

An example of "publication bias *in situ*": Alguacil and Silverman's (2004) study of smokeless tobacco and pancreatic cancer

Carl V. Phillips

University of Alberta School of Public Health
carl.v.phillips@ualberta.ca

May 2006

Keywords: Publication bias, Smokeless tobacco, Pancreatic cancer,

The author thanks Paul Bergen and Brian Guenzel for research assistance. This research was partially supported by an unrestricted grant from U.S. Smokeless Tobacco Company to the University of Alberta for the support of the research of Dr. Phillips. USSTC exercised no influence over the choice or conduct of this research and will only become aware of it upon its public release.

Abstract

When non-representative results from a study are emphasized in the abstract and conclusions, and the details of statistical analysis methods are chosen in order to increase associations, a biased literature results. Several forms of such biased analysis and reporting that seem to characterize many papers in the literature have been labeled "publication bias *in situ*". This paper presents an example of such bias in an article by Alguacil and Silverman (*Cancer Epidemiology, Biomarkers & Prevention*, 2004), which purported to demonstrate that the use of smokeless tobacco increases the risk of pancreatic cancer. That result has been and likely will continue to be influential in debates about whether smokeless tobacco is a viable reduced-harm alternative to cigarettes. But, as demonstrated in the present analysis, the association is entirely due to how the authors chose to analyze and report their data. While the crude association is actually negative (which the authors never mention), an unspecified adjustment for covariates makes it positive (other models would not). One dosage subgroup is emphasized because it has a large and statistically significant association, but the strong protective association for the complementary subgroup is never mentioned. This analysis demonstrates that a claim of a strong and convincing result can easily be extracted from data that actually shows very little or no association, and that the claim is likely to be taken at face value. This is probably a fairly common phenomenon, and it calls into question large parts of the health science literature.

Background

The goal of statistical analysis of data is generally to estimate, as accurately as possible, the relationships among the worldly phenomena that the data represent. Choosing the details of how to analyze a dataset and report the results in order to make an association appear particularly large does not fulfill that goal. Nevertheless, doing so is tempting -- stronger associations are more likely to be published and cited, and may support the authors' political agenda -- and very easy, a combination that probably makes it disturbingly common. Indeed, it may be so common in the health science literature that it substantially biases results away from the null and creates doubts about the quality of our knowledge. Thus, the practice should be challenged when observed.

There are several biased methods for analyzing data and reporting the results. I previously labeled a collection of these "publication bias *in situ*" (PBIS) to emphasize that the implications are similar to those of traditional publication bias (wherein studies with null or otherwise "uninteresting" results are never published), but unlike traditional publication bias, where the bias exists only at the level of the literature as a whole, the bias exists within the individual publication [1]. One fairly well understood form of PBIS is quite similar to traditional publication bias: Studies collect data on many variables, but only the "interesting" associations are emphasized in the publications while many (perhaps thousands) of other associations are calculated but judged uninteresting (or contrary to the analysts' biases) and downplayed or never published. Some attention has been devoted to a second form of PBIS: Authors make biased choices about reporting population subgroups, emphasizing (through highlighting in the title or abstract, or not reporting some of the subgroups) the more "interesting" results [1-5]. The most

subtle form of PBIS is quite possibly worst: Choices about functional form of the model (e.g., how to categorize variables, which covariates to control for, etc.) are made in ways that bias results in a preferred direction [1].

While the reporting aspect of PBIS may be easy to detect (e.g., it is quite common to see a study that has a variety of results across models or subgroups, but an outlier among them is the only result mentioned in the abstract and conclusions), PBIS in the statistical analysis is usually difficult or impossible to detect. It is seldom obvious that choices made by researchers about functional form, subgroups, etc. were based on teasing a preferred result from the data, rather than unbiased motives or arbitrariness. But when PBIS is apparent, it should not be allowed to pass without comment. This is the case in the 2004 paper by Alguacil and Silverman (A&S), "Smokeless and other noncigarette tobacco use and pancreatic cancer: A case-control study based on direct interviews," published in *Cancer Epidemiology, Biomarkers & Prevention* [6]. (Note: the present paper was originally submitted to *CEBP*, but was rejected based on a one-paragraph review, with the phrase "we will not entertain any further correspondence on this paper". While confessing to a rejection is probably not in an author's best interest, I believe that scientific debates should play out in the same pages and wanted to explain why this one did not. Moreover, explicit references to some reviewer comments are quite useful in this analysis, as presented below.)

