds of studies; research goals must not interfere with clinical management of study participants.

The exclusion of precancers from control groups in case-control studies can increase power. Terry and Neugut¹⁹ found stronger association between heavy smoking and risk of colon cancer after excluding rather than including controls with previous diagnosis of adenoma than with a study base of subjects with previous colonoscopy. They interpret their findings as evidence supporting a true smoking-colon cancer effect, which was attenuated in studies including controls with adenomas, possibly caused by smoking. Poole²⁰ points out correctly that the odds ratio for colon cancer based on exclusion of adenomas from the controls does not reflect the association in the underlying population; nor does the odds ratio calculated after the exclusion reflect the strength of the causal relationship. Nonetheless, the use of precancers provides stronger evidence for a causal effect of smoking in colon cancer etiology. As Potter noted,²¹ “improvement in diagnostic classification [might] provide further clarity.”

Pitfalls

Temporal complexities

In practice, all definitions of precancer and cancer are dependent on the timing of diagnosis, which is affected by screening practice as well as the sensitivity and specificity of the methods of diagnosis. Figure 1 includes a time axis as a reminder that the presence of exposure, precancer and cancer, and therefore any inference from a study, are all possibly modified by time of followup. For example, Weiss²² explicitly recommends that case-control studies of the effects of screening on prevention of cancer incidence use a definition of exposure (to screening) restricted to the interval when antecedents to the tumor (such as precancers) are present and detectable; analogously, he recommends that studies of screening for cancer to prevent mortality focus on the interval when the tumor is detectable. Careful attention to temporal issues, particularly length of followup to detect cancer after assessing the precancer, may explain discrepant results among different studies.

When early pathogenic effects affect vital status, willingness to enroll and eligibility for cohort membership, early pathogenic effects might be systematically excluded from followup. This *left-truncation*, sometimes called *left-censoring*, can cause bias in longitudinal studies generally,²³ and particularly in studies of precancers, when those excluded tend to be systematically at higher or lower risk of disease than those included. *Right-censoring* of followup due to the ethical requirement to use an effective intervention

after detecting a treatable precancer makes designing studies of natural history between diagnoses of precancers and colon and cervical cancer²⁴ not feasible.

Ascertainment of precancer

More complete and more accurate ascertainment of precancers will enhance their usefulness. The timing of screening can have a big effect: approximately 40% of detectable but undiagnosed CIN2 may regress over two years.²⁵ Screening decisions based on risk of the precancer or the cancer intrinsically effect ascertainment and thereby cause differential error in diagnosis: an extreme but probably quantitatively trivial example is comparing 5-year risk of CIN2+ after various cytology results, where any extra screening clinically indicated by more serious presentation might exaggerate the difference in risk of a histologically defined precancer.

Differential diagnostic misclassification can occur even in a randomized trial. Assignment of the causal HPV type when multiple types are found in a cervical lesion can be difficult, even when a sequence of earlier HPV specimens is available to compare durations of infections. In early studies of vaccine efficacy of the quadrivalent HPV vaccine, the practice of looking only for HPV16 and HPV18 in a lesion²⁶ may have led to misattribution of the type that caused the lesion. An incident HPV16 infection in a long-lasting lesion caused by HPV45, say, would be improperly counted as an HPV16-related endpoint in the placebo arm; a similar HPV45 lesion in the vaccine arm would likely not show presence of HPV-16 and would not be considered an HPV-16 related endpoint. The consequent differential misclassification of endpoint can exaggerate the benefit of the vaccine.²⁷

Zeal to avoid misclassification can reduce the value of a precancer by reducing homogeneity of risk. The American Society for Colposcopy and Cervical Pathology recently considered combining CIN2 and CIN3 into a single diagnostic category in standard cervical histology because CIN2 diagnosis is demonstrably irreproducible even by expert pathologists. Unfortunately, the consequent composite definition's advantages of reliability in diagnosis and increased sensitivity may lead to less usefulness in screening settings unless, unrealistically, the risks of developing cervical cancer in a reasonable time for those currently called CIN2 and for those currently called CIN3 are similar. Similarly, the positive predictive value for cancer after identification of a new rubric of CIN3 incorporating positivity only for types most likely to lead to cancer will likely be higher than for all CIN3.⁵

