

Meta-analysis of thirty-two case–control and two ecological radon studies of lung cancer

Ludwik Dobrzyński^{1,*}, Krzysztof W. Fornalski^{2,3} and Joanna Reszczyńska¹

¹National Centre for Nuclear Research (NCBJ), ul. A. Soltana 7, 05-400 Otwock-Świerk, Poland

²PGE EJ 1, ul. Mysia 2, 00-496 Warszawa, Poland

³Ex-Polon Laboratory, ul. Podleśna 81a, 05-552 Łazy, Poland

*Corresponding author. National Centre for Nuclear Research (NCBJ), ul. A. Soltana 7, 05-400 Otwock-Świerk, Poland. Tel: +48-22-273-1612; Fax: +48-22-77-93-481; Email: ludwik.dobrzyński@ncbj.gov.pl

(Received 11 April 2017; revised 14 June 2017; editorial decision 29 September 2017)

ABSTRACT

A re-analysis has been carried out of thirty-two case–control and two ecological studies concerning the influence of radon, a radioactive gas, on the risk of lung cancer. Three mathematically simplest dose–response relationships (models) were tested: constant (zero health effect), linear, and parabolic (linear–quadratic). Health effect end-points reported in the analysed studies are odds ratios or relative risk ratios, related either to morbidity or mortality. In our preliminary analysis, we show that the results of dose–response fitting are qualitatively (within uncertainties, given as error bars) the same, whichever of these health effect end-points are applied. Therefore, we deemed it reasonable to aggregate all response data into the so-called Relative Health Factor and jointly analysed such mixed data, to obtain better statistical power. In the second part of our analysis, robust Bayesian and classical methods of analysis were applied to this combined dataset. In this part of our analysis, we selected different subranges of radon concentrations. In view of substantial differences between the methodology used by the authors of case–control and ecological studies, the mathematical relationships (models) were applied mainly to the thirty-two case–control studies. The degree to which the two ecological studies, analysed separately, affect the overall results when combined with the thirty-two case–control studies, has also been evaluated. In all, as a result of our meta-analysis of the combined cohort, we conclude that the analysed data concerning radon concentrations below $\sim 1000 \text{ Bq/m}^3$ ($\sim 20 \text{ mSv/year}$ of effective dose to the whole body) do not support the thesis that radon may be a cause of any statistically significant increase in lung cancer incidence.

Keywords: Bayesian; case–control; ecological; low radiation; lung cancer; radon

INTRODUCTION

Radon, a radioactive gas naturally emanating from Earth's crust, contributes about one-half of the effective dose of ionizing radiation to humans. For this reason, the potential effect of radon on human health has been the subject of many studies worldwide. Of particular interest is whether any correlation exists between the local concentration of radon in the atmosphere and the number of locally observed cases of lung cancer.

In the USA, the recently published 'National Radon Action Plan: A Strategy for Saving Lives' [NRAP <http://www.lung.org/assets/documents/healthy-air/national-radon-action-plan.pdf>] (developed through the collaborative efforts of American Lung Association, American Association of Radon Scientists and Technologists, American Society of Home Inspectors, Cancer Survivors Against

Radon, Children's Environmental Health Network, Citizens for Radioactive Radon Reduction, Conference of Radiation Control Program Directors, Environmental Law Institute, National Center for Healthy Housing, U.S. Environmental Protection Agency, U.S. Department of Health and Human Services, U.S. Department of Housing and Urban Development)] stated that 'radon causes 21 000 lung cancer deaths every year' and that 'radon exposure is the second leading cause of lung cancer'. Both statements were questioned by Siegel *et al.* [1], who presented an analysis of selected epidemiology data; these two citations are in fact quoted from their publication.

In our previous work [2, 3], we analysed 28 published radon studies using the Bayesian approach. We found it difficult to unequivocally correlate the risk of lung cancer with radon concentration. In this

work, we have scrutinised 34 independent radon studies published so far [4–37]. In the present meta-analysis of this published data we used the same Bayesian techniques as applied earlier, but have also included some classical methods of data analysis.

MATERIALS AND METHODS

Meta-analysis

Review of original data from published studies offers an opportunity to make some general observations. To avoid any bias in our analysis, we decided to consider all the reliable datasets published over the years in peer-reviewed journals. All these publications were carefully studied with regard to the applied methodology of evaluating radon concentration, to dosimetry techniques, and to the evaluated health end-point, i.e. the odds ratio (OR) or relative risk (RR). From each study, adjusted RR or OR estimates and their 95% confidence intervals (CIs) for the mean values over given categories of radon concentrations in Bq/m^3 , were obtained. In all, some 28 000 cases and 900 000 controls were covered, as shown in Table 1, where a summary of the features of the analysed publications is given. It is apparent from this table that all these data could be included in our further analysis, as no lack of precision nor any inconsistencies in any of these studies are apparent. Cases and controls were subject to well-defined criteria, and residential histories were given for all periods considered—whether long or short, periods of measurements were provided, as were clear indicators of smoking habits, sex and mean age for the selected groups, together with information about the applied methodology of group selection. All the data that we considered in our further analysis were taken from these original publications satisfying the above criteria. Our aim was to test the hypothesis that presence of residential radon increases the risk of lung cancer occurrence. We also wished to evaluate the coherence between the published studies that we analysed.

Data

We have given equal importance to all the data, following the authors' specifications regarding time spent indoors and of in-house measured radon levels. The control groups were carefully chosen to match the exposed groups as accurately as possible. Radon exposure periods ranged between 5 and 25 years. Five years were considered to be the minimum period for cancer occurrence. Confounding factors were taken into account, as given in Table 2.

In view of the considerable scatter and generally rather high uncertainties of values of ORs or RRs in the published results, we deemed it to be inappropriate to fit the OR/RR vs radon concentration (or vs absorbed dose as both quantities are proportional) dependence by mathematical functions more complicated than a 2nd order polynomial. Therefore the following functions (models) only were considered: constant (no risk), linear (by two approaches) and quadratic (parabolic or linear-quadratic, as some might prefer). A specific linear relationship called the Linear No-threshold Model (LNT) is often used in the literature as if it was the best null hypothesis. While in our opinion LNT is not the best null hypothesis, we nevertheless decided not to neglect any possible hypothesis, but only to compare the results obtained.

The studies listed in Table 2 can be generally divided into two main groups: case–controls studies (thirty-two papers) and

ecological studies (two papers) (see Fig. 1a). We did not consider any of the radon studies of miners because their methodology is different (e.g. different confounding factors are present, especially the strong effects of smoking and dust), and because their environmental conditions are not representative of those typical for residential radon data. One should note that some of the publications listed in Table 2 (as noted in their description) have already been used in pooled studies [38–42], under different analyses.

This work consists of two parts: (i) separate classical (least-square) and Bayesian analysis of OR/RR in case–control data, and (ii) meta-analysis of all data using both classical and robust Bayesian methods, for case–control and for ecological data—separately and in combination.

Concerning the first part, we have concentrated on ORs of cancer morbidity [5, 10, 12, 13, 15, 16, 23, 24, 28, 29, 31, 32, 35]. Similarly, there are also publications in which RRs of cancer morbidity are considered [4, 6–9, 22, 27, 30, 34]. The ORs of mortality were also studied [14, 17, 21, 33]. As a rule, authors of those papers used multivariate logarithmic regression analysis with a preferred linear trend. The RRs of deaths due to lung cancer have also been considered [19]. However, due to difficulties in assigning doses and due to limited data on the RR of mortality, we took the latter paper [19] into account only in the second part of this work.

The four distinctly analysed datasets include (see Fig. 1b):

- (i) cancer mortality ORs (13 data points). Mean radon concentration up to 325 Bq/m^3 .
- (ii) cancer morbidity ORs (38 data points). Mean radon concentration up to 299.5 Bq/m^3 .
- (iii) cancer morbidity RRs (20 data points). Mean radon concentration up to 475 Bq/m^3 .
- (iv) meta-analysis of the above three datasets (71 data points).

In what follows, we shall use the notion of Relative Health Factor (RHF) for all these three risk factors. Thus, the RHF shall be understood as either mortality OR, or morbidity OR, or morbidity RR, depending on the context.

The above datasets form a subset of the original data that also contained higher mean values of radon concentrations; however, these were unspecified (listed, e.g. as $>500 \text{ Bq/m}^3$). In order to be as consistent as possible, in our primary analysis we had omitted such data points, as well as several points with an unclear description, testing however their influence on the final results in the second part of our work. To that end we assumed that the possible range of radon concentrations was that given in ref. [2]. This explains the origin of points with abscissa $>500 \text{ Bq/m}^3$ in Fig. 1a.

Ecological data were not taken into account in this part of our analysis in order to avoid potential ecological fallacy (bias), especially due to the results of Cohen [11, 43].

To facilitate comparison with our previous work [2], in the second part of our analysis all data were divided into different datasets according to the type of study (case–control/ecological) and to the dose/concentration range. We used the following commonly accepted radon dose-concentration conversion factors: $1 \text{ Bq/m}^3 = 0.179 \text{ mSv/year}$ of annual equivalent dose to lungs (H), and effective (whole-body) dose, $E = 0.12 H$ [44, 45].