A&S claimed to show that smokeless tobacco (ST) use is associated with pancreatic cancer (PC). Such a claim is significant because ST is the subject of intense debate: Because it is immensely less harmful than smoking and could be a substitute for some smokers, some of us advocate

promoting the substitution, a strategy of "harm reduction" (see, e.g., [7-12]). But every publication that claims to find any ST-disease association is used as a bludgeon against this suggestion (regardless of the actual absolute risk estimate, the quality of the information, the greater context of other studies, or anything else). A few weeks after A&S was published, several presenters at the Society for Research on Nicotine and Tobacco meetings who were mentioning ST for other reasons took the time to interject comments along the lines of "and now it has been shown that it causes pancreatic cancer too". Scientific articles [13], [14], [15] have cited A&S as showing an association and anti-ST advocacy publications have picked up on the claim. Even though another study [15] (which has its own glaring issues [16-19]) seems to have displaced A&S as the source for quasi-scientific advocacy claims of a ST-PC association, it seems likely that A&S will be used in anti-harm-reduction arguments. Thus, it is worth showing that A&S's data do not actually support their conclusion, which is the result of overt (and probably also hidden) PBIS.

Analysis

A&S reported case-control data for PC in a population that excluded cigarette smokers. Contingency table cell counts (from Tables 1 and 2 of their article) appear in Table 1. The obvious first limitation is that any analysis of ST users will be based on at most 7 exposed cases, though we should not fault researchers for contributing what information they can -- only for overplaying their limited results. Nor should we fault the authors for reporting in the abstract that, "Subjects who used smokeless tobacco regularly had a 40% increased risk of pancreatic cancer (95% CI: 0.5-3.6) compared with nonusers of tobacco" [6](p.55). Though it is conventional to describe this as "no association was observed" (because the confidence interval

includes the null), this conventional description is false: an association was observed. However, the phrasing is interesting: an odds ratio (OR) of 2.4 is seldom reported as a 140% increase (presumably "40%" is intended to sound more impressive than 1.4) and "regularly" refers to someone who ever used a certain quantity weekly for 6 straight months. More important is the PBIS, within the abstract itself, represented by this choice to report a non-statistically-significant result: The phrasing unapologetically changes for the OR of 0.6 for pipe smokers, which is described as "no increased risk" [6](p.55) rather than a 40% decrease.

Table 1 - Contingency table data from Aguacil and Silverman, 2004

Exposure	cases	noncases
No tobacco	123	682
Cigar use (ever)	16	85
Only cigars	9	37
Pipe use (ever)	9	62
Only pipe	1	24
Smokeless tobacco use (ever)	7	44
Only smokeless tobacco	5	25
≤2.5 ounces/week of ST	1	22
>2.5 ounces/week of ST	6	22

Data included only non-users of cigarettes.

"Only" categories refer to individuals who used no other product on this list.

The reporting becomes more curious when we notice that the crude OR for the ever ST users (compared to non-users of tobacco) is 0.88 but this is never reported; the reported OR is the 1.4, which is adjusted by race, gender, geographic site (cases came from three different registries), cigar smoking (highly correlated with pipe smoking), and age. It is not shocking to see

adjustment for covariates to change a relative risk change by more than 50%, and even change from a negative (protective) to positive association. For example, age can easily have such an effect (though the controls were selected to have a similar age distribution to the cases). What is odd and suspicious that the authors did not mention the crude OR or explain which covariate(s) accounted for the large adjustment. The careful reader is left especially curious because the "only ST" exposure had both a crude and adjusted OR of 1.1; thus the entire reported increased risk is driven by the effects of covariate adjustment, which itself results entirely from adding 2 exposed cases and 19 exposed noncases. We should be rather hesitant about drawing any conclusions from such unstable statistics, but the authors of the original paper leave most readers unaware of the instability of the data and that the result is entirely due to the chosen covariate adjustments.

