

RESEARCH

Open Access



# Low dose radiation risks for women surviving the a-bombs in Japan: generalized additive model

Greg Dropkin 

## Abstract

**Background:** Analyses of cancer mortality and incidence in Japanese A-bomb survivors have been used to estimate radiation risks, which are generally higher for women. Relative Risk (RR) is usually modelled as a linear function of dose. Extrapolation from data including high doses predicts small risks at low doses. Generalized Additive Models (GAMs) are flexible methods for modelling non-linear behaviour.

**Methods:** GAMs are applied to cancer incidence in female low dose subcohorts, using anonymous public data for the 1958 – 1998 Life Span Study, to test for linearity, explore interactions, adjust for the skewed dose distribution, examine significance below 100 mGy, and estimate risks at 10 mGy.

**Results:** For all solid cancer incidence, RR estimated from 0 – 100 mGy and 0 – 20 mGy subcohorts is significantly raised. The response tapers above 150 mGy. At low doses, RR increases with age-at-exposure and decreases with time-since-exposure, the preferred covariate. Using the empirical cumulative distribution of dose improves model fit, and capacity to detect non-linear responses. RR is elevated over wide ranges of covariate values. Results are stable under simulation, or when removing exceptional data cells, or adjusting neutron RBE. Estimates of Excess RR at 10 mGy using the cumulative dose distribution are 10 – 45 times higher than extrapolations from a linear model fitted to the full cohort. Below 100 mGy, quasipoisson models find significant effects for all solid, squamous, uterus, corpus, and thyroid cancers, and for respiratory cancers when age-at-exposure > 35 yrs. Results for the thyroid are compatible with studies of children treated for *tinea capitis*, and Chernobyl survivors. Results for the uterus are compatible with studies of UK nuclear workers and the Techa River cohort.

**Conclusion:** Non-linear models find large, significant cancer risks for Japanese women exposed to low dose radiation from the atomic bombings. The risks should be reflected in protection standards.

**Keywords:** Low dose radiation, Cancer, Women, Thyroid, Uterus, A-bomb, GAM

## Background

Analyses of cancer mortality and incidence in Japanese A-bomb survivors continue to play a key role in establishing radiation risk estimates. The 1958–1998 incidence cohort was analysed by Preston and co-workers [1], whose main model assumes a linear dose response. Their study found some evidence of radiation risk at doses below 150 mGy, and higher risks for women. Previously Pierce and Preston [2] found a significant dose response for all solid cancer incidence in the 0 – 100 mGy range. However,

according to the EPA Blue Book [3] “it has not been possible to demonstrate and quantify risk in the Lifespan Study (of A-bomb survivors) at doses below about 100 mGy”. The ICRP [4] recommended annual dose limit for occupational exposures is 20 mSv (whole body dose).

Some cancer studies find non-linear responses to radiation [5–8] and there are some significant results at low doses [9–16]. Low dose radiation risk was reviewed in 2003 [17] and remains a focus of research [18].

I consider cancer incidence in female Japanese A-bomb survivors in the dose ranges 0 – 500 mGy, 0 – 100 mGy, and 0 – 20 mGy, using the grouped anonymous (public) incidence data for the 1958 – 1998 cohort.

Correspondence: gregd@gn.apc.org  
Independent researcher, Liverpool, England

Attained age is the single most powerful predictor of cancer incidence. Models can be refined using birth cohort, age-at-exposure or time-since-exposure, city, urban/rural residence, and dose. The Not In City (NIC) group of Hiroshima and Nagasaki residents who were out of the area at the time of bombing can be included.

The analysis uses the Generalized Additive Model [19] implemented in R [20] through the programme *mgcv* [21]. GAM is a flexible technique for fitting non-linear responses and covariate interactions. Models are investigated for all solid cancers, and then applied to the most common anatomical or histological groups and specific sites within the female incidence data: digestive, genital, respiratory, adenocarcinoma, other epithelial, squamous, stomach, colon, liver, pancreas, gallbladder, rectum, uterine, cervical, uterine corpus, ovarian, lung, breast, thyroid, and central nervous system.

