

SMOKELESS TOBACCO AND ORAL CANCER;
OVERSTATEMENT OF THE ASSOCIATION
AND PUBLICATION BIAS *IN SITU*

Carl V. Phillips, MPP PhD

University of Texas School of Public Health and Center for Clinical Research
and Evidence Based Medicine, UT Medical School
and
Center for Philosophy Health and Policy Sciences, Inc.

poster presentation at Society for Research on Nicotine and Tobacco
19 February 2004

Background

Publication Bias *in situ* (PBIS)

I coined this term and explored its implications in a previous paper (Phillips 2003a; manuscript under review, available at www.epiphi.com/papers).

In short, it is the tendency of researchers to pick and choose among the results they might report from a study to emphasize those that are bigger (further from the null), statistically significant, or that conform to prior beliefs or hypotheses.

Researchers frequently desire to have striking and conclusive results so that a "better" journal will want to publish them, they will please funding agencies who prefer to support researchers with definitive results, and researchers can further their own policy goals.

For a given data set, there are countless ways to analyze data and enough random variation to let authors choose preferred results. (I previously presented examples of choosing functional forms (e.g., choosing cutpoints for categorizing a continuous variable or subgroups to analyze.) Since there will generally be no clearly right way to analyze the data, and cataloging all possibilities is impractical, readers will seldom be able to identify when bias has been created by selective reporting.

Implications of PBIS

- Systematic bias in the literature toward bigger or more striking results.
- It is much worse than traditionally defined publication bias (where the results of an "uninteresting" study are forgotten in a file drawer rather than published) because it is more difficult to detect and probably more prevalent.
- Can be substantially responsible for conclusions not supported by the data, as in the present example.

Smokeless Tobacco (ST) and the Risk of Oral Cancer (OC)

Most people (consumers, clinicians, and researchers included) "know" that the use of smokeless tobacco (ST) is a substantial risk factor for oral cancer (OC), and thus a major health hazard. This is the overwhelming message in popular communication on the point (Phillips, Guenzel B, and Wang C 2004).

But the epidemiologic evidence tends to disagree with this claim. The two published studies that use modern methods and have sufficient sample size, Schildt et al. (1998) and Lewin et al. (1998) provide strong support for the claim that ST is not associated with OC or the association is too small to detect.

The widespread beliefs and unrelenting claims by public health opinion leaders are presumably buoyed by the studies that did report a positive association.

(Note that ST currently refers mostly to dipping moist snuff, though it also includes chewing tobacco and dry snuff, which were more popular historically. The products in question are primarily U.S. and Swedish. Other smokeless products that combine tobacco with other ingredients that are popular elsewhere in the world have substantially different epidemiology.)

Studies Reporting an Association of ST and OC

Winn et al. (1981) discussed in more detail below, is the source of most claims that there is an association (see also our other poster in this session).

In the 23 years since that publication, there have been about 20 published articles or abstracts (depending on exactly what is counted) presenting epidemiologic data on the ST-OC relationship.

(Bjelke E and Schuman L 1982; Winn DM et al. 1982; Wynder EL et al. 1983; Young TB, Ford CN, and Brandenburg JH 1986; Stockwell HG and Lyman GH 1986; Spitz MR et al. 1988; Blot WJ et al. 1988; Franco EL et al. 1989; Sterling TD, Rosenbaum WL, and Weinkam JJ 1992; Maden C et al. 1992; Marshall JR et al. 1992; Zahm SH, Heineman EF, and Vaught JB 1992; Mashberg A et al. 1993; Kabat GC et al. 1994; Bundgaard T et al. 1995; Muscat JE et al. 1996; Schildt EB et al. 1998; Lewin F et al. 1998; Schwartz SM et al. 1998; Bouquot and Meckstroth 1998; Lee JL et al. 2000; Accortt et al. 2002)

Of these, 8 can be characterized as claiming to support a positive association.