Some readers might reply, "aren't we supposed to control for covariates in observational studies, and report the adjusted result?" (or as the previous reviewer of the present analysis put it, "the author appears to prefer the reporting of crude results as opposed to adjusted results"). We should, of course, control for confounders, but controlling for variables that are not confounders is as likely to move the estimated value away from the true value as toward it (and for certain classes of non-confounders is almost guaranteed to do so) [20]. Thus, a blind presumption that an adjusted OR is a more accurate measure than the crude OR is not justified. Unfortunately, a common mistake in epidemiology is to throw in whatever covariates you happen to have, and declare that confounding has been "controlled for". Worse still is the PBIS from picking and choosing which covariates to use, especially when the number of covariates is large compared to the effective sample size.

The crude OR is not "preferred" to an optimally adjusted OR, of course, but it may well be a better estimate of the true effect than a *particular* adjusted OR, and is an honest representation of the data. Even when the adjusted OR is a better estimate (which is usually the case, except for the most glaring mal-adjustments), the failure to mention what a huge effect the covariate adjustment had, let alone to check if the result was robust or even mention its source is suspicious. (It would be easy to verify the extent to which such suspicion is justified using the original data. Shortly after the original publication, my colleague sought the data from the authors, who said they would only share the data in response to a Freedom of Information Act request. We filed a request, which was acknowledged, but apparently never acted upon. Fortunately, the published information alone is enough to demonstrate the PBIS.)

The authors never explain what functional form they used for the covariates in their logistic regression; gender is obvious and cigar smoking seems to be dichotomous, but there are several choices for age and race. As with most published health reports, the reader has no way of knowing the full methodology and can only speculate about whether the choice of functional forms created PBIS. (Discussion of whether a black-box analysis of secret data constitutes genuine scientific publishing is left for another forum.) Even if the functional forms were reported, they would probably not appear any better or worse than countless other choices that could have been made. Only if authors report results for alternative functional forms (perhaps those used by previous studies on the subject [1]), which is almost never done, or if someone reanalyzes the data can we know if the functional forms chosen produce an particularly large result.

All doubt about PBIS is removed by the high-use subgroup analysis. A&S report in the abstract, "participants who used >2.5 oz of smokeless tobacco a week had an OR of 3.5 (95% CI: 1.1-11)." [6](p.55) There is again the curious issue of the crude OR, only 1.5, but the fact that covariates more than doubled the OR, let alone which covariates had that effect is never mentioned. Moreover, the cutpoint of 2.5 (which is nowhere explained or justified) seems to have been chosen to get a large OR and squeeze out statistical significance. It is apparent that only a cutpoint that included 6 or 7 of the exposed cases in the high-use category would be statistically significant (but including all 7 would make the gerrymandering too obvious, not to mention that further elevating the OR, already greater than that normally attributed to smoking, might make the result even less plausible).

Most important, A&S never mention in the prose of the article (let alone the abstract, which is will be read far more times than the tables) that low exposure (≤ 2.5 ounces/week) showed a protective association that roughly balances the reported positive association for high use, either the reported corrected OR of 0.3 (95% CI: 0.04-2.5) or the crude OR of 0.25. This balance is, of course, inevitable given that the overall association in the data is close to the null.

If the A&S analysis is taken as evidence that using >2.5 oz/week of ST more than triples risk for PC then it provides a similar level of evidence that lower use is protective by about the same ratio. Some might argue that the 3.5 is statistically significant and thus should be believed, while the 0.3 is not and thus should be dismissed, but this reflects a naive misunderstanding of statistics. Even setting aside the fact that a small study has a much higher power to "detect" (get

a statistically significant result) an increased risk compared to a decrease of the same magnitude (and sufficiently small studies or subgroups can yield *only* positive significant results, never negative), any student of modern epidemiology knows that lack of statistical significance should not be interpreted as a lack of association. A&S apparently realize this (recall that they reported the non-significant OR of 1.4 as an increase), but choose to ignore a non-significant result when it is inconvenient. Their result of 0.3, with the large majority of the confidence interval well below 1.0 means that either their analysis strongly suggests a protective effect for low-level ST use, or that the implausibility of such a protective effect suggests that we should also not believe the 3.5 that came from the same stratification of a near-null association.