The analysis aims to test for linearity of the dose response, explore the interaction of dose with attained age, age-at-exposure and time-since-exposure, adjust for the highly skewed distribution of dose, test whether the response is significant below 100 mGy, and estimate the risks at 10 mGy.

## Methods

The public dataset *LSSinc07* [22] was downloaded from the Radiation Effects Research Foundation (RERF) [23]. Records with unknown dose were excluded, leaving 25,570 cells with grouped data on 17,448 solid cancers among 105,427 survivors with 2,764,735 person-years of follow-up. The covariate **py** denotes person-years observation in each cell. **logage** is the log of **age**, the attained-age covariate which RERF obtained by weighting each individual contribution by its p-y within the cell. Likewise **agem** is the p-y weighted mean age-at-exposure, **since** is the difference between **year**, the p-y weighted mean calendar time, and the exposure date (depending on city), and **birth** is defined as year - age. The dataset also contains sex, city, and **distcat**, the latter taking the values 1 (urban), 2 (rural), and 3 (Not in City). The grouped data is stratified by these (and other) categories. The dataset is then restricted by sex = 2, and by **distcat** < 3 if the NIC group is excluded; city and **distcat** are converted to factor variables **fc** and **fd**. For convenience, covariate names are introduced in bold, but then used in normal font.

When modelling all solid cancers, the instantaneous flash dose is the p-y weighted mean colon dose in mGy, defined as **coldose** = 1000\*cola02w10 using the DS02 dosimetry and a Relative Biological Effectiveness (for neutrons) of 10, unless otherwise specified. The dataset is then restricted by **coldose** < 100 (or < 20 or < 500). For convenience  $A^{+}$ ,  $B^{+}$  and  $C^{+}$  refer to 0 –

20 mGy, 0 – 100 mGy and 0 – 500 mGy subcohorts including or excluding NIC. Weighting each coldose by the py of its data cell, the mean coldose in  $A^{+}$ ,  $B^{+}$  and  $C^{+}$  is 2.43, 3.70, 9.84, 13.85, 36.44, 48.93 mGy respectively.

More generally, modelling uses the appropriate organ dose as in [1] and *LSSinc07*. Tissue weighting factors are not applied. When considering a cancer group or site, **dose** refers to the relevant p-y weighted organ dose in mGy, and the response variable **obs** is the number of incident cases of that cancer in each data cell. The case numbers by site and dose range are shown in Table 1. Note that  $A^{+}$  etc. are defined from coldose, no matter which cancer site is under consideration.

Modelling seeks to predict the cancer risk, i.e. the expected value of obs/py, as a function of covariates and their interactions. Each data cell is regarded as an independent source of Poisson distributed incident cases, and penalised regression with cubic splines, using a log link, is used to estimate the expected value of obs. Briefly (see [21]):

$obs_j \sim \text{Pois}(\mu_j)$  where j indexes the data cells

$\log(\mu_j) = \sum_i \beta_i f_i(x_j)$  where i indexes the model components f and coefficients  $\beta$ ;  $f_i$  is a function of covariates x, whose form is specified in the model but whose value varies with  $x_j$  in cell j; and  $\beta_i$  is a numerical coefficient of  $f_i$ , independent of j, to be estimated during model fitting.

The  $f_i$  can include offsets, whose  $\beta_i$  is fixed at 1; factors; parametric terms; splines derived from the covariates; and tensor products of such splines.

A cubic regression spline with components  $f_0, f_1, \dots, f_{k-1}$  is obtained from covariate x by placing k knots, for example at equally spaced quantiles of x, forming the natural cubic splines with value 1 at the specified knot and 0 at the other knots, using a centering constraint to achieve  $\sum_j f_i(x_j) = 0$  when  $i > 0$ , and setting  $f_0 = 1$ .

