Narodowe Centrum Badań Jądrowych
National Centre for Nuclear Research
ŚWIERK

EX-POLON

On methodological problems in seeking for a correlation of lung cancer with radon level

J. Reszczyńska [1], M. Pylak [1,2], K.W. Fornalski [1,3], S.J. Mortazavi [4], L. Dobrzyński [1]

[1] National Centre for Nuclear Research (NCBJ), ul. A. Sołtana 7, 05-400 Otwock-Świerk, Poland
[2] Institute of Physics, Polish Academy of Sciences (IF PAN), Al. Lotników 32/46, 02-668 Warszawa, Poland
[3] Ex-Polon Laboratory, ul. Podleśna 81a, 05-552 Łazy, Poland
[4] Fox Chase Cancer Center, Philadelphia, PA, USA

Abstract

The relationship between lung cancer morbidity and radon concentration is an open subject of many researches all over the world. Secondary risk factors may limit classical statistical analysis, thus many authors do not support ecological results. To address this problem, the authors of this poster focused on the methodological aspects used for interpretation of such results when a rich collection of various explanatory parameters is available. This is e.g. the case of the county-level dataset collected by Simeonov and Himmelstein in 2015 [1], that contains geographically aggregated data on cancer risk factors, environmental features, demographics, collected for 3 142 U.S. counties and county-equivalents.

Apart from the Least-Square and Bayesian linear regression analysis, for the first time the Maximum Entropy Method (MEM) was used for interpretation of the influence of the correlated concentration of radon, altitude and UVB level on the lung cancer morbidity. An importance of binning the data is also shown. The binning of the data may be critical for final conclusions. For example, the negative slope of the fitted line can change to the positive one by a slight modification of the data binning.

It follows from both, LSQ and Bayesian methods of analysis that a decrease of cancer morbidity with increasing radon concentration is minute in the group of the lowest smoking prevalence, and is statistically sound in the group of high smoking prevalence. This conclusion follows from both, LSQ and Bayesian, methods of analysis. The use of MEM provided much richer picture: a clear trend of decreasing morbidity of lung cancer with increasing radon density level independently of the altitude of the place of residence and the level of UVB. Moreover, this trend does not depend on the prevalence of smoking and the sex.

Results

Data for selected subgroups [1], after reduction of the altitude band to 200-275 m, were initially carefully analyzed using the classical least squares regression (LSQ) method and robust Bayesian regression methodology [2]. The only models taken into account for the dose-effect dependence were a linear dependence and constant - no dependence on radon concentration.

In all cases, except the population with the lowest smoking percentage, a negative slope is observed in both methods (see Fig. 3). One notes that in all 3 subgroups cancer incidence is substantially higher in men than in women. Nevertheless, reducing the data to a certain percentage of smokers, not essentially different for men and women, one can still observe a statistically significant stronger decrease in lung cancer with increasing radon concentration in men (Fig. 4) which indicates that women exhibit lower risks than men.

All those findings were supported by the analysis performed by the Maximum Entropy Method (Fig. 5).