In the reporting of the 0.3 and 3.5, A&S arrange them in two rows of a table, without juxtaposing the unexposed reference group, creating the illusion of a dose-response trend (in fairness, this may just be an innocent attempt to save space in a table, but it is still misleading). If the referent were there, the non-monotonicity of (1.0, 0.3, 3.5) would make the implausible dose-response apparent, even to readers who believe the high-dose OR estimate. Instead of making that clear, A&S present a very misleading p-value of 0.04 for the test for trend. This statistic for the non-trend is clearly driven entirely by the high-dose group (after all, a statistical analysis for the "trend" between two points simple repeats the statistical analysis of the binary comparison of the two, and we already know the 3.5 is statistically significant compared to the referent). The low-use group effectively disappears for two reasons: (1) The confidence interval resulting from the cell count of 1 is so wide as to not have much influence on the trend line. (2) By treating every subject in a stratum as having the average exposure level for that stratum (as A&S do), the low-use category likely has an exposure level quite close to zero, so in effect the middle category just

pulls down the origin of the trend, leaving it nearly linear, rather than pulling it non-monotonic in the middle. (Unfortunately, due to the lack of detail about the huge effects of covariates, it is impossible to simulate the A&S result to formally demonstrate these points.)

Apart from that last more technical point, realizing there is no real trend is trivial; nevertheless, naive readers who just look at the (0.3, 3.5) trend and the p-value could actually believe there is a dose-response trend, and A&S seem to encourage this error. Evidence that some readers will misinterpret comes from the reviewer of the previous version of this: "the author fails to mention the significant trend test associated with amount of tobacco use, which could be considered a type of publication bias in situ in his own commentary" (it is true that the previous version did not mention the test statistic, attending instead to the actual effect estimates, which should always be preferred). I appreciate the concern about PBIS (though not the labeling of a careful quantitative analysis as a "commentary" [21]), but I especially appreciate the rare documentary evidence of how an ostensibly careful reader was confused by the misleading presentation of the study results.

In summary, it is likely that many readers and advocates will interpret A&S's results as showing a strong association between ST use and PC, an interpretation that is clearly encouraged by the way the results are presented. The consequences of this are potentially quite important, since any result that is mistakenly interpreted as showing a high risk from ST will tend to undermine efforts to promote ST as a harm-reducing alternative to smoking and to try to convince ST users to not switch to the much more dangerous use of cigarettes [22].

A&S's results provide only equivocal support for a small increased risk, and even that is based on one particular way of analyzing of the data. Though we were unable to reanalyze the data, it is clear that many alternative functional forms for the multivariate model would show null and protective associations (this conclusion follows immediate from the fact that the crude OR shows the latter). Given only 7 exposed cases (before stratifying) and a model that used at least 6 covariates, the estimated relative risk is obviously very sensitive to the exact choice of functional form. Alternative choices, which would be completely invisible to the reader given the lack of detail in the reporting, could probably generate most any result an author wanted. Given the pains they took to report positive associations in their results, it is easy to surmise what result A&S wanted.

Discussion

A&S may have done things in their analysis that are outside the bounds of accepted practice in epidemiology, but there is no compelling evidence that they exceeded those bounds. More's the pity for the accepted practice in epidemiology. Since among the unfortunate accepted practices is publishing methods sections that do not actually explain your methods [23], this, along with many of the observations above, must be phrased speculatively. Readers of scientific papers should not have to speculate about what the researchers did and what the results really mean.

In the spirit of challenging manipulation of data, it is worth a paragraph to explicitly challenging political manipulation of scientific credibility. Some anti-ST advocates will likely make the ironic argument that the present analysis of bias must be biased because my research is partially support by the smokeless tobacco industry (ignoring that this funding is the most unrestricted and