During model fitting, the  $\beta_i$  are estimated by minimising the penalised deviance,  $\text{Dev} + \beta^T S \beta$ , where the positive semi-definite penalty matrix S is a direct sum of blocks generated from the splines, each block incorporating a smoothing parameter  $\lambda_b$ . For tensor product splines, there are separate smoothing parameters for each margin. Dev is the Deviance, given in the Poisson case (omitting subscripts j) by

$-2 \sum [\text{obs} \cdot \log(\mu) - \mu - I_{\text{obs}} \cdot (\text{obs} \cdot \log(\text{obs}) - \text{obs})]$   
where  $I_{\text{obs}} = 1$  if  $\text{obs} > 0$  and 0 otherwise

Minimising the penalised deviance is a compromise between approximating the data (lowering Dev), and the “wiggleness” of the fitted model as measured by the penalty. Higher values of the smoothing parameter(s) impose stronger penalties, making the fitted model smoother with lower “effective degrees of freedom” (a penalised version of the number of independent param-

**Table 1** Case numbers by site and dose range

Site	A <sup>-</sup>	A <sup>+</sup>	B <sup>-</sup>	B <sup>+</sup>	C <sup>-</sup>	C <sup>+</sup>	Mid age	Mid agex	Mid since
Solid	4239	6271	5474	7506	6712	8744	68.4	33.1	36.9
Adeno	2477	3722	3188	4433	3928	5173	67.0	28.3	38.6
Othepe	1016	1460	1310	1754	1583	2027	73.3	38.4	32.7
Squam	550	823	718	991	881	1154	63.5	33.2	28.6
Digest	2206	3263	2825	3882	3418	4475	71.7	33.5	38.6
Genital	660	1008	871	1219	1035	1383	61.8	31.6	28.6
Resp	324	487	441	604	558	721	73.2	36.5	38.6
Stom	997	1459	1285	1747	1564	2026	71.6	36.6	33.6
Colon	373	578	471	676	570	775	71.7	29.3	44.1
Liver	287	416	356	485	433	562	71.7	32.6	41.4
Gall	181	266	234	319	273	358	73.2	36.8	38.6
Panc	134	192	172	230	208	266	73.4	36.6	38.6
Rect	183	283	245	345	291	391	68.5	31.6	41.3
Uterus	519	806	689	976	821	1108	61.7	31.5	28.6
Cervix	384	598	509	723	603	817	61.8	31.7	28.5
Corpus	84	132	108	156	126	174	61.1	23.1	38.5
Ovary	115	170	143	198	171	226	63.5	31.6	33.6
Breast	441	667	554	780	713	939	59.1	22.4	38.6
Lung	290	437	396	543	503	650	73.3	36.4	38.6
Thyroid	158	219	203	264	271	332	66.7	28.2	31.9
Cns	93	117	125	149	151	175	66.2	26.7	37.0

Mid age (agex, since) is the age (agex, since) below which 50% of cases occur. Values are shown for B<sup>+</sup>, and are similar for other dose ranges

ters), but allowing it to stray further from the individual data points. As  $\lambda_b \rightarrow \infty$ , the model becomes log-linear (only the linear components of the spline terms are unpenalised by S).

Model fitting involves optimising the smoothing parameters affecting S, through a criterion such as Generalized Cross Validation, Restricted Maximum Likelihood, or Marginal Likelihood (see [21, 24] for details). After fitting, inference is based on the Bayesian posterior covariance of  $\beta$ .

With the fitted  $\hat{\beta}_i$ , the expected value of  $\text{obs}_j$  is  $\mu_j = \exp(\sum_i \hat{\beta}_i f_i(x_j))$ . The fitted model also predicts the expected value  $\mu_{j,0}$  if the dose covariate is reset to 0 while preserving all other covariates and the  $\hat{\beta}_i$ . In each cell, baseline risk is estimated internally from the fitted model as  $\mu_{j,0} / \text{py}_j$  and Relative Risk is  $\text{RR}_j = \mu_j / \mu_{j,0}$ .