Table 1. Data for the 34 studies analysed in the present meta-analysis, including number of individuals, measurement methodology, personal characteristics and habits, and number of adjusted factors in OR/RR calculations

Author/First author	Year	Cases	Controls	Measurement type (as published)	Smoking habits	Age	Sex	Number of confounding factors
Blot	1990	308	356	radon detectors	+	+	women	4
Hystad	2014	2390	3507	mapping	+	+	+	19
Torres-Duran	2014	192	329	alfa track detector	never-smokers	+	+	3
Bochicchino	2005	384	404	radon detectors	+	+	+	5
Sandler	2006	1474	1811	alfa track etch detectors	+	+	+	–
Wilcox	2008	561	740	alfa track monitoring	+	+	+	6
Alavanja	1994	538	1183	alfa track detectors	non-smoking	+	women	1
Alavanja	1999	512	546	CR-39 alpha-particle detectors	+	+	women	5
Barros-Dios	2002	163	241	alfa track detector	+	+	+	6
Barros-Dios	2012	349	513	alfa track detectors (CR-39, Radosys)	+	+	+	3
Auvien	1996	517	517	alfa track passive detector	+	+	+	4
Brauner	2012	589	52 692	model based predictions	+	+	+	10
Baysson	2004	486	984	2 Kodalpha LR 115 detectors	+	+	+	5
Conrady	2002	72	240	The ALTRAC dosimeters	non-smoking	+	women	1
Darby	1998	484	1637	small passive NRPB's radon detectors	+	+	+	5
Field	2000	413	614	Radtrak alpha track detectors	+	+	+	4
Kreienbrock	2001	1449	2297	solid-state nuclear track detector	+	+	+	2
Letourneau	1994	738	738	dosimeters with polyethylene-lined cap	+	+	+	2
Lagarde	2001	258	487	radon dosimeters	never-smokers	+	male	–
Pershagen	1992	210	408	alfa—track detectors	+	+	women	3
Pershagen	1994	1360	2847	solid-state alpha track detectors	+	+	+	5
Pisa	2001	138	291	Dosimeters with two LR115 trace revealers	+	+	+	3
Ruosteenoja	1991	238	434	solid-stale nuclear track detectors	+	+	male	2
Ruosteenoja	1996	164	331	radon dosimeters	+	+	male	2
Tomasek	2001	173	3221	2 integral Kodak detectors LR115	–	–	–	–
Wichmann	2005	2963	4232	alfa track detectors	+	+	+	6
Thompson	2008	200	397	etch-track detectors	+	+	+	5
Turner	2011	3493	811 961	data from various sources: EPA SRRS, U.S. NRRS, LBL and Cohen's.	+	+	+	19
Sobue	2000	28	36	alfa track detectors	+	+	+	4

Continued

Table 1. Continued

Author/First author	Year	Cases	Controls	Measurement type (as published)	Smoking habits	Age	Sex	Number of confounding factors
Wang	2002	768	1659	alfa track detectors	+	+	+	5
Schoenberg	1990	433	402	alfa track detector	+	+	women	5
Oberaigner	2002	-	-	-	-	-	+	-
Conrady	1996	2155	no information	nuclear tracking detector	+	0-99	women	1
Cohen	1995	1601 data points (ecological)		data from PITT, EPA, STATE	+	+	+	56

Hereafter, we shall use the term ‘effect’ either in the sense of the OR or the RR, depending on the context.

Uncertainties

In both parts of our analysis we used the values of uncertainties calculated by the authors of their original papers as the 95% confidence interval (CI) under the assumption of log-normal distribution. However, standard deviation related to 68% CI is normally used in classical least-square fits. Therefore, we have verified whether the use of 68% CI instead of 95% CI has any influence on the values of the fitted parameters. In none of the quoted papers did we find any precise information on the manner by which the asymmetric uncertainties were calculated. Therefore, our analysis was also carried out assuming average radon concentrations, i.e. $\frac{sl+su}{2}$, where sl and su are the lower and the upper bounds (respectively) of the range shown in the original paper. This approach may be considered to be an element of sensitivity analysis showing to what extent any shift in radon concentration values may influence the final results.

The published radon data are typically presented in bins (concentration ranges) of various widths. In order to include both vertical and horizontal uncertainties in our analysis, a rectangular distribution of radon concentration within each bin was assumed. In this approximation, radon concentration variance is equal to $\sigma_x^2 = \frac{1}{12}W^2$, where W is the bin width. Although this value of σ_x , if considered in the usual context ($\pm 2\sigma_x = 95\%$ of the CI), may underestimate the actual uncertainty, we have shown in our analysis that this approximation does not lead to any substantial divergence of the final results.

Because the widths of radon concentration bins were relatively large in all datasets while the effect (vertical) error bars were quite asymmetric with respect to their assigned values, we tested the sensitivity of the result on the choice of weights representing these error bars for every data/model combination. In particular, we fitted our models assuming artificially symmetric vertical error bars (arithmetic averages of the original minimum and maximum values), asymmetric error bars (as given in the original papers for 95% CI), and, additionally, estimating that those uncertainties represent 68% of the CI value, as is typically used in experimental physics. While any of these approaches can be justified as being mathematically

correct, if considered as a sensitivity analysis, they can reveal to what degree one can trust the final conclusions.

Models

In the first part of the presented paper the following four models of the effect on radon concentration dependence were used to analyse every dataset:

- (i) no dependence (‘zero effect’): effect = a (constant)
- (ii) linear non-threshold dependence (LNT): effect = $1 + br$
- (iii) linear dependence: effect = $a + br$
- (iv) quadratic dependence: effect = $a + br + cr^2$
where r is radon concentration.

In the first part of our analysis, all four models were tested; however, based on the results obtained, in the second part of our analysis, the LNT hypothesis was discarded. The remaining three hypotheses (constant, linear, quadratic dependence) were fitted using the classical (least squares) and the robust Bayesian regression methods for all 34 studies, case-control and ecological studies being treated separately. Next, the relative strengths of the hypotheses were compared with each other, applying the Bayesian model selection method used earlier [2]. In this context, we recall the conclusion from that earlier work of ours (meta-analysis of 28 studies), namely that the ‘zero-effect’ hypothesis was ~90 times more likely than the LNT hypothesis, and that ‘zero-effect’ was also clearly preferred over the more general linear model.

Bayesian analysis

Within robust Bayesian regression analysis, it is assumed that each experimental point E_i and its original uncertainty σ_{0i} are given by appropriate Gaussian distributions (Bayesian likelihood functions) and that the prior uncertainty function is $p(\sigma_i) = \sigma_{0i} \sigma_i^{-2}$ [46, 47]. Introduction of such a prior follows from observing large data scatter, not justified by the claimed uncertainties σ_{0i} . Consequently, values of effective uncertainties σ_i become equal to or larger than the original ones, σ_{0i} . This results in a modification of the joint probability of obtaining the given dataset, expressed usually as $\exp(-\sigma^2/2)$, where σ^2 is the classical misfit function. This probability has now to be

Table 2. List of 34 studies analysed in the present meta-analysis

Country/region/group	Source	Type of data	Type of study
China, Shenyang	Blot <i>et al.</i> 1990 [4]	RR—MB	C-C ^{a,c}
Canada	Hystad <i>et al.</i> 2014 [36]	OR—MB	E
Spain, Galicia	Torres-Durán <i>et al.</i> 2014 [37]	OR—MT	C-C
Italy, Mediterranean	Bohicchio <i>et al.</i> 2005 [28]	OR—MB	C-C ^b
USA, Connecticut and Utah	Sandler <i>et al.</i> 2006 [30]	RR—MB	C-C
USA, New Jersey II	Wilcox <i>et al.</i> 2008 [31]	OR—MB	C-C
USA, Missouri I	Alavanja <i>et al.</i> 1994 [10]	OR—MB	C-C ^a
USA, Missouri II	Alavanja <i>et al.</i> 1999 [16]	OR—MB	C-C
Spain	Barros-Dios <i>et al.</i> 2002 [23]	OR—MB	C-C ^b
Spain, Galicia II	Barros-Dios <i>et al.</i> 2012 [35]	OR—MB	C-C
Finland I	Auvinen <i>et al.</i> 1996 [13]	OR—MB	C-C ^{a,b}
Denmark	Bräuner <i>et al.</i> 2012 [34]	RR—MB	C-C
France	Baysson <i>et al.</i> 2004 [27]	RR—MB	C-C ^b
Germany, Schneeberg	Conrady <i>et al.</i> 2002 [25]	OR—MT	C-C
England, south-west	Darby <i>et al.</i> 1998 [15]	OR—MB	C-C ^b
USA, Iowa	Field <i>et al.</i> 2000 [18]	OR—MB	C-C
Germany, western	Kreienbrock <i>et al.</i> 2001 [20]	RR—MB	C-C ^b
Canada, Winnipeg	Letourneau <i>et al.</i> 1994 [8]	RR—MB	C-C ^a
Sweden I	Lagarde <i>et al.</i> 2001 [22]	RR—MB	C-C ^b
Sweden II	Pershagen <i>et al.</i> 1992 [7]	RR—MB	C-C ^{a,b}
Sweden III	Pershagen <i>et al.</i> 1994 [9]	RR—MB	C-C ^{a,b}
Italy, Alps	Pisa <i>et al.</i> 2001 [21]	OR—MT	C-C
Finland II	Ruosteenoja 1991 [6]	OR—MB	C-C ^a
Finland III	Ruosteenoja <i>et al.</i> 1996 [12]	RR—MB	C-C ^b
Czech Republic	Tomášek <i>et al.</i> 2001 [19]	RR—MT	C-C ^b
Germany	Wichmann <i>et al.</i> 2005 [29]	OR—MB	C-C ^b
USA, Worcester	Thompson <i>et al.</i> 2008 [32]	OR—MB	C-C
USA II	Turner <i>et al.</i> 2011 [33]	OR—MT	C-C
Japan, Misasa	Sobue <i>et al.</i> 2000 [17]	OR—MT	C-C
China, Gansu	Wang <i>et al.</i> 2002 [24]	OR—MB	C-C ^c
USA, New Jersey	Schoenberg <i>et al.</i> 1990 [5]	OR—MB	C-C ^a
Austria	Oberaigner <i>et al.</i> 2002 [26]	RR—MB	C-C ^b