(Blot WJ et al. 1988; Winn DM et al. 1982; Kabat GC et al. 1994; Bjelke E and Schuman L 1982; Marshall JR et al. 1992; Maden C et al. 1992; Stockwell HG and Lyman GH 1986; Spitz MR et al. 1988)

Analysis of PBIS in the Literature on ST and OC

In addition to the usual incentives for researchers to find "better" results, the positions of most government and health-advocacy organizations are strongly anti-ST, so there is a lot of incentive to try to claim there is an association to please funding agencies, generate attention, or further the advocacy goals of the researcher.

All 8 post-Winn studies that purport to find an association between ST and OC are underpowered. The small sample size creates statistically unstable results, making it particularly likely that some identifiable subgroup will have a high relative risk. This creates substantial opportunity to distort the results by PBIS.

I examined Winn et al. (1981) and the 8 subsequent articles that claim to show an association for signs of PBIS. In three cases, it was possible to clearly identify the PBIS and calculate unbiased results from data presented, though the authors chose not to report the unbiased results and the reviewers and editors apparently did not ask for them. In another case, there appeared to be PBIS, but numbers could not be calculated. PBIS as a result of choosing to report only certain subgroups could be ruled out in only a few of these publications.

(Because of the many undetectable choices authors make about which results to report, PBIS cannot generally be ruled out entirely.)

(The analysis is restricted to more recent studies because older studies have poor methodology (e.g., no attempt to control for the effect of smoking, the largest risk factor for OC) so it is difficult to make sense of them, and exposures were to products that are substantially different than popular current products. The choice of the Winn publication as the cutpoint is somewhat arbitrary, chosen because that study was the landmark event in this literature.)

(Note: The discussion of significance and confidence intervals is meant to emphasize how they contribute to PBIS, not to imply that they are very useful statistics, a topic I discuss elsewhere. (Phillips 2003b; Phillips and LaPole L.M. 2003; Phillips and Goodman 2001) For present purposes, it is sufficient to recall that the set of all statistically significant results is biased away from the null, so reporting only such results guarantees a bias away from the null (and reporting only those that are significant and in the "right" direction further increases this bias).

Winn et al. (1981)

- The source most widely cited to support claims that smokeless tobacco causes oral cancer (Phillips, Guenzel B, and Wang C 2004).
- Women in central North Carolina in the 1970s, a population who, among the elderly, had a particularly high history of ST exposure (mostly dry snuff produced in the first half of the 20th century).
- Matched case-control design, with 255 cases and 502 noncases.
- Based on Winn's doctoral dissertation.

The two most commonly presented values for the effect of ST on the risk of OC are the two numbers reported in the abstract. (see Phillips & Guenzel, other poster in this session and poster presentation from Society for Epidemiologic Research, June 2003, available at www.epiPhi.com/papers)

First Commonly Reported Value: Odds Ratio of 4.2

The 4.2 (95% CI 2.6-6.7) is the crude estimate for ST users compared to non-tobacco users (often rounded to 4 when reported). This is the relative risk for all anatomic subsites in the study (oral and some pharyngeal) for nonsmoking white (more precisely, non-black) women.

This is the largest estimated OR of any smoking/race stratum in the data.

Unlike the more egregious cases of PBIS presented above, the authors do report other stratum-specific results in their table, but they otherwise ignore them. Singling out the largest result for the text discussion and the abstract (and thus most of what the world would take away from this study) is an example of PBIS.

- The excluded black women had a much lower relative risk and there is no apparent reason for excluding them (indeed, there is seldom much justification for stratifying results based on race, but that is a different topic). The estimated OR for black nonsmokers who used ST (compared to black non-tobacco users) was only 1.5 (95% CI 0.5-4.8). Combining all races would have given an OR of 3.6 (95% CI 2.3-5.6). This is still elevated, but reporting 4.2 as the main result while not mentioning the 1.5 outside the table and not reporting the 3.6, is clearly an upward bias.
- ST users who also smoked had lower risk. Smoking is a confounder with a large effect on OC risk, so failure to control can substantially elevate the apparent effect of ST in many studies when many subjects use both. Thus, stratifying on it is justified. But the relative risk for smokers (compared to nonsmokers) in this data is strangely low. ST users who also smoked had an estimated OR of a bit above 3 compared to non-tobacco users. Thus, if users of both products had been counted as ST users (keeping the denominator those who used neither product) the resulting estimated OR would have been lower. Controlling for smoking status is probably the best choice, but the fact that the estimated OR would have been lower (rather than the expected higher) without doing it, and that the OR for smoking was so low, should have been mentioned.

