

The Effectiveness of Chemoprevention Agents Is Underestimated When Lesion Sizes Are Rounded

Thomas H. Taylor^{1,2,3}, William B. Armstrong^{3,4,5}, and Frank L. Meyskens^{3,4,6}

Abstract

Change in the area of premalignant lesions is an end point in estimating the efficacy of chemopreventive agents. When examiners round measurements of lesion length and width, they introduce variability, which perturbs the relative percent change in lesion area and, consequently, the percent of subjects showing a clinical response. We use simulations to illustrate the resulting bias when the agent under test is effective in reducing lesion area. We simulated 500 oral leukoplakia lesions per run, with 2,500 runs at each of five levels of agent effectiveness, namely, true relative percent reduction in area of 25%, 45%, 50%, 55%, and 75%. Realistic values of lesion lengths and widths were generated randomly and then rounded to the nearest multiple of five. The product is the distribution of mean relative percent change in lesion area and the corresponding percent of subjects showing a clinical response. Even the fifth percentile of the distribution of mean relative percent change in lesion area consistently underestimated the true value, by about 6 percentage points. The percent showing a clinical response was underestimated by 50%, 37%, and 11% for true values of reduction in lesion area of 50%, 55%, and 75%, respectively. This could easily double the required sample size for a modest phase II study. We suggest that it is cost-effective to train observers of lesion length and width to eschew rounding of measurements in the chemoprevention setting. *Cancer Prev Res*; 3(2); 136–9. ©2010 AACR.

Introduction

Changes in the size of premalignant lesions remain an important end point for gauging the effect of cancer prevention agents. Whereas better understanding of tissue transformation at the molecular level sustains hope for surrogate end points useful for the evaluation of chemoprevention agents (1–3), at present, clinical end points seem necessary for formal assessment of response. Clinical end points often include relative percent change in the area of premalignant lesions and the percent of cases who exhibit at least a partial response derived from the relative percent change in lesion area. Underlying these clinical end points are estimates of lesion length and width made by clinical examiners. We have noted a tendency for some clinical examiners to round measures of lesion length and width. Here, we report simulation results to illustrate the consequences of rounding lesion measurements on relative percent change in lesion area and the percent showing a

clinical response, when treatment is effective in reducing lesion areas.

In our experience with oral leukoplakia as a model of precancerous tissue (4), we noticed that some examiners record lesion lengths and widths predominantly as multiples of five (i.e., 5, 10, 15, 20, etc.), at the expense of plausible values close to multiples of five (e.g., 4, 6, 9, 11, 14, 16, etc.). Because the borders of oral leukoplakia can be indistinct and the underlying tissue is elastic, examiners may be tempted to round measurements of lesion extent. Some examiners may feel that rounding errors will wash out across lesions, time points, and subjects. Although rounding may have little effect on estimates of the mean length or mean width of lesions, it does affect measures based on lesion area.

We simulate what may be cast as a single-arm trial of a hypothetical agent that diminishes the area of oral leukoplakia lesions. Lesions are measured at two time points, corresponding to pre- and post-treatment. There are five levels of agent effectiveness: relative reduction in lesion area of 25%, 45%, 50%, 55%, or 75%. This could represent dose escalation of a single agent or agents of differing potency. Primary end points are relative percent change in lesion area from pre- to post-treatment and the corresponding proportion exhibiting at least a partial clinical response. Results from rounded measurements are compared with nominal results, and implications for study sample size and power are illustrated.

Materials and Methods

For each of the five levels of agent effectiveness, we ran 2,500 iterations of 500 simulated lesions each.

Authors' Affiliations: ¹Genetic Epidemiology Research Institute, ²Department of Epidemiology, ³Chao Family Comprehensive Clinical Cancer Center, ⁴Department of Medicine, ⁵Department of Otolaryngology/Head and Neck Surgery, and ⁶Department of Biological Chemistry, University of California-Irvine, Irvine, California

Note: Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

Corresponding Author: Thomas H. Taylor, Department of Epidemiology, University of California-Irvine, 224 Irvine Hall, Irvine, CA 92697-7550. Phone: 949-824-6903; Fax: 949-824-1343; E-mail: thtaylor@uci.edu.

doi: 10.1158/1940-6207.CAPR-09-0114

©2010 American Association for Cancer Research.

Pretreatment lesion lengths and widths were from a uniform, random distribution, such that 80% were 1 mm through 20 mm, and 20% were 21 mm through 30 mm. This corresponds well to real-world measures of oral leukoplakia (4).

To incorporate agent activity, posttreatment lengths and widths were set to the corresponding pretreatment values divided by the positive square root of a factor achieving the desired reduction in lesion area. For reductions of 25%, 45%, 50%, 55%, and 75%, each posttreatment dimension was divided by the positive square root of 1.333, 1.818, 2.000, 2.222, and 4.000, respectively. The simulations do not incorporate individual variability in response to treatment nor the natural waxing and waning of oral leukoplakia.