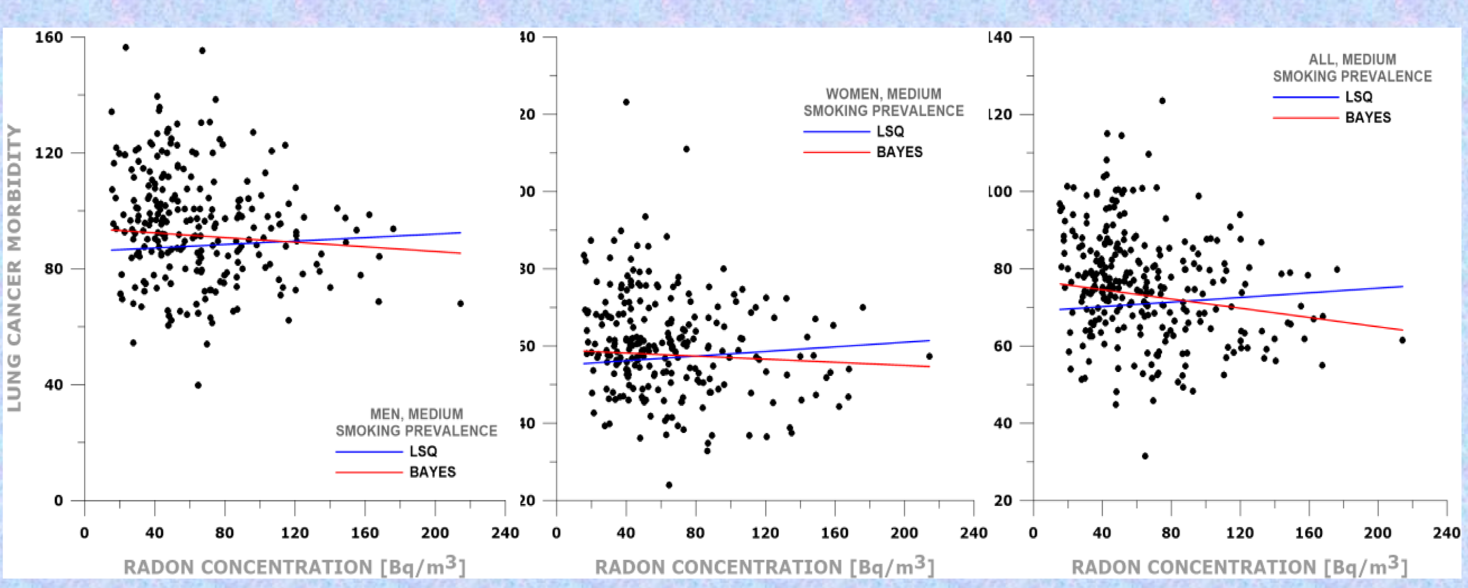

Fig.3 An assumed linear regression between lung cancer morbidity (per 100,000 cases) was performed using a classical and Bayesian regression method for the medium smokers group: men, women and general cohort data.

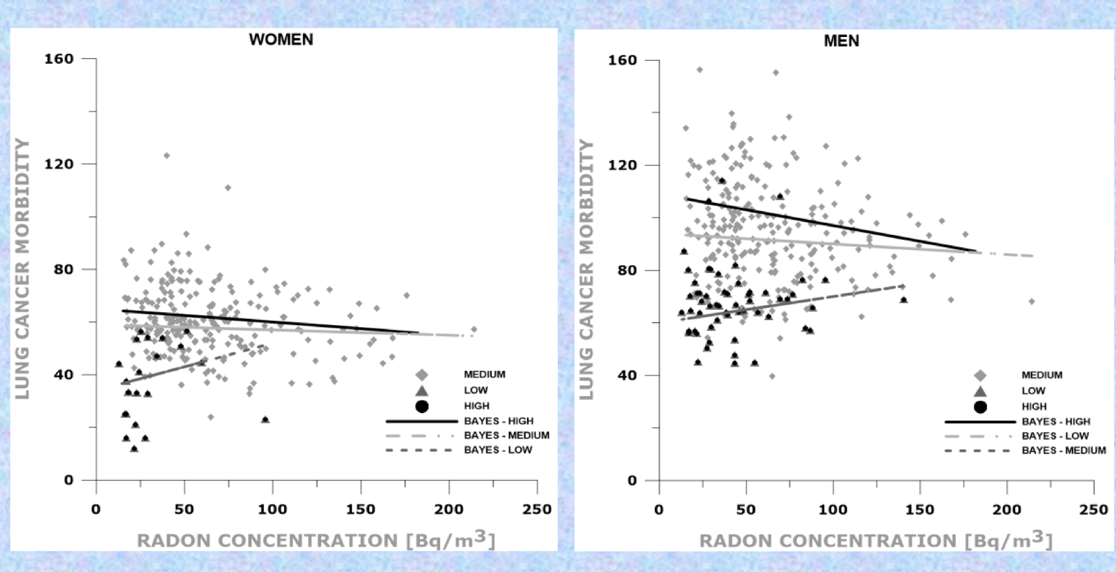

Fig. 4 Lung cancer incidence (per 100 000 cases) and radon concentration (Bq/m^3) for men and women. Solid lines represents best linear fit, for 3 smoking prevalence by the robust Bayesian regression method