Second commonly reported Value: Odds Ratio of 50

From the abstract: "among chronic users the risk approached 50-fold for cancers of the gum and buccal mucosa". The table reporting this shows a higher OR for the gum and buccal mucosa (with almost no increase at other sites) and among them, a higher OR for the 50+ year users compared to nonusers (OR=47.5, rounded to 50 in the abstract) than for 25-49-year users or 1-24-year users (both with ORs of about 13). The number comes from an analysis that was presumably intended to help argue that the association was causal by (a) comparing the oral sites with most direct contact with snuff to other sites to support biological plausibility and (b) looking at a dose-response relationship. (The goal of this analysis has to be guessed at because it is not explained.)

- Even if this result were unbiased, its reporting in the abstract is, in itself PBIS, and has led to widespread misinformation. Reporting the result for the extreme exposure duration (in this category, the average exposed subject started before age 10, in the 1910s, continuing to the 1970s) as if it is an estimate of the average effect in chronic users is misleading. This is made worse by many secondary sources reporting this number as the OR for the effect of ST on OC in general, even though the gum and buccal mucosa sites account for only a small portion of all OC. (The misinterpretations cannot be blamed wholly on the biased reporting in the article, but it is certainly facilitated by it.)
- The astonishingly high OR comes from choosing only a portion of the data, such that the unexposed group included very few cases. By restricting this analysis to nonsmokers and eliminating 17 of the 41 unexposed cases by looking only at the subsample obtained through hospital records (rather than death certificates), only 2 of unexposed subjects had cancer at the particular subsites. This compares to 34 of the 169 unexposed noncases remaining after the cuts.

- It is not clear why duration of use was chosen as the measure of dose. The data contains information sufficient to calculate the total lifetime quantity of snuff consumed by a subject, an actual measure of dose. Defining the highest dose category as ≥ 2340 cans of snuff used (chosen to include the same number of cases as in the longest-duration group), the crude OR is only 19 (or 16 following the methodology of the article, which added .5 to each cell in the 2x2 table before calculating). Other cutpoints for the highest dose produce similar results. This is obviously still high, but not nearly as high as the published result.
- Alternative cutpoints for duration of use show different results. For example, defining the cutpoint for the highest dose category as ≥ 45 years yields an OR in the order of 20 rather than 50, and the resulting dose-response is almost flat. A cutpoint of ≥ 55 produces results slightly lower than ≥ 50 . (It is possible to get higher ORs still but increasing the age cutpoint because only one control in this subsample had ≥ 60 years of exposure while several cases remain.)

- Various other alternative ways to analyze the data also tend to attenuate the highest OR. For example, counting anyone who quit using snuff by 1940 as a nonuser reduces the OR by about half, by doubling the number of unexposed cases in the subsample to 4. (In the interests of pointing out that this choice is not PBIS, changing that threshold to 1930 or 1950 would lower the OR further still.)

In short, while not every possible alternative way of analyzing the data and reporting the results produces less dramatic results than appear in the article, most reasonable alternatives do so. Almost all such results still show an elevated risk, but magnitude matters, especially when a study is the source of most claims about the topic.

(This is not to say that this article is unusual; making choices that produce "better" results may be more common than not in the health literature.)

As is often the case when there is reason to suspect PBIS, these choices are defensible: The hospital subsample permitted an interview of the subject rather than proxy data, so the authors said it was higher quality data. Perhaps years of use is the best measure of dose (though no such argument appears in the article). Perhaps there is a good reason why the highest category should be ≥ 50 rather than 45.

It is impossible to ever know whether the authors chose to report certain results because they were larger or because they believed they were the right way to analyze the data. Indeed, the nature of human psychology means that the authors themselves might not even be able to answer that. However, it seems safe to assume that authors who have several years to pour over a dataset (recall that this was dissertation research) run many alternative analyses, and the failure to report them is the heart of PBIS. PBIS occurs when the alternative measures of dose, the much lower result for 45, etc. are not even mentioned.