For each lesion, a set of rounded pre- and post-treatment measures was made by rounding each measure to the nearest multiple of five, substituting 1 for results of 0. Pre- and post-treatment estimates of both original and rounded lesion areas were formed from the product of the appropriate length and width measures. Relative percent change in lesion area was calculated separately from rounded and unrounded measures, as 100 times posttreatment area minus pretreatment area, all divided by the pretreatment area. A partial response was recorded if the relative percent change was a reduction of at least 50%.

Simulations were programmed in SAS software (5). Power and sample size estimates are for Fisher's exact test of proportions (6), calculated using the nQuery program (7). For mathematical considerations of the net effect of rounding and consequences of correlated lengths and widths, we repeated the work, randomly generating lesion

lengths from 3 mm through 32 mm (80% 3-22 mm, 20% 23-32 mm) and setting baseline widths to baseline lengths. The results are given in Supplemental Data.

Results

Table 1 shows, for each level of agent effectiveness, the values of mean relative percent change in lesion area and the proportion showing at least a partial response. Results from the unrounded data as well as those from the rounded data are given. As expected, there is no variability in the results from unrounded data. From unrounded data, both the mean relative percent change in lesion area and the percent achieving at least a partial response are nominal across the effectiveness levels.

To show variability in outcomes from the rounded data, Table 1 gives the average and the 5th, 50th (median), and 95th percentiles of both the distribution of mean relative percent change and the proportion showing at least a partial response from rounded measurements. The true values of mean relative percent change are consistently underestimated by the rounded data, as reflected by the mean or any percentile shown. A simple linear regression of the mean relative percent change score from rounded data against the true relative percent change yields a coefficient of 0.937 ($P < 0.0002$) and an R^2 of $>99\%$. (For the fifth percentile: coefficient = 0.941, $P < 0.0002$, $R^2 > 99\%$.) Thus, consistent with appearances, the values of relative percent change from rounded data underestimate the true value by about 6 percentage points.

Table 1 also gives results for the percent of lesions showing at least a partial response (here defined as $\geq 50\%$ or

Table 1. Effect of rounding lesion measurements on average relative percent change in lesion area and percent showing at least a partial response, from simulations

True percent reduction (built into simulation)	0.25	0.45	0.5	0.55	0.75
Unrounded data					
Mean rel % change	-25	-45	-50	-55	-75
% PR or better	0	0	100	100	100
Rounded data					
Mean rel % change	-18.7	-40.6	-43.2	-48.8	-65.8
5th percentile	-20.1	-42.3	-44.9	-50.4	-67.4
50th percentile	-18.7	-40.6	-43.2	-48.8	-65.8
95th percentile	-17.4	-39.0	-41.6	-47.3	-64.3
Mean % PR or better	10.4	43.2	48.5	62.3	88.8
5th percentile	8.2	39.6	44.8	58.8	86.4
50th percentile	10.4	43.2	48.6	62.2	88.8
95th percentile	12.6	46.8	52.2	65.8	91.0

NOTE: Five hundred lesions per iteration; 2,500 iterations at each level of true percent reduction.
Abbreviation: PR, partial response.

Table 2. Estimated effect of rounding measures of lesion length and width on observed proportion achieving partial response or better, and consequences for study power and required sample size

	Placebo arm	Treated arm				
		Attenuation of observed PR or better due to rounding (appx)				
		0%	10%	20%	30%	40%
Showing PR or better (<i>n</i>)	7	22	20	18	15	13
% PR or better	11.7	36.7	33.3	30.0	25.0	21.7
Not showing PR or better (<i>n</i>)	53	38	40	42	45	47
<i>n</i> *	60	60	60	60	60	60
Power (%)		86	76	62	38	22
<i>n</i> * required for minimum 80% power		51	65	85	143	234

NOTE: Fisher's exact test (two-tailed), $\alpha = 0.05$.

*Sample size in each of the two arms.

better) when measures of length and width are rounded. All levels of effect size yielded some partial (or better) responses, even when the effect size is smaller than the threshold defining a partial response (i.e., relative percent changes of -25% and -45%). At effect sizes at which 100% of lesions should show at least a partial response (i.e., relative percent changes of 50%, 55%, and 75%), rounding has attenuated the observed proportion. Thus, rounding of length and width data overestimates the proportion responding at lower levels of true efficacy and underestimates the proportion responding at higher levels of true efficacy. Whereas inspection shows some variability, it seems that rounding of lengths and widths underestimates proportion, showing a partial (or better) response by about 50%, 37%, and 11% for mean true percent reductions in area of 50%, 55%, and 75%, respectively.

Table 2 shows the effects of an attenuated proportion achieving at least a partial response on study power and required sample size as might result from rounded measures of lesion dimensions. Figures in the table are for Fisher's exact test, assuming that 11.7% of patients in the placebo arm show a partial (or better) response, against several attenuated values for the treated arm. When 36.7% of the treated subjects show a partial (or better) response, then 60 subjects per group yields 86% power. As the observed percent showing at least a partial response in the treated arm declines due to attenuation, power deteriorates. Table 2 also shows the estimated sample size to maintain 80% power in each of the depicted situations. That sample size is more than doubled when rounding has attenuated the observed proportion responding by 30%.