Difficulty of inference from effect on precancer to effect on cancer

Etiologic studies using precancers can be informative about pathogenesis. Studies of the association between the exposure and the precancer, however, cannot capture effects of determinants of transition from precancer to cancer. This problem can cause errors of interpretation, as is well-documented in the literature in the analogous area of evaluating therapy with an endpoint of a precancer like cancer progression instead of death.^{9, 28} Note that validity of inference to cancer from a study of a precancer fully justifies the validity of the same surrogate neither for a different intervention nor for a different endpoint.

In addition, the studies with precancer endpoints do not allow direct estimation of the effect of an intervention or exposure on the cancer. Herrero *et al.*³ emphasized the variation in age-specific reduction in risk of cervical precancer after vaccination, but could not provide parallel information about cervical cancer risk reduction; in fact, the number needed to vaccinate to prevent a single cervical cancer case is not directly estimable from the vast HPV vaccination literature, because the proportion of precancers progressing to cancer is unknown.

Theoretical framework

In this section, I discuss quantitative measures more fully and discuss applicability of some of the ideas here to more general case than the strict precancer focus until now.

Quantitative criteria

Population attributable fraction, Penetrance, Phenocopy fraction—As noted by Schatzkin and colleagues,²⁹ population attributable fraction (PAF) can be a measure of the usefulness of a precancer. Indeed, the fraction of the cancers with indirect causal pathways (figure 1) is an upper bound for the percentage of cases that will be prevented by fully effective prevention or treatment of the precancer. In an appendix, I show that the standard definition of PAF as the ratio of the difference between the crude rate of cancer and the rate of cancer in those without the precancer and the crude rate establishes that PAF for precancer is percentage of cancer cases caused by indirect pathways. When a potential precancer is not associated with cancer, the PAF is 0, and the putative precancer has no clinical value.³⁰ PAF is lower than sensitivity because PAF, unlike sensitivity, does not count the diseased cases with the precancer only incidentally, just as the percentage of breast cancers attributable to a BRCA mutation is less than the percentage of breast cancer cases carrying the mutation. When the precancer is necessary, the PAF equals the sensitivity because there are no *phenocopies*, or cases of disease whose causal pathway does not include the precancer; in general the phenocopy fraction is the complement of the population attributable fraction for the precancer.

A countervailing measure of a precancer is *penetrance*, or the proportion of precancers that will progress to cancer by an indirect pathway, that is, as progression of the precancer. Just as PAF is more appropriate measure of value of a precancer than sensitivity, this definition of penetrance is more appropriate than positive predictive value: the penetrance of the precancer is the risk of only those cancers caused by indirect pathways (figure 1), i.e., those that will be affected by an intervention targeting the pathway that includes the precancer. In this definition of penetrance, sporadic breast cancer does not contribute to the penetrance of a BRCA1 mutation. In general, broadening the definition of the precancer to include similar states less likely to progress to disease will reduce penetrance, even as PAF increases. When the precancer is sufficient, penetrance is 1. Of course, defining the precancer to exclude lower risk lesions but will lower the PAF.

Generalizations

Non-clinical states: genetic, somatic, epigenetic, behavioral

The functions and pitfalls discussed in this paper apply to any state for which direct intervention will reduce disease; thus they apply far more broadly than the restricted definition of precancer reserved for pathologic states and requiring direct and proximal relation to cancer. A range of possible examples are BRCA1 mutation; early somatic changes; or epigenetic alterations like methylation or gene expression that could be considered as precancers. In fact, even a behavior like smoking at least 20 life-time pack years of tobacco parallels with a precancer. On the other hand, a biomarker state like high prostate specific antigen (PSA) does not fit into this framework because it is not on the causal pathway: reducing high PSA directly would not affect prostate cancer risk.