Blot et al., (1988)

- Case-control study of OC that primarily reported relative risks from smoking and alcohol use
- 1114 cases (762 male, 352 female) and 1268 noncases (837 male, 431 female); only about 121 used ST ("about" because they report only percentages, so there is rounding error in calculating the exact count)
- Authors address ST only in the prose (not in tables), reporting exposure in 6% of male cases (accounting for rounding error, this is in the range 42 to 49), 7% of male noncases (55 to 62), 3% of female cases (9 to 12), and 1% of female noncases (3 to 6, though the following information narrows it to 4 to 6).

The reported result is a perfect case of PBIS: The authors report only one odds ratio (OR) for ST use, 6.2 (95% CI 1.9-19) for nonsmoking females based on 6 exposed cases and 4 exposed noncases. The authors say nothing to hint that this subgroup's result is not representative of the association in the whole study population.

But calculating the exposure-disease distribution of the population as a whole from the percentages given, we get an OR of 1.0 (95% CI 0.7-1.5), which is highly incompatible with the reported OR of 6.2, and is supportive of the claim of approximately zero elevated risk.

Since the OR for the population as a whole is very close to 1.0, the elevated risk reported for the one subgroup must be mirrored by a reduced risk for other subgroups. The fact that the omitted results were not statistically significant in no way excuses the omission. Indeed, a practice of reporting only statistically significant results means that studies that tend to show a null relationship (like this one) never report that finding, guaranteeing that there will be publication bias (of the traditional or PBIS variety).

It might be argued that Blot et al. emphasized the result for nonsmoking women because this was the population that produced the main result in the Winn study (the articles have three authors in common, including Winn and Blot). However, that was the main result in the earlier study because Winn studied only women and very few of the ST users smoked (or perhaps because the smokers had a lower relative risk), not because of any biological theory that the effect of ST on OC is particular to nonsmoking women. It seems unlikely that had that subgroup shown a protective effect, with the population as a whole showing a null association, that the authors would have singled it out for emphasis.

Winn et al. (1982)

- Published only as an abstract (from a presentation at the Society for Epidemiologic Research), shortly after the major Winn et al. article.
- Mortality in a large cohort of U.S. military veterans from the 1950s and 1960s, for whom tobacco use status had been determined by survey.
- Very little information to work with.

We can surmise the PBIS: The language is very conclusive, stating the "association between smokeless tobacco and mouth and throat cancer" as a given in the last sentence. However, the only relevant result reported was, "Among chewers who were very light smokers, the cancer risk was eight-fold for the pharynx and 1.5 for the oral cavity." Even if "chewers" refers to all ST use, the choice to report on only "very light smokers" is strongly suggestive of PBIS. If "chewers" excludes snuff dippers (ST use is defined in the first sentence as chewing and snuff use), it suggests further PBIS, choosing the particular exposure with a stronger effect.

The choice of very light smokers was presumably driven by the reported zero OC deaths among nonsmokers. The language is ambiguous as to whether this zero was for all nonsmokers or just nonsmoking ST users. If it was the latter, it would mean a calculated relative risk of 0.0 among nonsmokers.

There is no further information on the ST-OC association provided, so it is impossible to prove the existence of PBIS, as in the previous example.

However, the reported results strongly are suggestive of PBIS from choosing which subgroup to report. PBIS is such a problem for the literature because in most cases we do not have enough information to be sure it has occurred or correct for it.

Kabat et al. (1994)

- Case-control study, focusing on risk factors other than ST.
- 1560 OC cases (1097 male, 463 female) and 2948 (2075 male, 873 female) noncases. Few ST users.
- Results about ST are reported only in the text

A clear case of PBIS: The only OR reported for snuff is 34.5 (95% CI 8.49-140.10) for nonsmoking females. This is actually infinity (0 exposed controls, 4 exposed cases); the finite result came from replacing the cell count of 0 with 0.5. The PBIS is evident in the failure to report the OR for nonsmoking men, which is 0.0, based on the same number of exposed subjects (0 exposed cases, 4 exposed controls), and is more precise due to the larger number of unexposed men.