Continued

Table 2. Continued

Country/region/group	Source	Type of data	Type of study
Germany, Saxony	Conrady & Martin 1996 [14]	OR—MT	C-C
USA	Cohen 1995 [11]	OR—MT	E

E = ecological study, C-C = case-control study, RR = relative risk, OR = odds ratio, MB = morbidity, MT = mortality.

^aThis paper is also a part of 8 pooled studies by Lubin and Boice [38].

^bThis paper is also a part of 13 pooled European studies by Darby *et al.* [40].

^cThis paper is also a part of pooled Chinese studies by Lubin *et al.* [39].

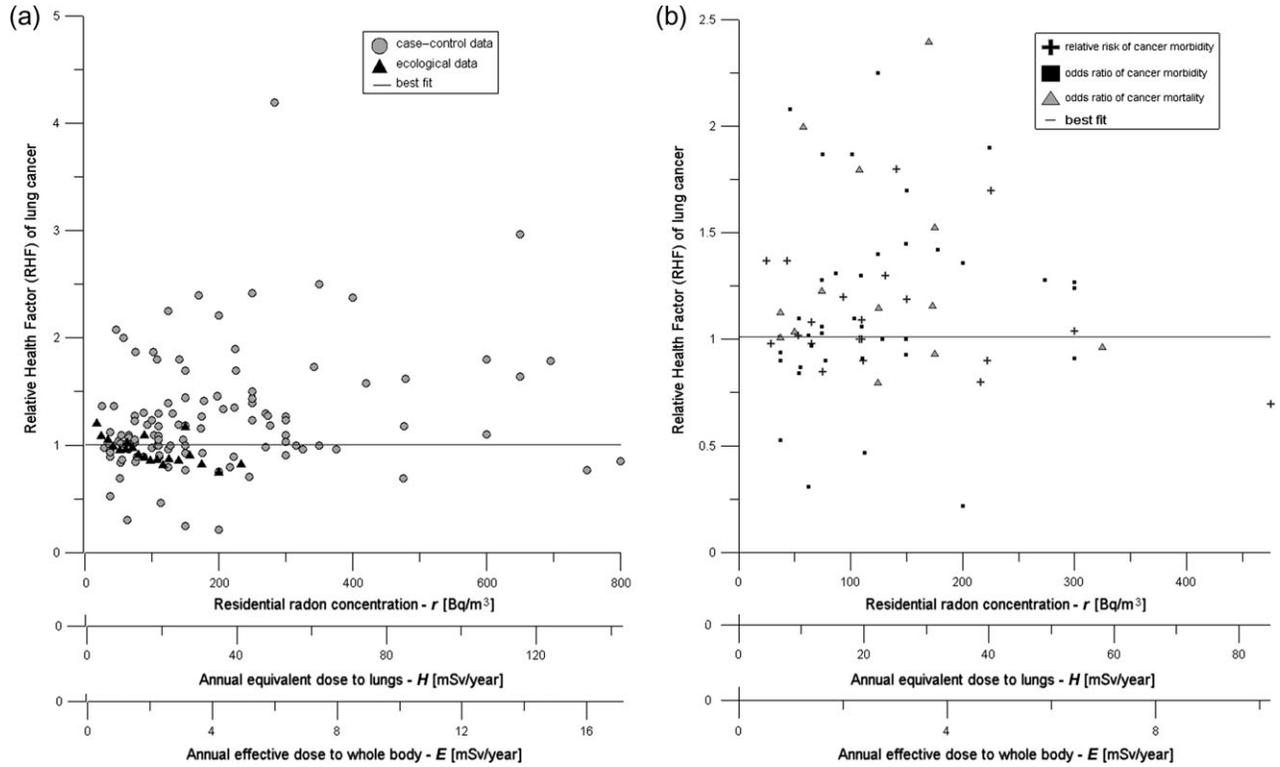


Fig. 1. (a) Raw data points from 34 radon studies (a total of 134 points) listed in Table 1 and 2. For legibility, uncertainty bars are intentionally omitted. The 21 points from ecological studies and 113 of case-control studies are marked differently. The solid line represents Model 1 ($RHF = \text{const}$). The best fit line is somewhat below the center of gravity of the data points because the data uncertainties are typically asymmetrical, with smaller uncertainties downwards and larger upwards. (b) Data set limited to 71 data points, shown as cancer mortality ORs (13 points), cancer morbidity ORs (38 points) and cancer morbidity RRs (20 points).

multiplied by $p(\sigma_i)$ and integrated over σ_i within the limits (σ_{0i}, ∞) . Following this procedure, one arrives the general posterior probability distribution $P = \prod P_i$:

$$P = \prod_i \frac{\sigma_{0i}}{(M_i - E_i)^2 \sqrt{2\pi}} \left[1 - \exp\left(\frac{-(M_i - E_i)^2}{2\sigma_{0i}^2}\right) \right] \quad (1)$$

where M corresponds to the model (expression, or curve) being fitted to the existing data points. This probability is next minimized with respect to the searched parameters, as in the conventional maximum

likelihood method. The interesting advantage of such an approach is that the outliers in the dataset acquire substantially lower weights, compared with the classical minimization of the χ^2 function and thus their influence on the fitted values is substantially limited. The disadvantage of this approach is that minimization of Eq. (1) requires more computational work than minimization of the χ^2 function, and that the uncertainties of the fitted parameters are different than those obtained using the conventional method.

Having established the posterior probability distribution, one can adapt it to the model selection algorithm [46, 47] and calculate the plausibility of the tested model M , as:

$$P(\text{Model}|\text{Data}) \propto P(\text{Data}|\text{Model}) \approx \sum_{i=1}^N P_i(\text{Model}_M) \times \prod_{\lambda=1}^n \frac{\sigma_i \sqrt{2\pi}}{\lambda_{\max} - \lambda_{\min}} \equiv N_M \quad (2)$$

where λ corresponds to the parameter(s) of the model [47]. The relative likelihood of two models displaying individual plausibilities N_A and N_B is given by:

$$W = \frac{N_A}{N_B} = \frac{P(\text{Model}_A|\text{Data})}{P(\text{Model}_B|\text{Data})} \quad (3)$$

If this ratio exceeds 1, model A is more likely than model B.

We also used the classical least squares method to fit the models to the same datasets.

The differences between classical and Bayesian approaches have been extensively discussed elsewhere [46, 47].