Transitions between sequences of precancers

Although standard case control studies of cervical cancer cannot evaluate factors that alter the chance of infection with HPV from cofactors of persistence of infection or progression,³¹ identifying a sequence of precancers on the continuum between health and

cancer can allow investigation of the determinants of the transitions that confer greater risk of cancer. With precise definition, infection, persistence and progression³² could also form a sequence of precancers. In principle, one could consider transitions between any of a series of precancers after meiosis and ending with death. For example, the similarities between smoking behavior and a true precursor are helpful when considering possible intermediates between genetic variation in chromosome 15q and lung cancer.³³

A sequential set of precancers allows examination of the determinants of each transition without amplification or interference from earlier or later transitions. A well-defined, well-measured sequential set of precancers useful in studies of transitions should show increasing population attributable fraction, increasing penetrance or increasing both for the disease of interest.⁴ Additional division into finer categories might further increase the value of this approach by allowing more homogeneity when the frequency of each category is sufficiently large. By contrast, including those who test positive on a truly more sensitive diagnostic test than the standard one can increase heterogeneity if the newly detected precancers tend to confer a lower risk than the average using a standard precancer test; consequently, the specificity will be lower and the new positive predictive value might be too low to justify an aggressive treatment appropriate for those at highest risk.³⁴

For example, Guan et al.²⁴, figure 2 inferred the greater virulence of HPV16, and therefrom, its important heterogeneity among HPV infections, from its successively higher fraction in precancer closer to cancer. Longitudinal studies might be more informative than sets of cross-sectional studies, but are often not feasible because of requirement for long-term followup and need for censoring after early steps in the sequence where useful preventive treatments are known.

When one or more precancer is identified, a simple but computationally-intensive robust and efficient test proposed by Zhang et al.⁴ can identify genetic or other etiologic risk factors that influence specific steps in the progression process; in particular, the method might provide increased power by reducing attenuation if the exposure affects only a single transition or be sensitive to a factor that might act in opposite directions at different transitions. The method extends the case-case study³⁵ to include three or more precancer categories considered in sequence.

Diseases other than cancer

Figure 1 with more general terminology (“precursor” instead of “precancer” and “disease” instead of “cancer”) can describe pathogenesis of diseases other than cancers. Atherosclerosis can be considered a precursor of myocardial infarction, and intermediate in a sequences beginning with their genetic determinants, continuing with early steps such hyperlipidemia and hypertension through atherosclerosis to myocardial infarction or stroke. HIV infection and low CD4 count, alone or together, are precursors for AIDS.

Precursors of death

Many trials have used precursors of death as endpoints in randomized treatment trials. Using precursors, such as disease progression, as a substitute for death can mislead, despite the potential value of quicker studies of new interventions to reduce mortality. Fleming and DeMets²⁸, Table 1, p. 607 list several studies with a survival endpoint “in which biological markers were correlates of clinical outcomes but failed to predict the effect of treatment on the clinical outcome.”

In epidemiologic studies of mortality, cancer at a specified site is a better precursor for death from cancer at that site than for an endpoint of all mortality. All studies of the effect of obesity on mortality³⁶ possibly conflate two disparate effects of increased weight: higher

risk of developing the precursors to cardiovascular disease and cancer and increased survival after diagnosis due to greater ability to withstand the wasting effects of the disease, particularly for cancers that cause cachexia.

Conclusion

Precancers serve many functions in epidemiologic studies. Precancers provide nuance to the disease process beyond the simple dichotomy of case versus control. In fact, one might see a precancer as a tool to reduce misclassification of the endpoint in epidemiologic and clinical studies, thereby allowing more precise estimates of effects and increased power, just as do ordered categories of an exposure like smoking instead of dichotomization into smokers and non-smokers or ever vs. never smokers, or at an arbitrary threshold, like 20 cigarettes per day. Special epidemiologic and clinical longitudinal studies of natural history that ascertain precancer endpoints, or, when feasible, both precancer and cancer endpoints, can capture some of the dynamic elements of the disease process and identify determinants of disease that affect the risk of precancer and the transition between the precancer and cancer. Precancers from cervical cancer and colon cancer, emphasized in this paper, accelerated progress and contributed to exceptional translational benefits at these two sites.