The paragraph also reports one OR for chewing tobacco, 2.24 (95% CI 0.69-7.34) for nonsmoking males. Not reported are the ORs of 1.2 for all males and 1.1 for male smokers. Numbers for females are impossible to determine from the information provided.

(The information on ST is presented in one very dense and complicated paragraph in the text which contains at least two typos. But it is possible to extract information from that paragraph that is internally consistent and consistent with other information in the paper, and thus appears to be correctly interpreted.)

In fairness to the authors, it should be noted that they do not mention the results for ST anywhere outside this paragraph, and probably thought they were not meaningful. Nevertheless, their publication in the form presented contributes to the PBIS in this literature.

Other publications

For completeness, the following are all other post-Winn publications that report a positive association between the use of at least one smokeless tobacco product and at least one oral cancer.

Bjelke and Schuman (1982) (published as an abstract only) and Marshall et al. (1992) (few exposed, very unstable result) report in prose a positive association, but with no indication of magnitude or methodology information that could be used to detect PBIS. Maden et al. (1992) report a single OR with no obvious PBIS.

Stockwell and Lyman (1986) break out results into 7 oral sites, which is rather thin for 18 exposed cases, but they report them all, eliminating selective reporting of subgroups as a possible source PBIS (and they all have a positive association, though obviously with wide confidence intervals).

Spitz et al. (1988) broke out the 9 cases who dipped from the 25 who used any ST and reported a statistically significant association which does not exist for ST as a whole. Because they explicitly reported the null result for those who chewed (with or without dipping), this is not egregious PBIS, though the slight protective effect of exclusive chewers, which can be inferred, is not mentioned.

In these latter two cases, advocates trying to bolster claims of association could selectively cite more impressive results, but such bias cannot, in this case, be blamed on the scientific literature itself.

Discussion

PBIS typically exists in a form that is not apparent to readers (Phillips 2003a), such as choosing variables to control for based on which choices produce larger effects. Thus, the reported findings should be seen as a lower bound for how prevalent PBIS is in this particular literature. The fact that several cases were detectable by the reader supports the hypothesis that PBIS is a major problem in the health science literature.

The observation of such substantial PBIS in the literature on this exposure-disease relationship does not prove that an ST-OC association does not exist. However, the one article that is primarily responsible for the belief in an association shows clear exaggeration of its findings (even setting aside how those findings are further exaggerated when cited by others) and a substantial portion of the later articles that might be interpreted as support of the claim of association actually show no such association. This should certainly substantially diminish the strength of anyone's belief that ST is a significant risk factor for OC.

The evidence on the relationship of ST and OC is very often summarized with a statement like "ST increases the risk of OC by a factor of more than 4." But the present analysis makes clear that a more accurate summary is "one study found an increase by a factor of a bit more than 3 (for an unusual exposure pattern), but most subsequent studies have been inconsistent with such a large increase, and provide much stronger support for the claim that there is no elevated risk."

These observations do not prove that PBIS occurs throughout the literature. But the potential impact of PBIS is illustrated by the fact that a widely-accepted claim of an exposure causing a disease seems to be built almost entirely on PBIS. This is a bias that can invalidate entire segments of the literature.