hands-off funding that any researcher could ever hope to get). Such *ad hominem* attacks and appeals to "truthiness" should seem particularly pathetic to sophisticated readers in the present case: After all, in contrast to A&S's black-box analysis, everything presented here is completely transparent, so any reader who understands the science can verify everything. Moreover, it will be clear from my writings that my goals are to improve the honesty of epidemiologic research [1;24-26] and to analyze the potential of ST as an alternative to smoking. The former of these obviously calls for great care, but so does the latter. Accurately measuring the risks from ST is of great interest to those of us who recognize the tradeoff among negatives and positives; those who condemn ST with an unscientific fervor will not be interested in quantifying the tradeoffs. The U.S. National Cancer Institute (NCI), which employed A&S and, along with other major U.S. organizations, pushes a fervent anti-ST and anti-harm-reduction message that ignores the science [22]. Naturally, I am not asking readers to "follow the money" and disbelieve the A&S conclusions because their salaries were paid by a biased organization -- the reasons for disbelieving them appear above -- but merely to note this in the context of the inevitable *ad hominem* attacks that will follow. There is bias everywhere, and one of the best ways to deal with it is critical reanalysis.

The A&S analysis has several problems beyond PBIS. ST use (like smoking) is associated with exposures that increase risks of many diseases, including lower socioeconomic/educational status, less access to medical care, greater alcohol and other drug use, and poorer diet and body mass index (the latter pair are believed to be important risk factors for PC). Several of these exposures are very difficult to control for, so residual confounding will always be a problem. A&S report running models that controlled (presumably quite roughly) for some of the above

factors, but left them out because they "did not substantially modify any of the risk estimates" [6](p.56). But their data on obesity and total calories probably did not measure effects of diet very precisely (though it still should have been left in the model as a likely confounder). It is also further evidence that PBIS from choice of functional form is likely: the reported result was clearly not based on an *a priori* model (why would someone assume catchment site is a confounder, but not income or obesity) but rather on browsing multiple options.

The noted asymmetric reporting in the abstract, treating positive associations as real and negative associations as null, also appears in the literature review. A&S describe two studies [27;28] as reporting increased risk while another [29] showed "[n]o association", but actually the latter showed a negative association. (None of the three were statistically significant.) Almost all readers will have a strong prior that ST use is not protective against any cancer, but this should affect our ultimate worldly conclusion, not our interpretation of individual study results. If negative results are dismissed as uninformative, our ultimate conclusions are preordained: The positive results that some studies will inevitably get -- even if due to chance alone -- will lead to the conclusion that there is a true positive association. It is bad to have such a strong prior that the evidence could not possibly convince us there is a protective effect (it is always *possible*); it is much worse to use a "methodology" that almost guarantees that the evidence will be interpreted as showing a positive association. Yet the analysis in A&S (and a substantial portion of other epidemiology) implicitly uses just such a methodology.

Detailed reanalysis is rather involved, and thus is only possible to do for a tiny portion of the literature. In the present case, at least the most obvious and damning observations should be

immediately apparent to educated readers who stop to think (though our scientific education might be failing, and few readers stop to think). Of course, it should not be necessary for other authors to deconstruct published results like this; we should have an ethos in health science that would require more complete and honest reporting of results in the first place.

If ST use causes a measurable increase in PC, we certainly want to know it. If it does not cause PC at any detectable level, we want to know that too. But study reports that analyze data and describe results in ways that make the former conclusion very likely, even if the latter scenario is actually true, clearly do nothing to inform us.

Conflict of Interest Statement

The author is interested in promoting the use of smokeless tobacco as a highly-reduced-risk alternative to smoking, and in accurately measuring what risks might result from ST use. He is thus motivated to criticize analyses that appear to mislead about those risks. He has no interests that conflict with these goals. The author's research is partially supported by an unrestricted research grant from U.S. Smokeless Tobacco Company to the University of Alberta for the support of the research of Dr. Phillips and colleagues. The author has received consulting income from USSTC in connection with litigation. USSTC exercised no influence over the choice or conduct of this research and will only become aware of it upon its public release.