RESULTS

Preliminary analysis

Results of our preliminary analysis are compiled in Tables 3–5. First of all, it may be noted that the choice of error bars has a rather negligible influence on the values of the best-fitted parameters. Also introduction of horizontal error bars did not much affect the results obtained. The relatively low slope of the straight line (cf. linear models 2 and 3) is the main reason for the above observations; hence, the values of vertical error bars dominate in the least square fits. Low values of the misfit function, χ^2 :

$$\chi^2 = \frac{1}{N - N_p} \sum_i w_i (E_i - M_i)^2 \quad (4)$$

where w_i is the weight of the i th point, N is the number of points, and N_p is the number of parameters, are due to the 95% CI values assumed in these calculations. By accepting such uncertainties in

Table 3. Values of the best-fitted parameters of the ‘Zero effect’ Model 1 (effect = a = constant), using the classical least square method for 95% CI

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
symmetric uncertainties	$a = 1.02 \pm 0.05$	$a = 0.99 \pm 0.06$	$a = 1.04 \pm 0.08$	$a = 1.01 \pm 0.03$
	$\chi^2 = 0.27$	$\chi^2 = 0.53$	$\chi^2 = 0.31$	$\chi^2 = 0.42$
asymmetric uncertainties	$a = 1.04 \pm 0.05$	$a = 1.04 \pm 0.05$	$a = 1.04 \pm 0.05$	$a = 1.05 \pm 0.03$
	$\chi^2 = 0.63$	$\chi^2 = 0.55$	$\chi^2 = 0.29$	$\chi^2 = 0.44$

Table 4. Values of best-fitted parameters of the LNT Model 2 (effect = $1 + b r$), using the classical least square method for 95% CI

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
symmetric	$b = (0.53 \pm 4.56) 10^{-4}$	$b = (1.84 \pm 5.09) 10^{-4}$	$b = (-1.09 \pm 5.84) 10^{-4}$	$b = (0.60 \pm 2.93) 10^{-4}$
	$\chi^2 = 0.28$	$\chi^2 = 0.53$	$\chi^2 = 0.33$	$\chi^2 = 0.40$
asymmetric	$b = (4.36 \pm 4.51) 10^{-4}$	$b = (5.10 \pm 4.98) 10^{-4}$	$b = (6.99 \pm 6.40) 10^{-4}$	$b = (5.33 \pm 3.00) 10^{-4}$
	$\chi^2 = 0.67$	$\chi^2 = 0.45$	$\chi^2 = 0.35$	$\chi^2 = 0.43$

Table 5. Values of best-fitted parameters of the Linear Model 3 (effect = $a + b r$), using the classical least square method for 95% CI

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
symmetric	$a = 1.05 \pm 0.05$	$a = 0.90 \pm 0.05$	$a = 1.11 \pm 0.04$	$a = 1.02 \pm 0.02$
	$b = (-2.33 \pm 4.25) 10^{-4}$	$b = (8.59 \pm 4.44) 10^{-4}$	$b = (-6.59 \pm 3.17) 10^{-4}$	$b = (-0.73 \pm 2.07) 10^{-4}$
	$\chi^2 = 0.60$	$\chi^2 = 0.47$	$\chi^2 = 0.35$	$\chi^2 = 0.44$
asymmetric	$a = 1.05 \pm 0.08$	$a = 0.93 \pm 0.13$	$a = 1.12 \pm 0.06$	$a = 1.04 \pm 0.05$
	$b = (-1.81 \pm 7.11) 10^{-4}$	$b = (1.34 \pm 1.44) 10^{-4}$	$b = (-5.01 \pm 5.01) 10^{-4}$	$b = (1.57 \pm 4.84) 10^{-4}$
	$\chi^2 = 1.02$	$\chi^2 = 1.38$	$\chi^2 = 0.50$	$\chi^2 = 0.98$

classical least-square fitting, one strongly (by a factor of ~4) lowers the weights, thus decreasing the χ^2 value by a factor of ~4. We followed the methodology described by Cantrell [48] and York *et al.* [49] when fitting straight lines with vertical and horizontal error bars.

The best-fitted value of a in Model 1 (zero effect) is $a = 1.0$, for all three datasets (Table 3). This result remains unchanged if all data are aggregated and analysed as a combined dataset (Dataset 4).

It may be interesting to see how often one encounters a given value of the RHF in the data. Figure 2 shows an RHF histogram calculated for asymmetric uncertainties. The histogram obtained for symmetrized uncertainties is not much different, although the distribution obtained is of course wider. Histograms obtained for individual Datasets 1–3 are qualitatively not different from that shown in Fig. 2. In all cases, a distinct maximum within the (1.00, 1.05) bin is observed. Slight elevation of the mean value of RHF is also seen in Table 3.

The values of the a constant obtained using Bayesian methods are slightly—but consistently—larger than 1.00; for aggregated data, $a = 1.05 \pm 0.04$. The difference is clearly within uncertainty limits; however, it reflects well what is seen in the histogram of Fig. 2. Practically the same results are obtained for symmetrized uncertainties. This could be explained if one assumed that hormesis is at work at the lowest observed irradiation levels (radon concentrations). Such a possibility has already been considered by Doss [50].

Table 5 compiled for Model 3 (general linear model) shows that the a intercept value may be slightly larger or smaller than 1, depending on the data selected. Then, the slope becomes either negative or positive, and it becomes larger as the deviation of a from unity increases. This is typical for correlated variables. Nevertheless, slopes are generally negative (except for dataset 2), i.e. the OR/RR effect decreases with increasing radon concentration. Bayesian analysis yields $a = 1.04 \pm 0.29$ and $b = 0.00015 \pm$

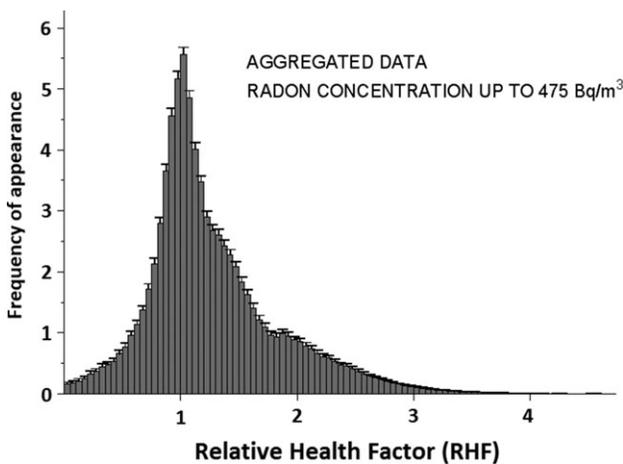


Fig. 2. Distribution of values of Relative Health Factor aggregated from datasets 1 to 3. Asymmetric uncertainties, as given by authors of the original publications, were assumed.

0.00165, which means that the slope of the line fitted within Linear Model 3 is practically zero, in spite of large uncertainties in both parameters. This is in agreement with the results of the conventional approach (see Table 5). Apparently, the ‘zero effect’ Model 1 is favoured. The a constant is larger than 1.00, just as in the case of Model 1.

In the LNT (Model 2), the a intercept is fixed at 1, by definition. The slope of the fitted line, see Fig. 3, is positive—not surprising in the light of the experience with the linear model discussed above. Again, the value of the slope bears a large uncertainty, its sign depending to a large extent on the method of analysis selected.

It is not strange that even much larger uncertainties in the fitted parameters appear if a 2nd order polynomial is fitted to the data (Quadratic Model 4). The obtained goodness-of-fit parameter value is reasonably acceptable, but the values of the fitted parameters are not. Moreover, Model 4 usually yields an inverted parabola, a physically unreasonable result. We conclude that Model 4 should be discarded.

The overall conclusion is that, irrespective of whether one analyses ORs or RRs of morbidity or mortality using the maximum likelihood or Bayesian approaches (with either asymmetric or symmetrized uncertainties), the scatter of experimental points and their high statistical uncertainties are similar; hence, it is reasonable to analyse all datasets in conjunction [2]. This option is further justified by observing that if the values of RR and OR are both close to unity, the OR does not differ much from the relative ratio [51]—thus, they can be treated jointly as RHF.

Analysis of the complete dataset

Results of our analysis of all the available data are compiled in Table 6. The relative health factor (RHF) is presented as a function of dose to facilitate comparisons with the mixed results of our previous paper [2]. Figure 1a shows 134 points representing the RHF of lung cancer plotted against three interrelated horizontal axes: annual equivalent dose to lungs H [mSv/year], annual effective dose (to the whole body) E [mSv/year], and average indoor radon concentration [Bq/m³]. All data points were taken directly from the 34 radon studies, with their original uncertainties (95% CIs). As mentioned earlier, the range of radon concentrations is broader than that in our initial analysis, as additional data points have been included and represent mean values in the outermost part of the relevant concentrations. As the authors of the original papers did not report uppermost concentrations, in such cases we have assumed that their range was the same as that of the preceding data points. This situation could have potentially affected the final results and favoured LNT. As it turns out, this was not the case.