From here on, models are defined and fitted using the R programme mgcv. For example, the initial Poisson model is specified as

```
P0 <- gam(obs ~ offset(log(py)) + (fc + fd)^2
+ ti(logage, k = 10) + ti(birth, k = 10)
+ ti(logage, birth, k = 10),
family = "poisson", method = "ML")
```

The reverse arrow <- defines the model by the subsequent expression, with gam denoting a generalized additive model. Here, obs is regarded as a sample from a Poisson distribution whose mean, varying across the data cells, is specified with a (default) log link. The offset term, with coefficient fixed at 1, ensures that the fitted value of  $\text{obs}/\text{py}$  is independent of  $\text{py}$ . The factor variables and their first order interaction are included as  $(\text{fc} + \text{fd})^2$ . The smooth term  $\text{ti}(\text{logage})$  refers to an unknown linear combination of splines derived from  $\text{logage}$ , and the tensor product term  $\text{ti}(\text{logage}, \text{birth})$  allows for interaction of  $\text{logage}$  with  $\text{birth}$ . In each smooth, the basis dimension is set at 10. By default, "cubic regression splines" are used with the  $\text{ti}$  terms. The smoothing parameter optimisation method indicated by "ML" is Marginal Likelihood. For details see [21] and ?gam and ?ti (within mgcv). Setting family = "quasipoisson" allows a mean-variance relation  $V = \mu\phi$  where V is the variance about mean  $\mu$  and  $\phi > 0$  is a scale factor. For the Poisson family,  $\phi = 1$ , but any positive scale can be set, or estimated as below.

The dose covariate is highly skewed and, in the grouped data, multimodal. Over 70% of  $\text{py}$  in the full data has dose < 20 mGy, or over 60% if NIC cells are excluded. Likewise over 84% has dose < 100 mGy, or 79% if NIC cells are

excluded. The grouped data cells are also stratified by dose category, an analytical decision embodied in the anonymous public data. For example, each cell in A<sup>-</sup> or A<sup>+</sup> will contain records with dose either below 5 mGy or between 5 and 20 mGy.

To adjust for this irregular distribution, when dose = D define **ecdos** as the proportion of cells with dose ≤ D, a monotone increasing function of dose with an approximately uniform distribution. It increases quickly where dose is concentrated, remaining static where dose is sparse. While max(**ecdos**)=1, min(**ecdos**) > 0 due to cells with dose = 0. Formally, **ecdos** is the empirical cumulative distribution function of dose, obtained in R as `ecdf(dose)(dose)`, with inverse transformation `quantile(dose)`. Figure 1 shows **ecdos** and dose in B<sup>+</sup>.

The initial model P0 (and its quasipoisson variant Q0) are enlarged with terms using **ecdos** or dose and interactions, as shown in Table 2.

The nested chains P0 ⊂ P2e ⊂ P3e ⊂ P5ae, P0 ⊂ P2e ⊂ P4ae ⊂ P5ae, P0 ⊂ P2e ⊂ P3e ⊂ P5se, and P0 ⊂ P2e ⊂ P4se ⊂ P5se give models of increasing complexity using **ecdos** with age<sub>x</sub> or since<sub>x</sub> respectively. Corresponding nested chains P0 ⊂ P2d etc. are defined using dose instead of **ecdos**. P2 refers to the pair P2e and P2d, etc.

The scale for a quasipoisson model M is estimated iteratively via Pearson  $\chi^2 = \sum (obs - fit)^2 / fit$ , where fit is the fitted value, by seeking  $\hat{\phi} \simeq \chi^2 / dfres$  where dfres is the residual degrees of freedom (# of records – effective degrees of freedom).

Additional file 1: Appendix A describes methods for Relative Risk RR, confidence intervals based on the Bayesian posterior distribution of  $\beta$ , dose response curves, and model selection. Simulation tests used to choose the preferred method of smoothing parameter optimisation and to validate the models are outlined in Additional file 1: Appendix B.