Reference List

1. Phillips CV. Publication bias in situ. *BMC Med Res Methodol* 2004;4:20.
2. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-3336.
3. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-359.
4. Kahn J. How a Drug Becomes "Ethnic": Law, Commerce, and the Production of Racial Categories in Medicine. *Yale Journal of Health Policy, Law, and Ethics* 2004;4:1-46.
5. Phillips CV. Publication Bias in situ. *American Journal of Epidemiology* 2003;157:S59.
6. Alguacil J, Silverman DT. Smokeless and other noncigarette tobacco use and pancreatic cancer: a case-control study based on direct interviews. *Cancer Epidemiol Biomarkers Prev* 2004;13:55-58.
7. Rodu B. *For Smokers Only: How smokeless tobacco can save your life*. Sulzburger & Graham Pub Co, 1995.
8. Tobacco Advisory Group of the Royal College of Physicians. *Protecting smokers, saving lives*. London: 2002.
9. Bates C, Fagerstrom K, Jarvis M, Krunze M, McNeill A, Ramstrom L. European Union policy on smokeless tobacco: A statement in favour of evidence-based regulation for public health. *Action on Smoking and Health* 2003;European Policy on smokeless tobacco.
10. Foulds J, Ramstrom L, Burke M, Fagerstrom K. Effect of smokeless tobacco (snus) on smoking and public health in Sweden. *Tob Control* 2003;12:349-359.
11. Rodu B, Stegmayr B, Nasic S, Asplund K. Impact of smokeless tobacco use on smoking in northern Sweden. *J Intern Med* 2002;252:398-404.
12. Rodu B, Cole P. The burden of mortality from smoking: comparing Sweden with other countries in the European Union. *Eur J Epidemiol* 2004;19:129-131.
13. Accortt NA, Waterbor JW, Beall C, Howard G. Cancer Incidence among a Cohort of Smokeless Tobacco Users (United States). *Cancer Causes Control* 2005;16:1107-1115.
14. Hatsukami DK, Lemmonds C, Zhang Y, Murphy SE, Le C, Carmella SG, Hecht SS. Evaluation of carcinogen exposure in people who used "reduced exposure" tobacco products. *J Natl Cancer Inst* 2004;96:844-852.
15. Boffetta P, Aagnes B, Weiderpass E, Andersen A. Smokeless tobacco use and risk of cancer of the pancreas and other organs. *Int J Cancer* 2005;114:992-995.
16. Ramstrom L. Re: "Smokeless tobacco use and risk of cancer of the pancreas and other organs" by Boffetta et al. *Int J Cancer* 2005.
17. Nilsson R. Possible carcinogenicity of smokeless tobacco. *Int J Cancer* 2005.
18. Rutqvist LE, Lewin F. Flawed methods. *Int J Cancer* 2005.

19. Rodu B, Cole P. A deficient study of smokeless tobacco use and cancer. *Int J Cancer* 2005.
20. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37-48.
21. Phillips CV, Goodman KJ, Poole C. Lead editorial: The need for greater perspective and innovation in epidemiology. *Epidemiol Perspect Innov* 2004;1:1.
22. Phillips CV, Wang C, Guenzel B. You might as well smoke; the misleading and harmful public message about smokeless tobacco. *BMC Public Health* 2005;5:31.
23. Phillips CV. Introducing article-processing charges and inviting "detailed methods sections" articles. *Epidemiol Perspect Innov* 2005;2:5.
24. Phillips CV, Maldonado G. Using Monte Carlo methods to quantify the multiple sources of error in studies. *American Journal of Epidemiology* 1999;149:S17.
25. Phillips CV. Quantifying And Reporting Uncertainty From Systematic Errors. *Epidemiology* 2003;14:459-466.
26. Phillips CV, LaPole L.M. Quantifying errors without random sampling. *BMC Medical Research Methodology* 2003;3:1-10.
27. Zheng W, McLaughlin JK, Gridley G, Bjelke E, Schuman LM, Silverman DT, Wacholder S, Co-Chien HT, Blot WJ, Fraumeni JF, Jr. A cohort study of smoking, alcohol consumption, and dietary factors for pancreatic cancer (United States). *Cancer Causes Control* 1993;4:477-482.
28. Heuch I, Kvale G, Jacobsen BK, Bjelke E. Use of alcohol, tobacco and coffee, and risk of pancreatic cancer. *Br J Cancer* 1983;48:637-643.
29. Farrow DC, Davis S. Risk of pancreatic cancer in relation to medical history and the use of tobacco, alcohol and coffee. *Int J Cancer* 1990;45:816-820.