Constant, linear and quadratic functions were fitted separately to points from thirty-two case-control studies only, and to points from two ecological studies that included Cohen’s data [11]. We aggregated all the points from these different sources on a single plot, as we did in our previous paper [2]. In fact, such a plot was displayed earlier in the UNSCEAR 2006 Report [44] (where Cohen’s data were presented in Fig. XV together with the results of case-control studies). This approach seems to be acceptable in view of the

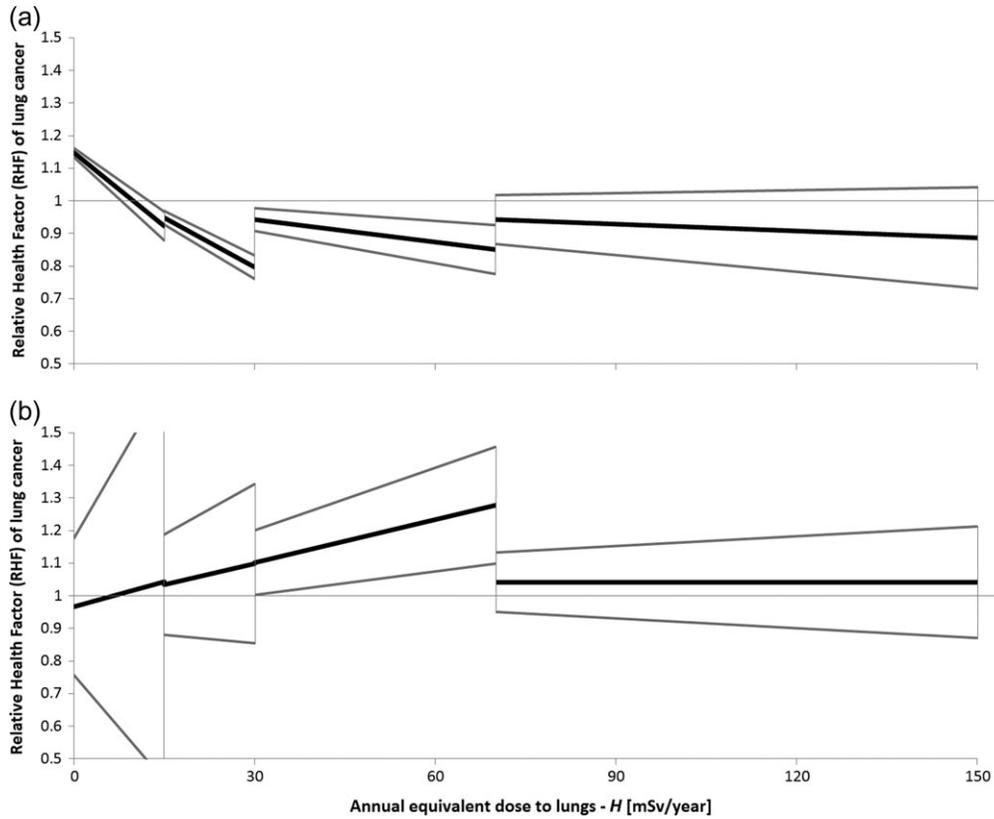


Fig. 3. Best-fitted linear dependences to the data from Fig. 1a using the Bayesian approach fitted separately within four ranges of radon doses for (a) all 34 studies, and (b) for thirty-two case-control studies analysed in the second part of this meta-analysis. Grey lines represent 95% CI. Regressions are based on the data in Table 5.

gathered statistics, and scatter of the data points and of their uncertainties.

In this part of our meta-analysis, the data were analysed over four regions of equivalent dose to lungs H : up to 15 mSv/year, up to 30 mSv/year, up to 70 mSv/year, and up to 150 mSv/year. The asymmetrical distribution of points along the dose axis justifies such an approach. Aggregated results are presented in Table 6. Figure 3 shows the results of Bayesian linear best fit to the data from Fig. 1a divided into these four dose sub-ranges. The constant model fit generally oscillates around $RR = 1$; deviations from $RR = 1$ are in most cases statistically insignificant. The uncertainties shown cover the 95% CI. The impact of Cohen's data [11] on the slope of the best-fitted straight lines is clearly visible over a substantial part of the radon concentration range.

From Fig. 3b, one may get an impression that the data in the range up to 70 mSv/year could support the LNT. This is not so. First, the slope of the fitted line is correlated with the $RR(0)$ value: the larger the intercept, the more negative the slope. An intercept below 1 results in a positive slope of the line. Second, the uncertainty margins are large; thus, such a result cannot be convincing (has insufficient statistical power). Finally, taking into account the complete dataset, one observes that there is no dependence of RHF on dose. On the other hand, Fig. 1a and Table 6 show that values of what is called a null dose actually represent the lowest doses

determined in the studies. This may introduce further uncertainties in the interpretation of the fitted parameters [52].

DISCUSSION

The first part of our meta-analysis was focused on separately analysing data given in individual studies. Thus end-points, such as ORs of lung cancer mortality or morbidity, and RRs of cancer morbidity, were treated separately, rather than jointly as RHF. There is no doubt that only the 'Zero Effect' Model 1 (i.e. no influence of radon concentration/dose on RHF) has delivered a reasonably stable fit for all such datasets. Results obtained for individual datasets using the LNT Model or the Linear Model are inconclusive. The classical least square approach and the Bayesian approach both lead to the same conclusion for aggregated data: no slope of the regression line other than zero has any statistical support. The correlation between the fitted parameters is too strong to accept the slightly positive slope obtained for LNT (Model 2) as an indication of a trend in RHF dependence on radon dose. Quadratic fits did not add any new value to this analysis, so we decided to abandon Quadratic Model 4 altogether. In view of the presented results, one gains confidence in analysing the data *en masse*, i.e. in treating all datasets as if they described the same RHF end-point.

The second part of our meta-analysis was conducted over different ranges of radon concentrations/doses. The entire dose range

Table 6. Values of best-fitted parameters for the Constant, Linear and Quadratic models, over four ranges of annual equivalent dose to lungs, H , and their respective likelihoods, as delivered by the Bayesian model selection algorithm

No. of studies	Model	Annual equivalent dose to lungs H								Model selection likelihood ^c N_M	
		Up to 15 mSv/year (84 Bq/m ³)		Up to 30 mSv/year (168 Bq/m ³)		Up to 70 mSv/year (391 Bq/m ³)		Full range ^d : up to 150 mSv/year (838 Bq/m ³)			
		Bayesian method	Classical method	Bayesian Method	Classical method	Bayesian method	Classical method	Bayesian method	Classical method		
34(all studies)	Constant ^a	a = 1.008 ± 0.003	a = 1.014 ± 0.013	a = 0.975 ± 0.003	a = 0.972 ± 0.010	a = 0.982 ± 0.003	a = 0.973 ± 0.010	a = 0.980 ± 0.003	a = 0.972 ± 0.010	0.3	
	Linear ^b	b = -0.016 ± 0.002	b = -0.016 ± 0.003	b = -0.010 ± 0.001	b = -0.011 ± 0.001	b = -0.002 ± 0.001	b = -0.006 ± 0.001	b = -0.001 ± 0.001	b = -0.002 ± 0.001	0.03	
		a = 1.148 ± 0.014	a = 1.149 ± 0.028	a = 1.100 ± 0.006	a = 1.105 ± 0.017	a = 1.012 ± 0.005	a = 1.048 ± 0.017	a = 0.992 ± 0.005	a = 1.003 ± 0.013		
	Quadratic ^c	c = 0.002 ± 0.001	c = 0.003 ± 0.001	c = 0.001 ± 0.001	c = 0.0008 ± 0.002	c = 0.000 ± 0.001	c = 0.004 ± 0.001	c = 0.000 ± 0.001	c = 0.000 ± 0.001	c = 0.0001 ± 0.0000	0.015
		b = -0.059 ± 0.018	b = -0.059 ± 0.017	b = -0.032 ± 0.006	b = -0.033 ± 0.005	b = -0.021 ± 0.002	b = -0.022 ± 0.002	b = -0.011 ± 0.001	b = -0.008 ± 0.001		
	a = 1.318 ± 0.083	a = 1.321 ± 0.072	a = 1.222 ± 0.035	a = 1.227 ± 0.032	a = 1.164 ± 0.011	a = 1.168 ± 0.020	a = 1.087 ± 0.006	a = 1.068 ± 0.018			
32 (case–controls)	Constant	a = 1.016 ± 0.027	a = 0.995 ± 0.037	a = 1.025 ± 0.023	a = 0.998 ± 0.031	a = 1.049 ± 0.017	a = 1.011 ± 0.028	a = 1.046 ± 0.016	a = 1.006 ± 0.027	1	
	Linear	b = 0.005 ± 0.027	b = 0.002 ± 0.011	b = 0.004 ± 0.006	b = 0.002 ± 0.003	b = 0.004 ± 0.002	b = 0.003 ± 0.001	b = 0.000 ± 0.001	b = 0.000 ± 0.001	0.05	
		a = 0.967 ± 0.210	a = 0.979 ± 0.104	a = 0.970 ± 0.064	a = 0.973 ± 0.045	a = 0.970 ± 0.039	a = 0.963 ± 0.029	a = 1.042 ± 0.021	a = 1.015 ± 0.022		
	Quadratic	c = 0.002 ± 0.008	c = 0.003 ± 0.005	a = 0.000 ± 0.001	c = 0.003 ± 0.005	c = 0.000 ± 0.001	c = 0.000 ± 0.001	c = 0.000 ± 0.001	c = 0.000 ± 0.000	0.1	
		b = -0.040 ± 0.134	b = -0.058 ± 0.091	b = 0.002 ± 0.022	b = -0.008 ± 0.017	b = 0.002 ± 0.007	b = -0.008 ± 0.017	b = 0.006 ± 0.002	b = 0.005 ± 0.002		
	a = 1.148 ± 0.617	a = 1.239 ± 0.408	a = 0.974 ± 0.148	a = 1.037 ± 0.113	a = 0.968 ± 0.067	a = 0.971 ± 0.054	a = 0.933 ± 0.041	a = 0.939 ± 0.033			
2 (ecological) ^d	Constant	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	a = 0.954 ± 0.003	a = 0.967 ± 0.011	1.2	
	Linear	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	b = -0.011 ± 0.001	b = -0.012 ± 0.001	0.8	
								a = 1.111 ± 0.006	a = 1.112 ± 0.020		
Quadratic	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	c = 0.001 ± 0.001	c = 0.001 ± 0.000	8.1		
							b = -0.026 ± 0.005	b = -0.026 ± 0.004			
							a = 1.200 ± 0.032	a = 1.199 ± 0.026			

All uncertainties represent 95% CI. (Aggregated results of the second part of the meta-analysis in this work—Bayesian and classical least squares methods).

