Reviewer's report

Title: Cumulative solar ultraviolet radiation exposure and basal cell carcinoma in a nationwide US cohort using satellite and ground-based measures

Version: 0 Date: 23 Aug 2019

Reviewer: Denis Hemon

Reviewer's report:

This paper provides a precise quantitative description of the association between exposure to ambient ultra violet radiation (UVR) and the incidence rate of Basal Cell Carcinoma (BCC).

The population studied is a large cohort of 63,912 white cancer-free US radiologic technologists from entry (1983-1998) to exit (2003-2005) with known ultraviolet irradiance at up to 5 residential locations; 2,151 cases of BCC were identified. Using generalized-additive models, the association between the incidence rate of BCC and ambient cumulative ultraviolet radiant exposure is described, using either ground-based National Solar Radiation database Average Daily Total Global data (AVGLO) or satellite-based National Aeronautics and Space Administration Total Ozone Mapping Spectrometer data (NASA-TOMS).

The excess absolute risk (EAR) and the excess relative risk (ERR) of BCC observed in this cohort were modeled using a Poisson regression and a linear, or linear-quadratic or log-linear function of the cumulative exposure to UVR, as estimated from the AVGLO and NASA-TOMS obtained from the residential histories.

The main conclusions drawn from these analyses are the existences of:

- a very marked increase of the EAR and ERR of BCC as a function of cumulative exposure to UVR (estimated by AVGLO or by NASA-TOMS);

- a significant upward curvature of the relationship between AER of BCC and cumulative exposure to UVR (AVGLO and NASA-TOMS) and no such significant curvature of the relationship between cumulative UVR exposure and ERR of BCC;

- a substantial variation in ERR with time after exposure (beginning of ? end of ?) and age at (onset of ?) exposure, whereas no such effect was clearly observed for the AER of BCC?

Overall assessment

This work is based on a large cohort, with good quality observation and a wide range of the exposure of interest. It addresses in detail the interesting question of quantifying the association between UVR and BCC, which was not addressed previously with this level of epidemiological information and analytical effort.
On the other hand, the reader who is firstly very impressed by the clear message given by figure 1 may become perplex when he goes deeper into the reading of this manuscript considering:

- the profusion of models and parameters that are used,
- the pending question of how to interpret results when algorithms do not converge,
- the capacity of this large set of data to allow the estimation of risk as a function of several parameters that are all likely to be tightly correlated together,
- the potential role of other BCC risk factors which are not taken into account in the present analysis.

It is therefore recommended:

- to propose to the authors to provide a revised version that will take into account the following comments,
- to accept this paper for publication in BMC Environmental Health, if these comments are adequately addressed.

Main comments

1) Is the cumulative exposure x EAR of BCC association linear or quadratic?

The authors underline in the abstract, results, conclusion, that the linear-quadratic model better fits the association between the EAR of BCC and cumulative exposure UVR than a simple linear model. However they also give as a summary statistics the slope of the simple linear models in the results, conclusion, and abstract.

As a matter of fact, looking carefully at Figure 1, it does not appear that there is a clear upward curvature of the exposure x risk observations. This may not be contradictory with the p-values of the quadratic component of the model as the number of person-years at risk and of BCC cases are very large: the p-value for including a quadratic term may be very small while the difference between the linear and linear-quadratic models may also be quantitatively small.

This point should be clarified for instance by giving a table with the estimated EAR and ERR of BCC and their 95% confidence intervals for a range of percentiles of the cumulative exposure distribution in the cohort (P10% P25% P50% P75% P90%).

From these figures, the authors - and the readers - will be able to appreciate if the linear and linear-quadratic models do differ materially in their capacity to estimate adequately the risk observed:
- if yes, it is appropriate to underline the existence of a curvature, but inappropriate to give the slope of the simple linear model in the abstract, results and conclusion. In this case the estimated EAR or ERR should be given for a few values of the cumulative exposure to UVR to illustrate both the order of magnitude of the EAR and the curvature of their variation with cumulative UVR.

- if not, it is not appropriate to underline the existence of curvature in the abstract and conclusion, but appropriate to summarize the association between EAR of BCC and cumulative exposure by the slope in a simple linear model.

2) Clarify notation in the text, not only in Table footnotes or supplementary information

The notation used in the text, even though they follow the common notation in the GAM, should be clarified in the text and not only in the Supplementary material. As an example page 6 and 7 when presenting formula (1): \( \lambda, t, \phi, \beta, \alpha, H \) should be defined in the text and not only in Supplementary Table B1.