Identification of useful precancers that might allow for effective intervention or screening programs to reduce morbidity and mortality is a continuing challenge in classical and molecular epidemiology. Quantitative criteria specially defined for pathways where the precursor is a causal mediator, including high population attributable fraction and high penetrance can help in evaluation or comparison of the value of proposed precancer definitions chosen to reflect disease progression while addressing problems of temporality and misclassification. Additional attention to the functions and the limitations of precursors might improve epidemiologic practice, and eventually lead to attendant translational benefits. The theory of causal models might profitably incorporate precursors within a broad theoretical model.

Schatzkin³⁷, p. 57, in the 2001 volume³⁸ on precursors, noted that “the large, long, expensive studies required to fully investigate potential surrogates are precisely the studies that surrogates were designed to replace.” More than a decade later, Schatzkin’s ironic point³⁷ is becoming a central tenet of classical and molecular epidemiology as we recognize that challenging studies to identify and characterize precursors provide important insight into natural history of pathogenesis before diagnosis of disease that is missed for cancers without epidemiologically useful precursors. In turn, designs that include ascertainment of appropriate precursors can avoid relying exclusively on disease diagnosis as the starting point or the endpoint, and thereby allow more powerful, more efficient and timelier epidemiologic studies.

References

1. Franco, EL.; Rohan, TE. Introduction. In: Franco, EL.; Rohan, TE., editors. Cancer precursors: Epidemiology, detection and prevention. New York: Springer-Verlag; 2001. p. 1-3.
2. Berman JJ, Albores-Saavedra J, Bostwick D, et al. Precancer: a conceptual working definition -- results of a Consensus Conference. *Cancer Detect Prev.* 2006; 30(5):387–394. [PubMed: 17079091]
3. Herrero R, Wacholder S, Rodriguez AC, et al. Prevention of persistent human papillomavirus infection by an HPV16/18 vaccine: a community-based randomized clinical trial in Guanacaste, Costa Rica. *Cancer Discov.* 2011; 1(5):408–419. [PubMed: 22586631]
4. Zhang H, Wacholder S, Qin J, Hildesheim A, Yu K. Improved genetic association tests for an ordinal outcome representing the disease progression process. *Genet Epidemiol.* 2011; 35(6):499–505. [PubMed: 21618605]

5. Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer*. 2012; 131(10):2349–2359. [PubMed: 22323075]
6. Gillison ML, Castellsague X, Chaturvedi AK, et al. Comparative epidemiology of HPV infection and associated cancers of the head and neck and cervix. *Int J Cancer*. 2013 in press.
7. Zandberg DP, Bhargava R, Badin S, Cullen KJ. The role of human papillomavirus in nongenital cancers. *CA Cancer J Clin*. 2013; 63(1):57–81. [PubMed: 23258613]
8. Vahakangas KH, Samet JM, Metcalf RA, et al. Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. *Lancet*. 1992; 339(8793):576–580. [PubMed: 1347094]
9. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989; 8(4):431–440. [PubMed: 2727467]
10. Schiffman MH, Bauer HM, Hoover RN, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst*. 1993; 85(12):958–964. [PubMed: 8388478]
11. Herrero R, Schiffman MH, Bratti C, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Rev Panam Salud Publica*. 1997; 1(5):362–375. [PubMed: 9180057]
12. Schatzkin A, Freedman LS, Dawsey SM, Lanza E. Interpreting precursor studies: what polyp trials tell us about large-bowel cancer. *J Natl Cancer Inst*. 1994; 86(14):1053–1057. [PubMed: 7802771]
13. Jacobson JS, Neugut AI. Re: Interpreting precursor studies: what polyp trials tell us about large-bowel cancer. *J Natl Cancer Inst*. 1994; 86(21):1648–1649. [PubMed: 7932831]
14. Bratti MC, Rodriguez AC, Schiffman M, et al. Description of a seven-year prospective study of human papillomavirus infection and cervical neoplasia among 10000 women in Guanacaste, Costa Rica. *Rev Panam Salud Publica*. 2004; 15(2):75–89. [PubMed: 15030652]
15. Rodriguez AC, Burk R, Herrero R, et al. The natural history of human papillomavirus infection and cervical intraepithelial neoplasia among young women in the Guanacaste cohort shortly after initiation of sexual life. *Sex Transm Dis*. 2007; 34(7):494–502. [PubMed: 17237737]
16. Rodriguez AC, Schiffman M, Herrero R, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst*. 2008; 100(7):513–517. [PubMed: 18364507]
17. Rodriguez AC, Schiffman M, Herrero R, et al. Longitudinal study of human papillomavirus persistence and cervical intraepithelial neoplasia grade 2/3: critical role of duration of infection. *J Natl Cancer Inst*. 2010; 102(5):315–324. [PubMed: 20157096]
18. Wentzensen N, Rodriguez AC, Viscidi R, et al. A competitive serological assay shows naturally acquired immunity to human papillomavirus infections in the Guanacaste Natural History Study. *J Infect Dis*. 2011; 204(1):94–102. [PubMed: 21628663]
19. Terry MB, Neugut AI. Cigarette smoking and the colorectal adenoma-carcinoma sequence: a hypothesis to explain the paradox. *Am J Epidemiol*. 1998; 147(10):903–910. [PubMed: 9596467]
20. Poole C. Controls who experienced hypothetical causal intermediates should not be excluded from case-control studies. *Am J Epidemiol*. 1999; 150(6):547–551. [PubMed: 10489992]
21. Potter JD. Invited commentary: old problem, new wrinkles. *Am J Epidemiol*. 1998; 147(10):911–913. [PubMed: 9596468]
22. Weiss NS. Case-control studies of the efficacy of screening tests designed to prevent the incidence of cancer. *Am J Epidemiol*. 1999; 149(1):1–4. [PubMed: 9883787]
23. Applebaum KM, Malloy EJ, Eisen EA. Left truncation, susceptibility, and bias in occupational cohort studies. *Epidemiology*. 2011; 22(4):599–606. [PubMed: 21543985]
24. McCredie MR, Sharples KJ, Paul C, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncol*. 2008; 9(5):425–434. [PubMed: 18407790]
25. Castle PE, Schiffman M, Wheeler CM, Solomon D. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol*. 2009; 113(1):18–25. [PubMed: 19104355]
26. Koutsky LA, Ault KA, Wheeler CM, et al. A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med*. 2002; 347(21):1645–1651. [PubMed: 12444178]

27. Wacholder S, Lubin JH, Dosemeci M, Gail MH. Bias despite masked assessment of clinical outcomes when an outcome is defined as one of several component events. *Control Clin Trials*. 1991; 12(4):457–461. [PubMed: 1657524]
28. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996; 125(7):605–613. [PubMed: 8815760]
29. Schatzkin A, Freedman LS, Schiffman MH, Dawsey SM. Validation of intermediate end points in cancer research. *J Natl Cancer Inst*. 1990; 82(22):1746–1752. [PubMed: 2231769]
30. Hartge P, Hayes R, Reding D, et al. Complex ovarian cysts in postmenopausal women are not associated with ovarian cancer risk factors: preliminary data from the prostate, lung, colon, and ovarian cancer screening trial. *Am J Obstet Gynecol*. 2000; 183(5):1232–1237. [PubMed: 11084571]
31. Wacholder S. Chapter 18: Statistical issues in the design and analysis of studies of human papillomavirus and cervical neoplasia. *J Natl Cancer Inst Monogr*. 2003; (31):125–130. [PubMed: 12807956]
32. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007; 370(9590):890–907. [PubMed: 17826171]
33. Wacholder S, Chatterjee N, Caporaso N. Intermediacy and gene-environment interaction: the example of CHRNA5-A3 region, smoking, nicotine dependence, and lung cancer. *J Natl Cancer Inst*. 2008; 100(21):1488–1491. [PubMed: 18957674]
34. Wentzensen N, Wacholder S. From differences in means between cases and controls to risk stratification: A business plan for biomarker development. *Cancer Discovery*. 2013 in press.
35. Fleming TR, DeGruttola V, DeMets DL. Surrogate endpoints. *AIDS Clin Rev*. 1997:129–143. [PubMed: 9305447]
36. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*. 2013; 309(1):71–82. [PubMed: 23280227]
37. Schatzkin, A. Intermediate makers is cancer research: theoretical and practical issues in the use of surrogate endpoints. In: Franco, EL.; Rohan, TE., editors. *Cancer precursors: Epidemiology, detection and prevention*. New York: Springer-Verlag; 2001. p. 46-59.
38. Franco, E.; Rohan, T. *Cancer precursors: Epidemiology, detection and prevention*. New York: Springer-Verlag; 2001.
39. Wacholder S. The impact of a prevention effort on the community. *Epidemiology*. 2005; 16(1):1–3. [PubMed: 15613938]