^aRelative health factor, RHF = a.

^bRelative health factor, RHF = b [year mSv⁻¹] H [mSv year⁻¹] + a.

^cRelative health factor, RHF = c [year² mSv⁻²] H² [mSv² year⁻²] + b [year mSv⁻¹] H [mSv year⁻¹] + a.

^dRange of doses in ecological studies was 42 mSv/year (235 Bq/m³) only, and there were too few points to divide the range into some sub-ranges.

^eSee Eq. (2); all N_M values were calculated for the full range of equivalent doses (up to 150 mSv/year, up to 42 mSv/year in ecological studies).

was divided into four sub-ranges, and the three simplest models were applied. The Bayesian approach was used to estimate the relative plausibility of each model. The results are shown in the last column of Table 6. It can be seen that the Bayesian likelihood factors usually favour the ‘Zero effect’ Model 1. Taking only constant and linear fits into consideration, it is worth pointing out that if all thirty-two studies are considered together, the constant model (no risk below some dose limit) is ~10 times more likely than the linear model (even if the linear slope values were negative).

In the case of the two ecological studies, quadratic Model 4 prevails due to the parabolic shape of Cohen’s data [11], which dominate within the two studies combined. It is important to note that ecological studies are burdened with certain inherent limitations, e.g. they are highly sensitive to confounding factors. For example, Cohen’s results [11] were criticized for improperly taking into account the effect of cigarette smoking as a confounding factor [53, 54]. In Fig. 4 of the paper by Heath *et al.* [54], the dependence of lung cancer mortality on radon concentration is almost parallel to the number of smokers in the respective counties. The same figure is redrawn in the UNSCEAR 2006 Report [44]. The aforementioned correlation, however, does not prove that improper conclusions were drawn by Cohen in his papers [11, 55]. In fact, Cohen himself considered as many as 54 socio-economic variables that could have served as confounding factors, smoking being the major component. However, this did not change his overall finding of an inverse correlation of radon levels with lung cancer. While Cohen’s specific aim was to find support for the LNT approach, his final conclusion was that LNT should be discarded [43]. Scott [56] independently evaluated Cohen’s data and showed that radon levels up to ~150 Bq/m³ did not increase the lung cancer risk in the exposed people (even in smokers), which he interpreted by

assuming that those people must have had some effective natural protection mechanisms. Identical correlations were also reported by Becker [57].

As Cohen indicated in his letter to Jaworowski [B.L. Cohen’s letter of 30 June 2003 to Prof. Zbigniew Jaworowski (copy available from L.D.)] and repeated in his later paper [55], even a perfect negative correlation between the number of smokers and the number of lung cancers due to exposure to radon would not validate LNT, an opinion shared by Henriksen [58], who also noted that only in Cohen’s data is radon inhaled both indoors and outdoors taken into account. Additionally, case studies by Thompson [32] have clearly evidenced some hormetic effects at low radon concentrations (i.e. below some 250 Bq/m³). Such effects are not seen (being ‘smeared over’) in studies covering much broader ranges of residential radon concentrations.

The so-called ‘ecological fallacy’ issue, which can affect the interpretation of the geographical data of Cohen, was discussed by several authors, including Seiler and Alvarez [59], and Hart [60]. In that context, special care must be exercised when interpreting ecological studies. This is why in our study we had separately analysed the thirty-two case-control studies and the two ecological studies. A nominally case-control study was presented in the work of Hystad *et al.* [36]. However, these authors evaluated radon concentration according to the ‘ecological’ method (i.e. using an area average for each residential address) and considered information collected at an individual level for each of the case and control subjects (including residential history, confounding factors such as smoking habits etc.). Therefore, for our purposes, we decided to classify that study as an ecological one.

A very interesting discussion about radon and its influence on lung cancer mortality is found in the report of Henriksen [58],

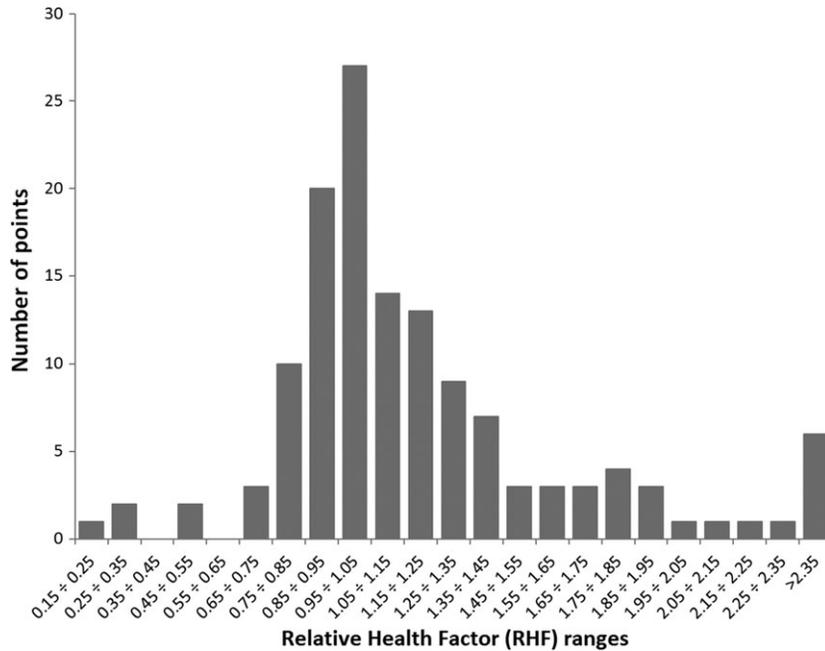


Fig. 4. Distribution of RHF values for the raw data points of Fig. 1a.

where attention is drawn to the important role of radon daughters in the lung/respiratory tract, and to the role played by the lung clearance system—so far never having been taken into account in any of the studies relating lung cancer to radon concentration. Ignoring such factors may result in substantial errors in the estimation of the dose due to radon. Therefore radon concentration would seem to be a more relevant and more representative descriptor of exposure. Even then, discussion of health effects is difficult because radon concentration is known to broadly fluctuate over time due to several environmental factors.

The values of N_M (plausibility of a given model, see Eq. 2) presented here differ from those presented earlier [2]. The differences reflect the larger number of data points (134) in the new analysis, and thus the slightly different ranges of λ parameters (Eq. 2). However, the clear preference of the constant (dose-independent) model persists, confirming the general conclusion of our earlier work (see Fig. 1a). A repeated meta-analysis of 28 studies [2] with present values of λ has yielded $N_M = 0.3, 0.007, 8 \times 10^{-6}$ for the constant, linear, and quadratic models, respectively.

Figure 4 shows the RHF distribution of raw data points from Fig. 1a. The maximum around RHF = 1 appears to be quite natural, because most of the points refer to low radon exposure levels, so RHF values close to 1 may be expected. The asymmetric character of this distribution suggests its non-Gaussian character, but as we have verified, it is not a log-normal distribution either. Two Gaussian curves, a narrow one centred around 1.0 and another one broader and centred around a value >1.0 offer a much better fit. This ‘contamination’ of a purely Gaussian distribution could arise from some unidentified but significant bias present in some publications. Figure 2 shows similar features. The histogram in Fig. 5 demonstrates features similar to those of Fig. 4; however, here each point is represented by a proper Gaussian distribution with symmetrized uncertainty. This representation is much more objective than that of Fig. 4. The data of Fig. 5 could be used as a Bayesian prior function in future studies of the health effects of radon.

Radon is claimed to be one of the most significant causes of lung cancer worldwide. Such claims are often supported by results of various case–control and ecological studies. The site of exposure—domestic or workplace (mostly mines)—is another distinction between these studies. However, in many cases, the conclusion that the influence of radon is detrimental to health has in fact been based on the only hypothesis considered—the LNT model. Support for this hypothesis has been claimed by Darby *et al.* [40], by Krewski *et al.* [61], and by Darby’s European study [62], apparently proving its validity (at least for residential radon). The range of radon concentrations studied by Darby *et al.* was rather limited. But what is more important, many authors [1, 63, 64] have indicated that the conclusions of Darby *et al.* have arisen from circular reasoning, i.e. from confirming a hypothesis that had been assumed as the starting point of their analysis. Even if the original data do not justify the LNT model, it is applied as a hypothesis in most studies. Many authors do not even discuss the possibility of interpreting their data through any other model.