Numerical results are given at 2.5 and 10 mGy in A<sup>+</sup>, 2.5, 10, 20, 40, 60, 80 mGy in B<sup>+</sup>, and at 2.5, 10, 100, 200, 300, and 400 mGy in C<sup>+</sup>. Confidence intervals are 90% (lower and upper 95% confidence limits LCL95% and UCL95%) unless otherwise specified, and “borderline” denotes  $0.98 \leq LCL95\% \leq 1$ .

Initially, no minimum value of the smoothing parameters was imposed. However, if the visual output was “too wiggly”, indicating overfitting, min.sp was increased until the output appeared fairly smooth, a subjective judgement.

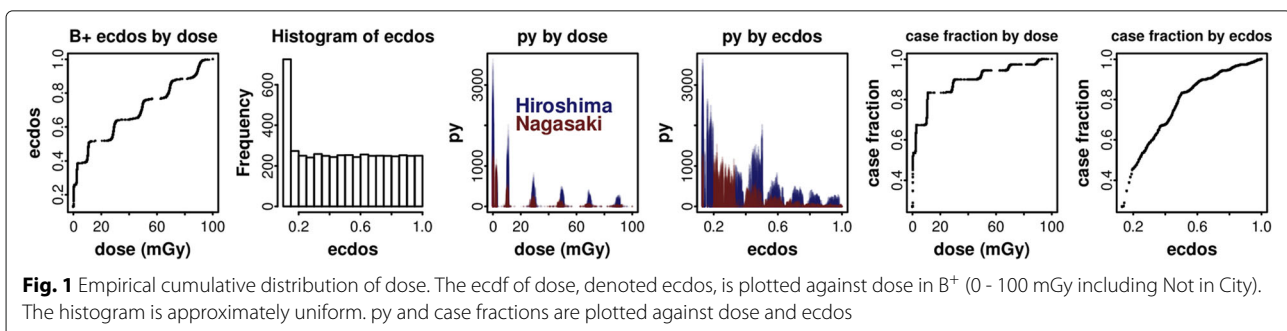
In each data cell, a fitted model gives predicted values  $\mu$ , with  $\sum \mu = \sum obs$ , and predicts values  $\mu_0$  if dose had been reset to 0, or **ecdos** reset to `ecdf(dose)(0)`. The attributable fraction  $1 - \sum \mu_0 / \sum \mu$  is the predicted number of additional cases above the baseline, as a proportion of those observed.  $\sum \mu / \sum \mu_0$  estimates the Standardized Incidence Ratio from the fitted model.

The age distribution of cases varies between cancer sites. Define the “mid” age (and age<sub>x</sub>, since) for a site and dose range as that below which half the cases occur. Table 1 includes the mid values by sites in B<sup>+</sup>; values for other dose ranges are similar. Results are reported at the mid age rounded to the nearest 5 years, and where appropriate at the rounded mid age<sub>x</sub> or since<sub>x</sub>.

For a fitted model M, set  $I_M = 1$  on any data cell for which  $LCL95\% > 1$ , and 0 otherwise. The proportion of obs (or py) for which  $I_M = 1$  measures the extent to which  $RR > 1$  is significant across the dose range. Since  $I_M = 0$  if dose = 0, the proportions are obtained for  $D_{0.5}$ , those cells with dose > 0, or for  $D_{0.5}$  (dose > 0.5 mGy). Corresponding proportions are obtained on  $S_{0.5}$  (since < 35 and dose > 0.5 mGy), or  $T_{0.5}$  (age<sub>x</sub> > 35 and dose > 0.5 mGy), or  $S_{0.5} \cap T_{0.5}$ . Likewise, the indicator  $J_M = 1$  on any data cell for which  $UCL95\% < 1$ , and 0 otherwise. The proportion of obs (or py) for which  $J_M = 1$  measures the extent to which  $RR < 1$  is significant across the dose range.

Selected models for all solid cancers were re-fitted after excluding exceptional data points as detected by Cook’s Distance, or with thin plate regression splines, or for urban residents, or with RBE as obtained in [25].

For a fitted model M, influence(M) is a vector of leverage at each cell, and plotting influence(M1)/influence(M2) against dose compares model sensitivities in different dose regions. The ability of **ecdos** and dose models to detect hypothetical dose-responses was compared by simulation (Appendix B). Further simulations considered the uncertainty in dose.