Time after exposure and age at exposure should be defined as all subjects in the cohort had a lifelong exposure to ambient UVR.

3) Do the AVGLO and NASA-TOMS cumulative exposure metrics describe equally well or not the observations?

Figure 1 might suggest that the EAR of BCC increases more steeply over the range of cumulative exposure to UVR when estimated by AVGLO (from 0 to about 3 / 10,000 PY) than when estimated by NASA-TOMS (from 0 to about 2/10,000 PY, that is 1 third less). On the other hand the values of the AIC criteria given in Table 2 suggests that both estimates of cumulative exposure to UVR provide equally acceptable linear-quadratic fits to the observed data as compared to the simple linear models based on these exposure metrics. The authors concluded:

- in the abstract (page 2) that both measures of cumulative exposure gave similar values for the risk,

- in the conclusion (page 11) that AGVGLO appeared to give a better description of the observations than NASA-TOMS.

Here again a coherent conclusion should be proposed, and for this purpose a tabulation of the estimates of EAR and their 95\%CI over a range of percentiles of the distribution of cumulative exposures (P10\%, P25\%, P50\%, P75\%, P90\%) will help to appreciate if both exposure measurements do or don't fit similarly the data.
A scatter plot of \((x_i = \text{AVGLO}, y_i = \text{NASA-TOMS})\) over the subjects \((i=1,n)\) for which these two metrics are available will also be important to compare the capacity of both metrics to capture the variation of EAR or ERR of BCC.

4) Final conclusion

The final conclusion of the abstract (page 2) and the text (page 11) states that:

« If confirmed in other datasets, our results suggests that interventions aimed at reducing risk of basal cell carcinoma should concentrate on those with the highest levels of ambient UVR exposure. »

While the originality of the work performed relied on the use of cumulative exposure metrics, the large body of observations available and their precise quantitative analysis, this final conclusion could have be drawn by the authors from the literature they cited and which demonstrates, convincingly, that the incidence of BCC do increases with exposure to UVR.

To provide new quantitative information on this question the authors proposed to introduce duration of exposure through the use of cumulative exposure. From this point of view, it would be interesting that the authors put into perspective (in the abstract, text, conclusion) the role of variability of exposure level (variation of ambient UVR between the dwellings of the subjects in the cohort, related to demography and the spatial variability of UVR over continental USA) as compared to the role of the variability of exposure duration (between subjects in the cohort) in the increase in BCC risk.

5) Non-convergence of several models

Several of the models adjusted to the data are given while the algorithm for calculating their parameters estimates did not converge. The authors should help us to disentangle:

- the models where this lack of convergence did not matter because the estimations of the EAR or ERR were "algorithmically" stable even though the estimations of the unknown parameters were not (and therefore these parameters could not be estimated meaningfully, and should not be given in results and tables)

- and where this lack of convergence should lead to ignore this model in the paper.

A discussion of the role of considering several variables, which are tightly correlated together, like age, cumulative exposure, age at first exposure, time after exposure, duration of exposure, should also be given by the authors to distinguish when a model can be meaningfully fitted to the data and when it cannot. From this point of view, analyzing the role the cumulative exposure within different age windows appears to be problematic and this part should be either omitted or discussed in detail.
6) Confounding

Page 19, lines 49 to 56, the authors indicate that "In addition to exposure to ambient UVR exposure, education, income, cigarette smoking, alcohol consumption, body mass index, hours exercise per week, eye color, skin complexion, ever sunburnt, number of blistering sunburns before age 15, skin reaction strong sunlight and number of dental X rays were all significantly associated with BCC risk".

Some of these characteristics may either vary with age or vary from place to place within USA. Some of them may therefore have a potential to be associated with cumulative exposure to UVR and therefore to confound the association between cumulative exposure to UVR and BCC.

This should address in the paper

In the footnote of Table two it is stated:

†All analysis used linear-quadratic model (B1') with adjustment to the baseline BCC rate for baseline questionnaire, ln[age], birth year, [birth year]2, [birth year]3, [birth year]4, [birth year]5.

What is baseline BCC rate? Is it adjusted for the different covariates that were listed as independent risk factors for BCC? If yes this should be clarified, if not this should be investigated.

7) Minor comments

- In the tables EAR and ERR should be given with just one digit after the comma
- Table 1, Lines 10 and 11: the mean (SD) of age should be moved in the good column (no BCC or BCC)

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable
**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests' below.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.