

Author's response to reviews

Title: Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: An application of generalized additive models to case-control data

Authors:

Veronica Vieira (vmv@bu.edu)

Thomas Webster (twebster@bu.edu)

Janice Weinberg (janicew@bu.edu)

Ann Aschengrau (aaschen@bu.edu)

David Ozonoff (dozonoff@bu.edu)

Version: 3 **Date:** 7 June 2005

Author's response to reviews: see over

Title : Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data

Authors: Veronica Vieira, Thomas Webster, Janice Weinberg, Ann Aschengrau and David Ozonoff

Dear Dr. Webster,

Thank you for submitting your revision of the above manuscript. We are very pleased with the changes made, and only some minor issues seem to remain before we can proceed toward formal acceptance of your paper.

Your manuscript was reviewed by one of the reviewers, who had asked to see the revision. His comments are copied below. We would like you to respond in the form of a revised manuscript, together with a point by point response.

When reading your current revision, it was not quite clear to me how you had dealt with Dr. French's comment in regard to the selection of controls. Your response indicates that you decided to refer to the previous publication by Aschengrau et al., but I was not sure whether this change was actually made. Please clarify.

Response: In the earlier manuscript we stated "See Aschengrau et al. [6, 11] for a detailed description of the methods used to define the study population." We altered this in response to Dr. French's comment to read "See earlier papers [10-12] for a detailed description of the methods used to define the study population, including the rationale for the method of control selection."

There is a small technical problem with Figure 1. For publication purposes, it needs to be at least 300 dpi. Please reupload so that it is legible at a width of 600 pixels. As always, your revised manuscript should be in strict accordance with our instructions for authors, cf. the pre-acceptance checklist at <http://www.biomedcentral.com/info/edgr-preacceptcheck.asp>.

Response: The figure has been reformatted.

When you have revised your manuscript in the light of the above comments and made the required changes to the format of the paper, please re-upload the revised files, and make any necessary changes to your manuscript submission forms on the Environmental Health web site, by clicking 'My Environmental Health', logging in with your user name and password and then clicking the manuscript title to go through to your submission screen. You should provide your point-by-point response to the reviewer comments, along with a summary of the changes you have made, by uploading them at the end of the submission process, using the 'Cover letter' box.

We would expect you to return the revised paper in the very near future, preferably within two weeks, but if you imagine it will take you longer to prepare the revision, please give us some estimate of when we can expect to hear from you.

Please don't hesitate to contact us if you have any problems or questions regarding your manuscript.

With best wishes,

Philippe Grandjean

environmental@health.sdu.dk

<<http://www.ehjournal.net/>>www.ehjournal.net

Olaf Berke:

· Page 8 last but one sentence: I disagree with the statement in this sentence and cannot see any indication for using the given reference to underpin the statement: "Unfortunately, the density ratio approach provides no easy method to adjust for covariates [23]." Why can logistic regression not be adjusted for covariates? Diggle is reviewing the methodology and gives examples in [23].

Response: The density ratio approach does not employ logistic regression. Instead, it separately estimates the density of the cases and the density of the controls, and then takes their ratio. Although the difficulty of adjusting for covariates with this approach is mentioned in the cited

reference, it is made more explicitly in Diggle's chapter in the book by Elliot et al 2000. We substituted the latter reference to make this clearer.

· One of the two references 30 or 33 should be deleted - there is no need to cite SPLUS twice!

Response: We disagree. The first reference refers to the software; the second refers to a discussion in a manual regarding a specific technical issue.

· Page, middle of 2nd paragraph: "We estimate...using ...SPLUS function predict.gam. As this function...can produce discrepancies..." => is the SPLUS implementation of the method the cause or the method? It was my objective in the first review to point out my position that you should distinguish between the "general use of methodology" and your "specific use of software products". (I understand that you are willing to share your SPLUS code, but scientific literature is to my opinion not a medium to advertise commercial products. The analysis can equally well be performed in R, which is freeware.) I believe, your critic in this specific case is about the general method and not on SPLUS functions and this should be made clear to the reader!

Response: As stated in the manuscript, we describe a method for doing these analyses with existing software (S-Plus) widely used by epidemiologists. In this part of the manuscript, we state that the current S-Plus implementation of *predict.gam* defines neighborhoods in a manner different from what many users might assume and/or desire. We are not attempting to "advertise" S-Plus. Several of us use R and are very supportive of freeware. To the best of our knowledge, at the time we did this work, R did not yet have all of the capabilities required to do our analysis. A new R gam package was released in March 2005; we are currently trying it.

· I asked for more information on the "permutation test" and you gave the information on page 11/12. First I think what you tried is better known as a Monte Carlo test (see Kelsall/Diggle or Waller/Gotway or Cressie). Second, I am afraid you were not clear enough in your description of the method: the 999 "permutations" should be Monte Carlo replications under the condition of "complete spatial randomness". If your permutations create disease clusters - the Monte Carlo p-values are incorrect (to your advantage I expect them to be overestimated => your cold and hot spots are too small)! And finally the permutations of "locations" should read permutations of "cases", since you condition on the locations these cannot be moved around.

Response: We edited this part of the manuscript to try to further clarify our procedure. We did a permutation test under the null hypothesis of no association between case/control status and location. However we did not obtain the full permutation distribution (as this would be unwieldy) and instead sampled from this distribution to obtain an approximate permutation p-value. Therefore this test may be considered a type of Monte Carlo test as well. The test is performed by conditioning on the number of cases and controls, the relationship between case/control status and covariates, and then randomly assigning individuals to the fixed locations, i.e., relabeling the residences. Some samples may result in disease clustering and extreme test statistics by chance (i.e. sampling in the tail of the null permutation distribution), however this is to be expected and the p-value is appropriate.

· Is it pure coincidence that the span (also known as bandwidth) parameter has the same values like the AIC for the two models on page 17? Be clearer on the use of map, model and data in connection with AIC and span/bandwidth: we don't get a AIC for a map but for a model fit to data and we generate a map from that model - therefore the AIC does not belong to the map, rather than the map belongs to the specific AIC (and epidemiologist shall be clear on cause and effect!).

Response: We assume you are referring to the fact that both the imputed and non-imputed data sets had AIC curves with local minima at 0.15 and 0.35. We doubt it is “pure coincidence” as the data sets are very similar. Our methodology section makes it clear that we select the span size first and then generate maps.