**Table 2** Model definitions

Model	Family	$\phi$	Model	Family	$\phi$	Terms
P0	poisson	1	Q0	quasipoisson	> 0	offset(log(py))+(fd+fc)^2+ti(logage) +ti(birth)+ti(logage,birth)
P1e	poisson	1	Q1e	quasipoisson	> 0	+.eccdos
P2e	poisson	1	Q2e	quasipoisson	> 0	+.ti(eccdos)
P3e	poisson	1	Q3e	quasipoisson	> 0	+.ti(eccdos)+ti(eccdos,logage)
P4ae	poisson	1	Q4ae	quasipoisson	> 0	+.ti(eccdos)+ti(eccdos,agex)
P4se	poisson	1	Q4se	quasipoisson	> 0	+.ti(eccdos)+ti(eccdos,since)
P5ae	poisson	1	Q5ae	quasipoisson	> 0	+.ti(eccdos)+ti(eccdos,logage) +ti(agex)+ti(eccdos,agex)
P5se	poisson	1	Q5se	quasipoisson	> 0	+.ti(eccdos)+ti(eccdos,logage) +ti(since)+ti(eccdos,since)
P1d	poisson	1	Q1d	quasipoisson	> 0	+.dose
P2d	poisson	1	Q2d	quasipoisson	> 0	+.ti(dose)
P3d	poisson	1	Q3d	quasipoisson	> 0	+.ti(dose)+ti(dose,logage)
P4ad	poisson	1	Q4ad	quasipoisson	> 0	+.ti(dose)+ti(dose,agex)
P4sd	poisson	1	Q4sd	quasipoisson	> 0	+.ti(dose)+ti(dose,since)
P5ad	poisson	1	Q5ad	quasipoisson	> 0	+.ti(dose)+ti(dose,logage) +ti(agex)+ti(dose,agex)
P5sd	poisson	1	Q5sd	quasipoisson	> 0	+.ti(dose)+ti(dose,logage) +ti(since)+ti(dose,since)

Models use cubic regression (cr) splines, with basis dimension 10

Model fitting by method ML

“+.”: augment the initial model P0 (or Q0) by the terms which follow

ti refers to a smooth term on the enclosed covariate(s)

These intensive computations were applied only for all solid cancers. Elsewhere, the Bayesian posterior CIs (Appendix A) are reported without further validation. Some codes with commentary are shown as additional files.

## Results

For all solid cancers, no minimum smoothing parameters were imposed. Basis dimension  $k=10$  was sufficient when tested by `gam.check` or modelling residuals against covariates. The ML method of smoothing parameter optimisation was preferred to alternatives, giving 95% CIs which performed well under simulation tests described in Appendix B and shown in Fig. 2. ML is discussed in [24].

P5ae, P5se, P5ad and P5sd were cross-validated (Appendix B) in  $A^{+-}$ , to test the proportion of deviance explained and the estimated  $RR_{|10\text{ mGy}}$  with 90% CIs, see Additional file 1: Table S1.

Additional file 1: Figure S1 shows the dose response in  $B^+$ . All plots show RR significantly raised, although the response is lower with the dose models. Plots with REML or GCV.Cp optimisation methods are similar. For the eccdos models, the appearance of the response reflects the relation between eccdos and dose (see Fig. 1); corresponding curves of RR against eccdos are regular.

Figure 3 compares the responses of P5se and P5sd in  $A^+$ ,  $B^+$  and  $C^+$ , evaluated at age 70 and since 35. Curves are similar over  $A^-$ ,  $B^-$  and  $C^-$ , and for P5ae and P5ad at age 70 and agex 35, with slightly weakened significance.