References

- Accortt NA et al. "Chronic Disease Mortality in a Cohort of Smokeless Tobacco Users." *American Journal of Epidemiology* 156(2002): 730-737.
- Bjelke E and Schuman L "Chewing tobacco and use of snuff: relationships to cancer of the pancreas and other sites in two studies." *Paper presented at: Proceedings of the 13th International Congress on Cancer [Abstract]*(September 1982).
- Blot WJ et al. "Smoking and drinking in relation to oral and pharyngeal cancer." *Cancer Research* 48(1988): 3282-7.
- Bouquot, J.E. and R.L. Meckstroth "Oral cancer in a tobacco-chewing US population-no apparent increased incidence or mortality." *Oral And Maxillofacial Pathology* 86(December 1998): 697-706.
- Bundgaard T et al. "Case-control study of squamous cell cancer of the oral cavity in Denmark." *Cancer Causes and Control* 6(1995): 57-67.
- Franco EL et al. "Risk factors for oral cancer in Brazil: a case-control study." *International Journal of Cancer* 43(1989): 992-1000.
- Kabat GC et al. "The role of tobacco, alcohol use, and body mass index in oral and pharyngeal cancer." *Int. J. Epidemiol.* 23(1994): 1137-43.
- Lee JL et al. "Predicting cancer development in oral leukoplakia: ten years of transitional research." *Clinical Cancer Research* 6(2000): 1702-10.
- Lewin F et al. "Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. A population-based case-referent study in Sweden." *Cancer* 82(1998): 1367-275.
- Maden C et al. "Human papillomavirus, herpes simplex virus, and the risk of oral cancer." *American Journal of Epidemiology* 135(1992): 1093-102.
- Marshall JR et al. "Smoking, alcohol, dentition and diet in the epidemiology of oral cancer." *Oral Oncology, European Journal of Cancer* 28(1992): 9-15.
- Mashberg A et al. "Tobacco smoking, alcohol drinking and cancer of the oral cavity and oropharynx among U.S. veterans." *Cancer* 72(1993): 1369-75.

Muscat JE et al. "Gender differences in smoking and risk for oral cancer." *Cancer Research* 56(1996): 5192-7.

Phillips,C.V. "Publication Bias in situ." *American Journal of Epidemiology* 157(June 2003a): S59. (manuscript under review, available from www.epi.umd.edu/papers)

Phillips C.V. "Quantifying And Reporting Uncertainty From Systematic Errors." *Epidemiology* 14(March 2003b): 459-66.

Phillips,C.V. and K.J.Goodman "The Messed Lessons of Sir Austin Bradford Hill." 2001 (under review, manuscript available at www.epi.umd.edu/papers).

Phillips,C.V., Guenzel B, and Wang C "Misinformation about the risks of smokeless tobacco relative to smoking on the world wide web." Unpublished, 4 A.D.

Phillips,C.V. and LaPole L.M. "Quantifying errors without random sampling." *BMC Medical Research Methodology* 3(June 2003): 1-10.

Schildt EB et al. "Oral snuff, smoking habits and alcohol consumption in relation to oral cancer in a Swedish case-control study." *International Journal of Cancer* 77(1998): 341-6.

Schwartz SM et al. "Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection." *Journal of the National Cancer Institute* 90(1998): 1626-36.

Spitz MR et al. "Squamos cell carcinoma of the upper upper aerodigestive tract." *Cancer* 61(1988): 203-8.

Sterling TD, Rosenbaum WL, and Weinkam JJ " Analysis of the relationship between smokeless tobacco and cancer based on data from the national mortality followback survey." *Journal of Clinical Epidemiology* 45(1992): 223-31.

Stockwell HG and Lyman GH "Impact of smoking and smokeless tobacco on the risk of cancer of the head and neck." *Head and Neck Surgery* 9(1986): 104-9.

Winn DM et al. "Snuff dipping and oral cancer among women in the southern United States." *New England Journal of Medicine* 304(1981): 745-9.

Winn DM et al. "Chewing tobacco and use of snuff in relation to cause of death in a large prospective cohort [Abstract]." *American Journal of Epidemiology* 116(1982): 567.

Wynder EL et al. "Oral cancer and mouthwash use." *Journal of the National Cancer Institute* 70(1983): 255-60.

Young TB, Ford CN, and Brandenburg JH "An epidemiologic study of oral cancer in a statewide network." *American Journal of Otolaryngology* 7(1986): 200-8.

Zahm SH, Heineman EF, and Vaught JB "Soft tissue sarcoma and tobacco use: Data from a prospective cohort study of United States Veterans ." *Cancer Causes and Control* 3(1992): 371-6.

Brian Guenzel provided research assistance. This research was supported by the non-profit Center for Philosophy, Health, and Policy Sciences, Inc. CPHPS is the recipient of an unrestricted gift from U.S. Smokeless Tobacco Company for support of research of Dr. Carl V. Phillips. This research was investigator-initiated. USST did not influence the content or see this material before it was publically released.

Note: This PDF is modified slightly from the poster as it appeared. One omitted reference has been added and one typo corrected.