Appendix

This appendix shows that population attributable fraction (PAF) is a useful measure of mediation between disease and exposure because it is the percentage of cases that manifest the precursor as part of pathogenesis. By the fundamental definition of PAF,

$$PAF = \frac{I_c - I_0}{I_c} = \frac{p_0 I_0 + p_1 I_1 - I_0}{I_c} = \frac{p_1 (I_1 - I_0)}{I_c}$$

where p_0 is the proportion of those contributing to the crude rate without the precursor and $p_1 = 1 - p_0$ is the proportion with the precursor, and I_0 and I_1 are the fractions of those without and with the precursor who develop disease. That is, the PAF numerator is a weighted average of the risk difference in the unexposed (0) and the exposed ($I_1 - I_0$), with the weight being the proportion at each exposure level, equivalent to the product of the proportion with the precursor and the relative increase in disease risk from having the precursor. That is, the PAF is the proportion of cases that can be prevented or treated from an intervention that is fully effective against the subset of disease for which the precursor condition is part of the same pathogenesis process; the numerator of PAF, $p_1(I_1 - I_0)$, called attributable community risk (ACR)³⁹, is the proportion of the population that will benefit from the fully effective intervention. PAF and ACR are valid causal measures as long as the risk difference is a causal estimate and the proportion of the population experiencing the precursor is accurate. Typically, the precursor is

underascertained, so the rate I_0 includes some with the precursor and therefore at higher risk for disease, and the true PAF and ACR are underestimates.

We can define the PAF directly from case control data in terms of sensitivity and specificity and their complements, $cSensitivity = 1 - Sensitivity$ and $cSpecificity = 1 - Specificity$, where sensitivity and cSpecificity are the proportion of cases and non-cases who have

experienced the precursor: $PAF = \frac{cSpec(OR-1)}{1+cSpec(OR-1)}$, where $OR = \frac{Sens}{cSens} \frac{cSpec}{Spec}$.

The PAF of CIN3 for epithelial cervical cancers is near 100%; the PAF of CIN3 for adenocarcinoma is much lower, reflecting the poorer performance of Pap-based screening for detecting adenocarcinomas.