This does not mean that other models have never been tested. In particular, the linear–quadratic, the log-linear, and the linear-with-threshold models were tested (along with LNT) in a pooled

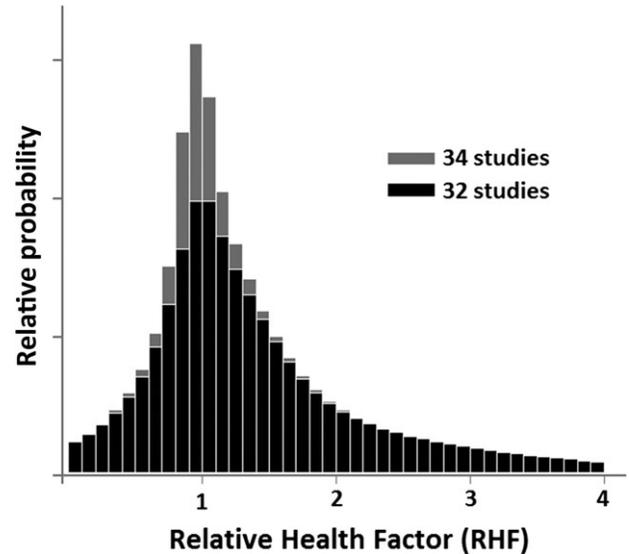


Fig. 5. The histogram of Fig. 4, where each data point is represented by the relevant Gaussian distribution within its uncertainties.

meta-analysis of 13 European case–control studies on lung cancer and residential radon exposure [62]. Results of individual fits did not differ significantly. Even so, the authors preferred the LNT model. However, in line with the Ockham’s razor principle (‘Among competing hypotheses, the one with the fewest assumptions should be selected’ [65]), the robust Bayesian model selection method has always supported the simplest (i.e. constant or simple linear) model [2].

Figure 3 reveals yet another feature, namely that linear fits to limited ranges of radon doses/concentrations are different over different sets of data (whether Cohen’s data [11] are included or not). However, both of these sets of data yield the same linear fit—in fact, as do similar fits of the constant model—if the full range of annual equivalent doses to lungs H up to 150 mSv/year is considered. Again, Model 1 seems to be the most natural hypothesis if the full range of radon concentration data is taken into account. If only the linear model is used, narrower ranges are less representative of the final conclusion.

The issue of using only the LNT model is in fact more complex. According to the LNT hypothesis, risk is proportional to concentration/effective dose, starting from zero Bq/m³ or mSv/year. However, there are no data to support the validity of this hypothesis over the whole range of concentrations/doses, particularly around their lowest values. All existing studies are subject to a number of limitations. The huge scatter in the published results makes it impossible in practice to draw any coherent conclusions [2, 3]. In fact, the LNT hypothesis has been fundamentally criticized in many independent studies (e.g. [63, 64]). Also, a recent letter to the editor by Fornalski *et al.* [63] and Henriksen’s report [58] provide many new arguments. One should note that 838 Bq/m³ [57]—the upper bound of the range of radon concentrations (corresponding

to about 18 mSv/year) is used in calculations of annual effective dose to the whole body (a standard procedure in radiation protection worldwide). This value lies below the annual dose limit for radiation workers (20 mSv/year) applied in most countries over the world, while in many countries the acceptable radon concentration level in homes is 200 Bq/m³ (which corresponds to an annual effective dose to the whole body of ~4 mSv/year). This should be confronted with the broadly accepted annual dose limit to the general population of 1 mSv/year above natural background exposure.

Our results are in line with results published by several authors, including Conrady *et al.* [67] and more recently by Cuttler and Sanders [66, 68]. The former so-called ‘Schneeberg study’ (discussed in [58]) demonstrated that OR lung cancer mortality in humans exposed to radon at concentrations up to 1000 Bq/m³ shows no tendency to increase, against the control group exposed to radon concentrations below 50 Bq/m³. A reverse effect is seen—their data may be interpreted as confirmation of hormesis in action. The authors of the quoted work [58] estimated, based on Cohen’s results, that the No Observable Adverse Effect Level (NOAEL) point occurs at ~2100 Bq/m³ (~376 mSv/year for lungs and ~45 mSv/year for the whole body). One may then postulate that there is no risk of lung cancer at indoor radon concentrations of <1000 Bq/m³ (~180 mSv/year for lungs and ~20 mSv/year for the whole body). It should also be mentioned that the detailed analysis of existing data on residential radon and the potential risk for uranium miners was previously made by Scott [69]. His conclusion regarding an impact of residential radon (limited to ~250 Bq/m³) on lung cancer was in full agreement with Cohen’s conclusion. The paper [69] states, even in the title, that ‘Residential radon appears to prevent lung cancer.’

CONCLUSIONS

Three different models and ‘blind statistics’ were applied to analyse data from 34 radon studies treated separately or in conjunction, to seek a model that could best represent these data, i.e. the most likely representative model. No statistical evidence could support the thesis that the linear model best fits the data over low radon concentrations. We have arrived at an opposite view, supported by several arguments, namely that the linear relationship has typically been pre-assumed in such analyses and that it should be discarded. The data are statistically too weak to accept any more complicated model than that of a constant (‘zero effect’) one. Thus the proper null hypothesis to be used in such analyses is that the effect (risk) is independent of dose. Before other models (functions or mathematical formulae) can be considered in the analysis of low radon exposures, the ‘zero-effect’ hypothesis must first be disproved.

By applying robust Bayesian statistical analysis, we have again found support for the hypothesis that the RR of lung cancer is independent of radon concentration below 838 Bq/m³. This is the most reliable and statistically acceptable conclusion arising from our meta-analysis of data from 34 studies.

ACKNOWLEDGEMENTS

The authors wish to thank Professor Michael P.R. Waligórski for carefully reading and commenting on the text. His invaluable help

in improving the English expression of our paper is greatly appreciated.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING

The research was funded by the National Centre for Nuclear Research (NCBJ), Poland.

REFERENCES

1. Siegel JA, Pennington ChW, Sachs B *et al.* Rectifying radon’s record: an open challenge to the EPA. *Int J Radiol Imaging Technol* 2016;2:1–5.
2. Fornalski KW, Dobrzyński L. Pooled Bayesian analysis of twenty-eight studies on radon induced lung cancers. *Health Phys* 2011;101:265–73.
3. Fornalski KW, Dobrzyński L. Response to Pawel and Puskin. *Health Phys* 2012;102:352–3.
4. Blot WJ, Xu ZY, Boice JD *et al.* Indoor radon and lung cancer in China. *J Natl Cancer Inst* 1990;82:1025–30.
5. Schoenberg JB, Klotz JB, Wilcox HB *et al.* Case–control study of residential radon and lung cancer among New Jersey women. *Cancer Res* 1990;50:6520–4.
6. Ruosteenoja E. Indoor radon and risk of lung cancer: an epidemiological study in Finland. *Dissertation*. Department of Public Health, University of Tampere, Finnish Government Printing Centre, Helsinki, 1991.
7. Pershagen G, Liang ZH, Hrubec Z *et al.* Residential radon exposure and lung cancer in Swedish women. *Health Phys* 1992; 63:179–86.
8. Letourneau EG, Krewski D, Choi NW *et al.* Case–control study of residential radon and lung cancer in Winnipeg, Manitoba, Canada. *Am J Epidemiol* 1994;140:310–22.
9. Pershagen G, Akerblom G, Axelson O *et al.* Residential radon exposure and lung cancer in Sweden. *N Engl J Med* 1994;330: 159–64.
10. Alavanja MCR, Brownson RC, Lubin JH *et al.* Residential radon exposure and lung cancer among nonsmoking women. *J Natl Cancer Inst* 1994;86:1829–37.
11. Cohen BL. Test of the Linear No-Threshold Theory of radiation carcinogenesis for inhaled radon decay products. *Health Phys* 1995;68:157–74.
12. Ruosteenoja E, Makelainen I, Rytomaa T *et al.* Radon and lung cancer in Finland. *Health Phys* 1996;71:185–9.
13. Auvinen A, Makelainen I, Hakama M *et al.* Indoor radon exposure and risk of lung cancer: a nested case–control study in Finland. *J Natl Cancer Inst* 1996;88:966–72.
14. Conrady J, Weniger MK. Modelle—spezifischere analytische Studien zum Radonrisiko in Wohnungen sind notwendig. *Bundesgesundheitsblatt* 1996;19:106–10.
15. Darby S, Whitley E, Silcocks P *et al.* Risk of lung cancer associated with residential radon exposure in south-west England: a case–control study. *Br J Cancer* 1998;78:394–408.