Discussion

These simulations show that rounding of lesion widths and lengths, in the presence of a treatment that reduces

lesion area, yields a net underestimate of mean relative percent change in lesion area at every level of agent efficacy examined. When the agent is producing a true reduction in lesion area of 50% or more, then the associated, observed percent showing a clinical response is underestimated. Sample size in chemoprevention studies is often based on anticipated clinical response. As we show in Table 2, the effect of attenuating observed clinical response in a typical phase II study can substantially reduce study efficiency. A 30% reduction in observed clinical response can easily halve the power of such a study and more than double the sample size to maintain study power.

Certainly, the act of rounding measurements increases their variability, thus the appearance of some partial (or better) responses when the true effect of the simulated agent is below the threshold defining a partial (or better) response (Table 1, built-in reductions of 25% and 45%). However, although the process is stochastic, the typical net effect of rounding is to attenuate the relative percent change in area and, consequently, the percent showing at least a partial response. An overestimated dimension amplifies the estimate of lesion area because of multiplication by the other dimension. Insofar as the treatment is effective in reducing true lesion area, then this amplification is greater for pretreatment lesions than for posttreatment lesions because the range of numbers for lengths and widths is greater at pretreatment than at posttreatment. Although, at first, this might be seen as a bias toward a positive result, the fact that the relative percent change involves division by the pretreatment area produces the opposite effect. The difference in lesion area between pre- and post-treatment is being divided by an amplified pretreatment figure. Thus, the bias is toward reducing the observed relative percent change in lesion area. Therefore, the proportion achieving at least a partial clinical response is underestimated.

These results are discomfoting for searches of modest size for agents of moderate chemopreventive potential. Of course, in the real world, the biases from rounded

measurements would mix with individual variation in response to treatment and any natural waxing and waning of lesion sizes, as well as depend on the proclivity of observers to round. All of these affect study efficiency. However, whereas variation in response to treatment and natural waxing and waning are beyond our control, we can encourage precision of measurements by clinical observers.

In our experience, accruing 120 people with oral leukoplakia into a phase II trial requires multiple performance sites because eligible participants are relatively rare. Multiple performance sites entail multiple observers of lesion extent. This situation is similar to that reported for blood pressure screening programs, in which many observers taking blood pressure measures from auscultation and sphygmomanometers rounded measurements (8, 9). The solution for blood pressure screening programs was careful training and the advent of more objective measurement tools.

In the chemopreventive setting, more objective measurement of chemopreventive activity awaits better understanding of premalignant tissues. At present, changes in size of premalignant lesions are a mainstay in evaluating putative chemoprevention agents. There is a need to train

clinical observers in the measurement of such lesions for the best possible study efficiency. Although not a trivial task, our simulations suggest that such training will pay off in fewer false-negative outcomes of chemoprevention trials.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr. Catherine Diamond and anonymous reviewers for comments on earlier drafts.

Grant Support

NIH/National Cancer Institute grants 5 U01 CA072294-11 and P30-CA62203 (F.L. Meyskens).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 6/8/09; revised 8/12/09; accepted 8/26/09; published OnlineFirst 1/19/10.

References

1. Kelloff GJ, Lippman SM, Dannenberg AJ, et al. AACR Task Force on Cancer Prevention. Progress in chemoprevention drug development: the promise of molecular biomarkers for prevention of intraepithelial neoplasia and cancer—a plan to move forward. *Clin Cancer Res* 2006;12:3661–97.
2. Armstrong WB, Taylor TH, Meyskens FL. Point: Surrogate end point biomarkers are likely to be limited in their usefulness in the development of cancer chemoprevention agents against sporadic cancers. *Cancer Epidemiol Biomarkers Prev* 2003;12:589–92.
3. Kelloff GJ, O'Shaughnessy JA, Gordon GB, Hawk ET, Sigman CC. Counterpoint: Because some surrogate endpoint biomarkers measure the neoplastic process they will have high utility in the development of cancer chemopreventive agents against sporadic cancers. *Cancer Epidemiol Biomarkers Prev* 2003;12:593–6.
4. Armstrong WB, Kennedy AR, Wan XS, et al. Clinical modulation of oral leukoplakia and protease activity by Bowman-Birk inhibitor concentrate in a phase IIa chemoprevention trial. *Clin Cancer Res* 2000;6:4684–91.
5. SAS Institute, Inc. SAS OnlineDoc®, version 8, February 2000. Copyright ©2000, SAS Institute, Inc. Available from: <http://v8doc.sas.com/sashtml/>.
6. Freeman DH. Applied categorical data analysis. New York: Marcel Dekker, Inc. 1987.
7. Elashoff JD. NQuery Advisor 4.0. Boston (MA): Statistical Solutions Ltd; 1995.
8. Hessel PA. Terminal digit preference in blood pressure measurements: effects on epidemiological associations. *Int J Epidemiol* 1986;15:122–5.
9. Thavarajah S, White WB, Mansoor GA. Terminal digit bias in a specialty hypertension faculty practice. *J Hum Hypertens* 2003;17:819–22.