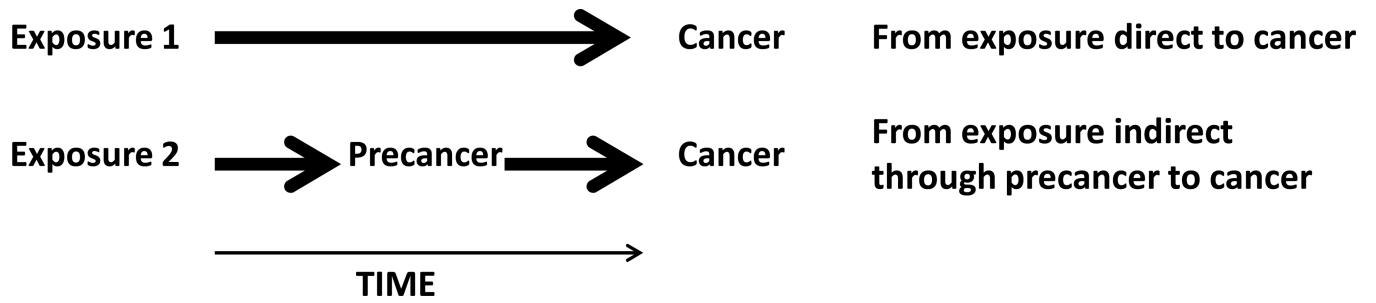


Figure 1. Depictions of simple *causal pathways* to cancer

The top causal diagram shows exposure 1 leading to cancer directly without passing through a known precancer step. The bottom causal diagram shows exposure 2 leading to cancer indirectly by causing the precancer, from which the cancer eventually emerges.

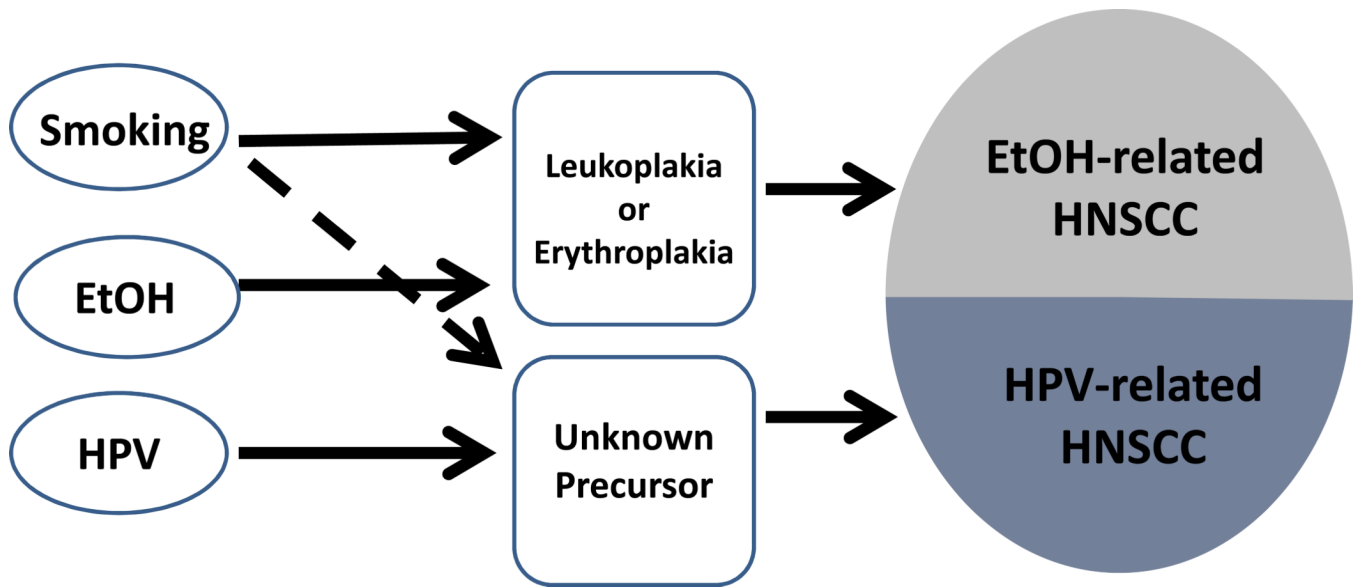


Figure 2. Different Precancers might distinguish two pathways for head and neck squamous cell carcinomas (HNSCC)

Leukoplakia and erythroplakia are precancerous lesions for HNSCC caused by alcohol (EtOH) and smoking. No precancer lesions of other HNSCCs, mostly caused by HPV are known. Easily detectable precancers could be the basis for effective screening programs to prevent HNSCC or could indicate different management strategies for patients after diagnosis.⁶