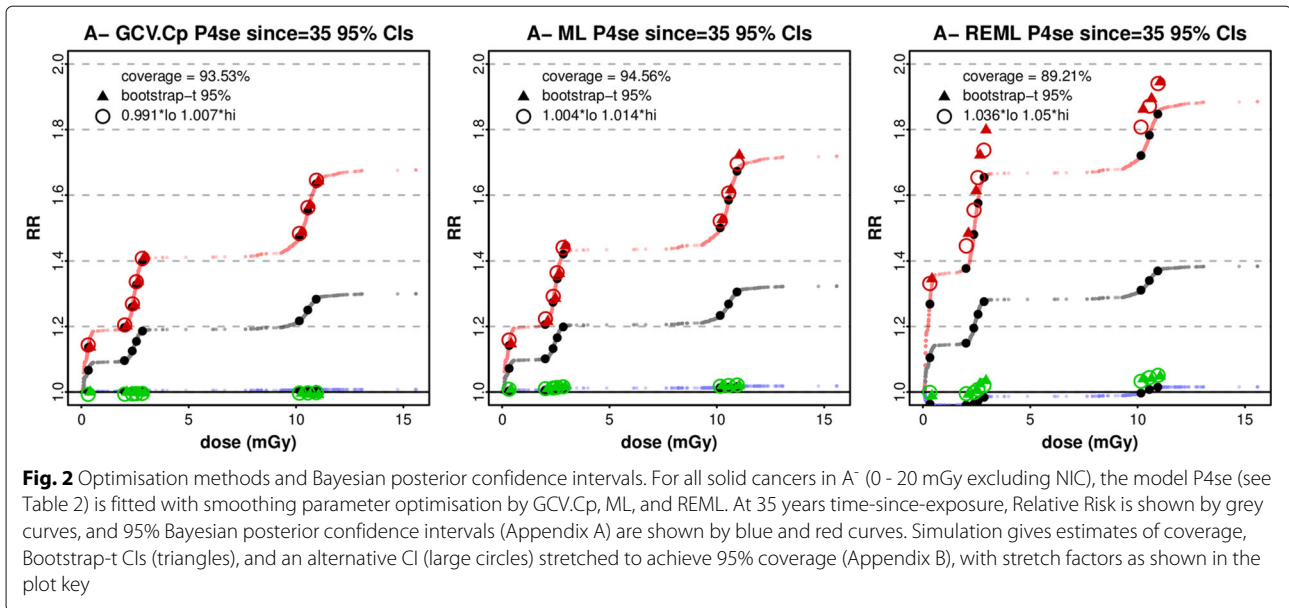
Data simulated over  $B^-$  from a variety of hypothetical forms of the true dose-response curve, without interactions, was fitted with P2e and P2d (Appendix B). Table 3 has results for 10 curves, comparing models by the mean coverage of their respective CIs across the dose range. Models with eccdos give better or equivalent coverage, unless the true dose-response is linear.

Figure 4 displays the simulated geometric mean fitted values and CIs for two of these curves (R1 and C2), showing that the dose model underestimates the true response while the eccdos model captures it correctly.

The increased sensitivity of eccdos models at low doses is seen in Additional file 1: Figure S2, plotting the ratio of influence with P5sd to that with P5se, against dose. In  $B^{+-}$  and  $C^{+-}$  low doses have less influence on P5sd, reducing its sensitivity to their effects.