Data

The lung cancer appearance is likely connected with primarily strongly correlated confounding factors consisting of radon concentration, altitude inhabited by people, and the level of UVB [1]. As an example Fig.1 shows the correlation between the two first parameters for the case of regions with medium and high smoking prevalence.

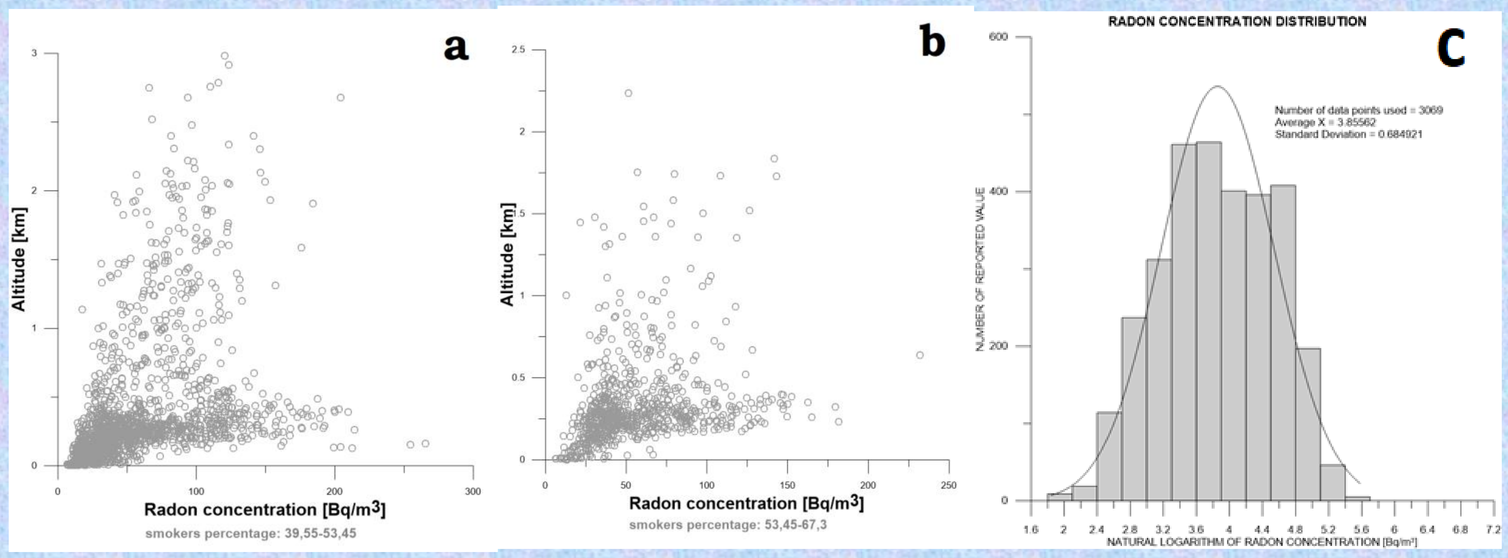

Fig.1 Correlations between the altitude and radon concentration for smoking prevalence of a) 39.55%-53.45% and b) 53.45%-67.30%. [1] c) Log-normal radon concentration distribution. The fit to a Gaussian shows that the median concentration of naturally occurring radon is 47.3 Bq/m^3 within 68% CI (23.8 - 93.7) and 95% CI (12.0 - 185.9)

The presented analysis refers also to the meta-analyses presented in our previous paper [2]. Both collections focus on data presented in approx. 30 separate case-control studies, where all data are re-analysed together. The result turns out to be strictly connected with the model (mathematical curve) which is fitted to the data.