16. Alavanja MC, Lubin JH, Mahaffey JA et al. Residential radon exposure and risk of lung cancer in Missouri. *Am J Public Health* 1999;89:1042–8.
17. Sobue T, Lee VS, Ye W et al. Residential radon exposure and lung cancer risk in Misasa, Japan: a case–control study. *J Radiat Res* 2000;41:81–92.
18. Field RW, Steck DJ, Smith BJ et al. Residential radon gas exposure and lung cancer. The Iowa Radon Lung Cancer Study. *Am J Epidemiol* 2000;151:1091–102.
19. Tomášek L, Kunz E, Muller T et al. Radon exposure and lung cancer risk—Czech cohort study on residential radon. *Sci Total Environ* 2001;272:43–51.
20. Kreienbrock L, Kreuzer M, Gerken M et al. Case–control study on lung cancer and residential radon in Western Germany. *Am J Epidemiol* 2001;153:42–52.
21. Pisa FE, Barbone F, Betta A et al. Residential radon and risk of lung cancer in an Italian alpine area. *Arch Environ Health* 2001;56:208–15.
22. Lagarde F, Axelsson G, Damber L et al. Residential radon and lung cancer among never-smokers in Sweden. *Epidemiology* 2001;12:396–404.
23. Barros-Dios JM, Barreiro MA, Ruano-Ravina A et al. Exposure to residential radon and lung cancer in Spain: a population-based case–control study. *Am J Epidemiol* 2002;156:548–55.
24. Wang Z, Lubin JH, Wang L et al. Residential radon and lung cancer risk in a high-exposure area of Gansu Province, China. *Am J Epidemiol* 2002;155:554–64.
25. Conrady J, Martin K, Lembcke J et al. The true size of the lung cancer risk from indoor radon: hidden behind a smoke screen? *Int Congr Ser* 2002;1225:253–8.
26. Oberaigner W, Kreienbrock L, Schaffrath Rosario A et al. *Radon und Lungen krebs im Bezirk Imst/Österreich*. Fortschritte in der Umweltmedizin. Landsberg am Lech: Ecomed Verlagsgesellschaft, 2002.
27. Baysson H, Tirmarche M, Tymen G et al. Indoor radon and lung cancer in France. *Epidemiology* 2004;15:709–16.
28. Bochicchio F, Forastiere F, Farchi S et al. Residential radon exposure, diet and lung cancer: a case–control study in a Mediterranean region. *Int J Cancer* 2005;114:983–91.
29. Wichmann HE, Schaffrath Rosario A, Heid IM et al. Lung cancer risk due to radon in dwellings—evaluation of the epidemiological knowledge. *Int Congr Ser* 2005;1276:54–7.
30. Sandler DP, Weinberg CR, Shore DL et al. Indoor radon and lung cancer risk in Connecticut and Utah. *J Toxicol Environ Health A* 2006;69:633–54.
31. Wilcox HB, Al-Zoughool M, Garner MJ et al. Case control study of radon and lung cancer in New Jersey. *Radiat Prot Dosimetry* 2008;128:169–79.
32. Thompson RE, Nelson DF, Popkin JH et al. Case–control study of lung cancer risk from residential radon exposure in Worcester County, Massachusetts. *Health Phys* 2008;94:228–41.
33. Turner MC, Krewski D, Chen Y et al. Radon and lung cancer in the American Cancer Society Cohort. *Cancer Epidemiol Biomarkers Prev* 2011;20:438–48.
34. Bräuner EV, Andersen C, Sørensen M et al. Residential radon and lung cancer incidence in a Danish cohort. *Environ Res* 2012;118:130–6.
35. Barros-Dios JM, Ruano-Ravina A, Perez-Rios M et al. Residential radon exposure, histologic types, and lung cancer risk. A case–control study in Galicia, Spain. *Cancer Epidemiol Biomarkers Prev* 2012;21:951–8.
36. Hystad P, Brauer M, Demers PA et al. Geographic variation in radon and associated lung cancer risk in Canada. *Can J Public Health* 2014;105:e4–10.
37. Torres-Durán M, Ruano-Ravina A, Parente-Lamelas I et al. Lung cancer in never-smokers: a case–control study in a radon-prone area (Galicia, Spain). *Eur Respir J* 2014;44:994–1001.
38. Lubin JH, Boice JD. Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies. *J Natl Cancer Inst* 1997;89:49–57.
39. Lubin JH, Wang ZY, Boice JD et al. Risk of lung cancer and residential radon in China: pooled results of two studies. *Int J Cancer* 2004;109:132–7.
40. Darby S, Hill D, Auvinen A et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case–control studies. *BMJ* 2004;330:223–6.
41. Pavia M, Bianco A, Pileggi C et al. Meta-analysis of residential exposure to radon gas and lung cancer. *Bull World Health Organ* 2003;81:732–8.
42. Zhang ZL, Sun J, Dong JY et al. Residential radon and lung cancer risk: an updated meta-analysis of case–control studies. *Asian Pac J Cancer Prev* 2012;13:2459–65.
43. Cohen BL. The linear no-threshold theory of radiation carcinogenesis should be rejected. *J Am Physicians Surg* 2008;13:70–6.
44. United Nations Scientific Committee on the Effects of Atomic Radiation. Sources-to-effects assessment for radon in homes and workplaces. *UNSCEAR Report 2006, Vol. II, Annex E*. Table 23 on p. 291 and Table 25 on p. 296, 2006.
45. Kendall GM, Smith TJ. Doses to organs and tissues from radon and its decay products. *J Radiol Prot* 2002;22:389–406.
46. Sivia DS, Skilling J. *Data Analysis—A Bayesian Tutorial*, 2nd edn. Oxford: Oxford University Press, 2006.
47. Fornalski KW, Dobrzyński L. The robust Bayesian approach to the model selection algorithm. *Research & Reviews: Journal of Statistics and Mathematical Sciences* 2015;1:8–12.
48. Cantrell CA. Technical Note: Review of methods for linear-squares fitting of data and application to atmospheric chemistry problems. *Atoms Chem Phys* 2008;8:5477–87.
49. York D, Evensen NM, López Martínez M et al. Unified equations for the slope, intercept, and standard errors of the best straight line. *Am J Phys* 2004;72:367–75.
50. Doss M. Linear no-threshold model vs. radiation hormesis. *Dose Response* 2013;11:480–97.
51. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ* 1998;316:989–91.
52. Scott B. Avoiding diagnostic imaging, not low dose radiation, is the real health risk. *J Am Physicians Surg* 2016;21:74–80.
53. Puskin JS. Smoking as a confounder in ecologic correlations of cancer mortality rates with average county radon levels. *Health Phys* 2003;84:526–32.

54. Heath CW Jr, Bond PD, Hoel DG et al. Residential radon exposure and lung cancer risk: commentary on Cohen's county-based study. *Health Phys* 2004;87:647–55.
55. Cohen BL. Test of the Linear-No-Threshold Theory: rationale for procedures. *Dose Response* 2005;3:369–90.
56. Scott B. Residential radon appears to prevent lung cancer. *Dose Response* 2011;9:444–64.
57. Becker K. Health effects of high radon environments in Central Europe: another test for the LNT hypothesis? *Nonlinearity Biol Toxicol Med* 2003;1:3–35.
58. Henriksen T. Radon, lung cancer and the LNT model. Research and review BMF-group, UiO. *Internal Report 2016*. Department of Physics, The Faculty of Mathematics and Natural Sciences, University of Oslo, www.mn.uio.no/fysikk/tjenester/kunnskap/straling/radon-and-lung-cancer.pdf (25 May 2017, date last accessed).
59. Seiler FA, Alvarez JL. Is the 'ecological fallacy' a fallacy? *Hum Ecol Risk Assess* 2000;6:921–41.
60. Hart J. On ecological studies: a short communication. *Dose Response* 2011;9:497–501.
61. Krewski D, Lubin JH, Zielinski JM et al. Residential radon and risk of lung cancer a combined analysis of 7 North American case-control studies. *Epidemiology* 2005;16:137–45.
62. Darby S, Hill D, Deo H et al. Residential radon and lung cancer—detailed results of a collaborative analysis of individual data on 7148 persons with lung cancer and 14,208 persons without lung cancer from 13 epidemiologic studies in Europe. *Scand J Work Environ Health* 2006;32:1–83.
63. Fornalski KW, Adams R, Allison W et al. The assumption of radon-induced cancer risk. *Cancer Causes Control* 2015;26:1517–8.
64. Sachs B, Meyerson G, Siegel JA. Epidemiology without biology: false paradigms, unfounded assumptions, and specious statistics in radiation science (with commentaries by Inge Schmitz-Feuerhake and Christopher Busby and a reply by the authors). *Biol Theory* 2016;11:69–101.
65. Wikipedia. Occam's razor. https://en.wikipedia.org/wiki/Occam%27s_razor (25 May 2017, date last accessed).
66. Sanders CL. *Radiation Hormesis and the Linear-No-Threshold Assumption*. Springer: Heidelberg, 2010.
67. Conrady J, Martin K, Poffijin A et al. High residential radon health effects in Saxony (Schneeberg Study). *Report (Contract F 14P-CT95-0027)*, European Commission, DG XII, Nuclear Safety Program, 1999.
68. Cuttler JM, Sanders CL. Threshold for radon-induced lung cancer from inhaled plutonium data. *Dose Response* 2015;13:1–4.
69. Scott BR. Residential radon appears to prevent lung cancer. *Dose Response* 2011;9:444–64.