The crucial importance in that situation is a proper binning of the data. Usually, when for a dataset one has some number of similar experimental points, it is possible to merge them into one single point to reduce the uncertainty and make the analysis simpler. Thanks to that, the number of data points for further analysis will be strongly reduced. Such a selection of proper bins is usually made in an arbitrary way which makes further conclusions problematic. Especially, the binning can be the problem of data manipulation (see Fig. 2) where the same data with different binning gives different results.

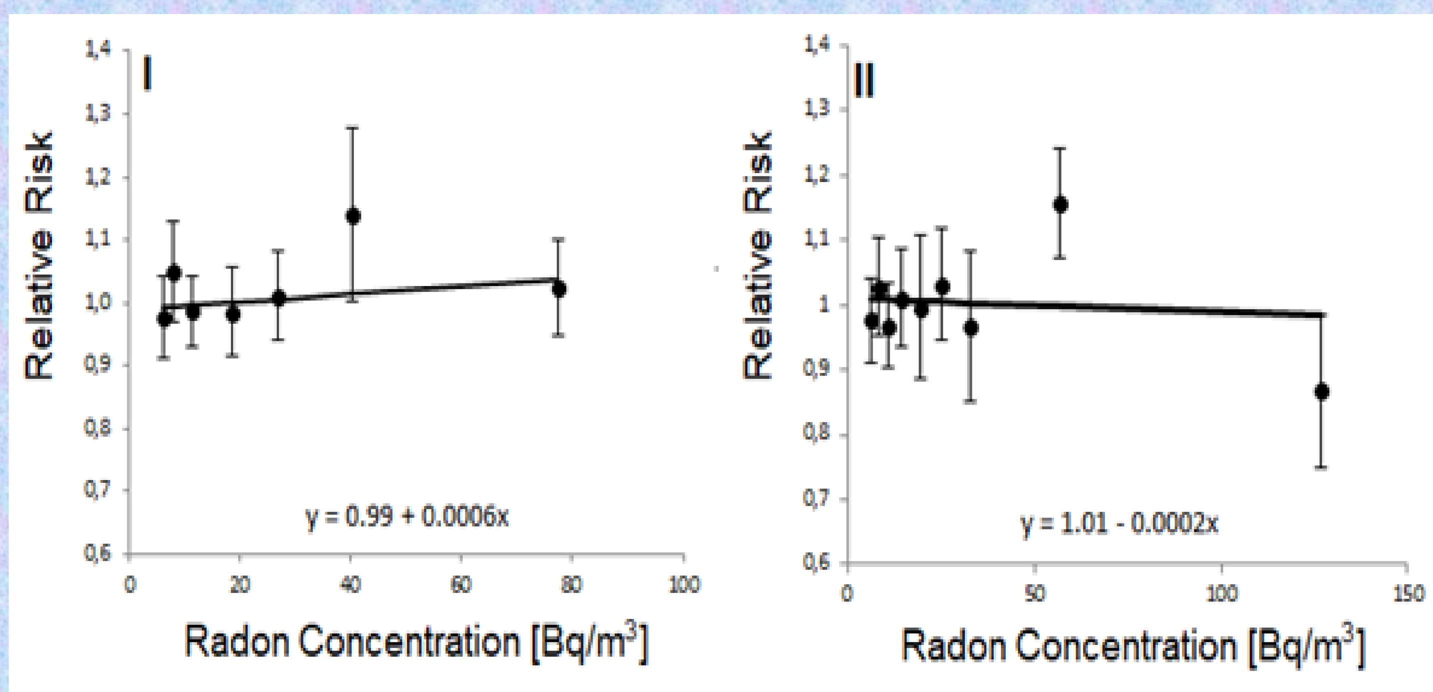

Fig. 2. A re-analysis of the 32 case-control data listed by Dobrzyński et al. [2] after merging groups of experimental points into single ones in bins with arbitrary boundaries: a) bin ranges of 0-37, 37-50, 50-75, 75-125, 125-175, 175-270 and 270-800 Bq/m^3, which results in a pro LNT conclusion; b) bin ranges of 0-37, 37-53.5, 53.5-65, 65-100, 100-124, 124-150.1, 150.1-200, 200-600 and 600-800 Bq/m^3, which results in a pro hormetic conclusion. Please note, that in both cases the same original data were used.

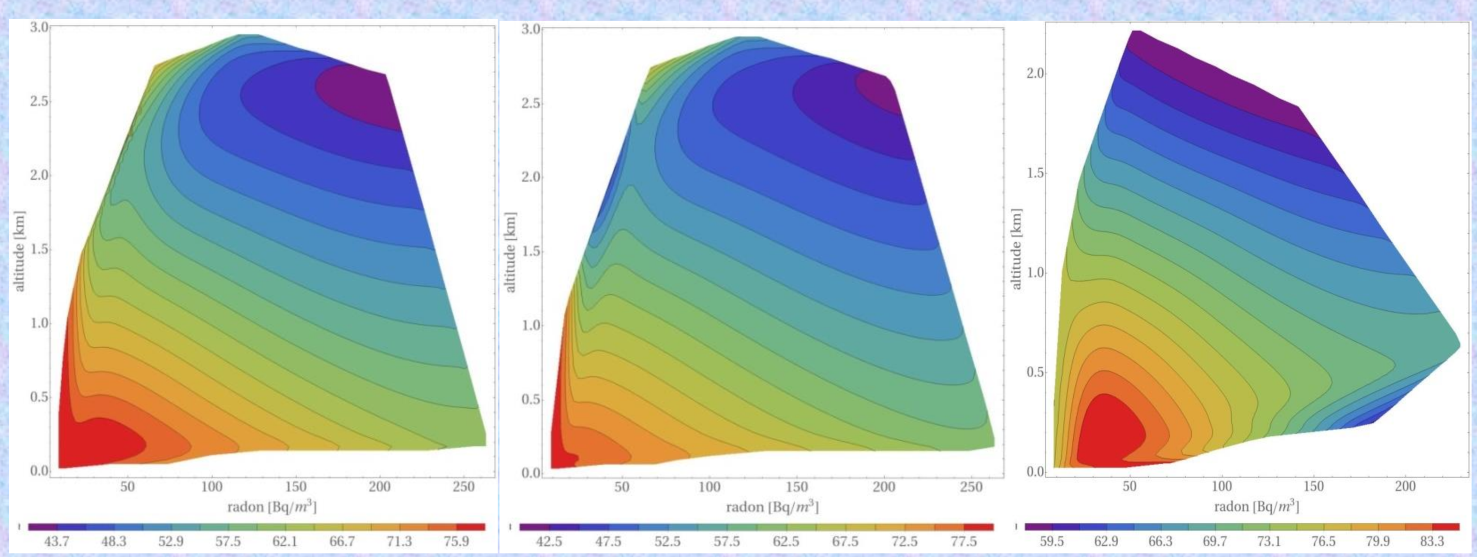

Fig. 5 Maps of lung cancer morbidity in the altitude– radon concentration plane for all population (a), medium smoking prevalence (b), and high smoking prevalence (c) [1]; maps were generated by the Maximum Entropy Method

The presented paper was submitted to the International Journal of Low Radiation (2019).

REFERENCES:

[1] - Simeonov KP, Himmelstein DS (2015) Lung cancer incidence decreases with elevation: evidence for oxygen as an inhaled carcinogen. Peer J 3:e705. https://doi.org/10.7717/peerj.705

[2] - Dobrzyński L, Fornalski KW, Reszczyńska J (2018) Meta-analysis of thirty two case-control and two ecological radon studies of lung cancer. J Radiat Res 59(2): 149-163.