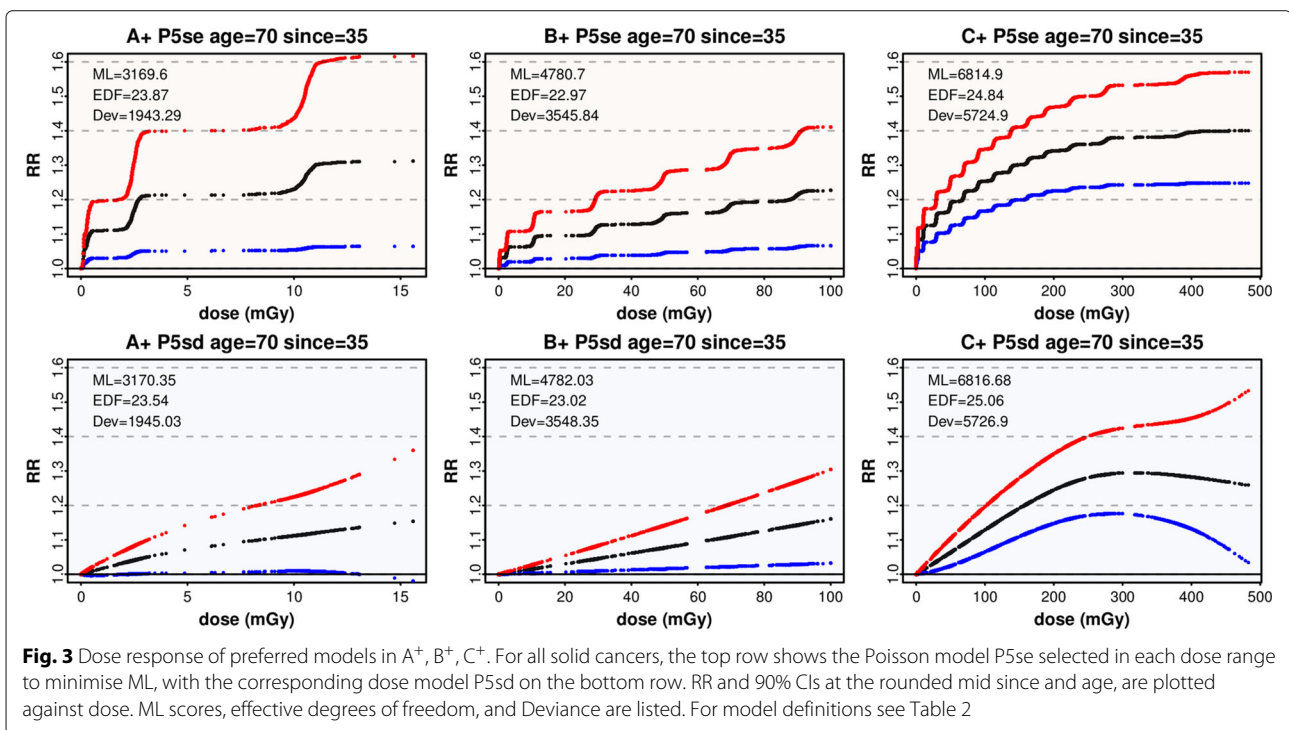
Splines constructed from eccdos have better variability than dose splines at very low doses, lowering the internal baseline estimates. Additional file 1: Figure S3 compares splines from P2d, P2e, and P2g constructed from gdos, a smooth approximation to eccdos, when modelling over  $B^+$ . Although the P2g response curve is simpler, the parameters of gdos vary with dose range and its distribution is multimodal, whereas eccdos has a universal definition and an approximately uniform distribution.

Interactions in  $B^+$  are shown by perspective plots in Fig. 5, fitting P5a and P5s with either eccdos or dose, and fixing age = 70 or agex = 35 or since = 35. Figure 6 shows  $RR_{|10\text{ mGy}}$  and 90% CIs at age 70, using P5a and P5s fitted



over A<sup>+</sup>, B<sup>+</sup> and C<sup>+</sup>. With P5a at age 70 RR increases with age at all doses in A<sup>+</sup>, B<sup>+</sup> or C<sup>+</sup>. With P5s at age 70, RR decreases (or is constant) with since at all doses. The decrease is confined to since > 35 in C<sup>+</sup>. With P5a at age 35 RR generally decreases (or is constant) with age. With P5s at since 35 RR increases (or is constant) with age in A<sup>+</sup> and B<sup>+</sup>, but decreases with age at higher doses in C<sup>+</sup>. Poisson model selection methods (Appendix A) were applied to choose preferred models. P5se is preferred in all

six dose ranges by method 1 (lowest ML score). The three methods chose models with ecdos in 17 of 18 cases. Additional file 1: Table S2 shows the preferred Poisson models for each range and selection method. As appropriate, RR is shown at age 70, age 35, or since 35. All preferred models show RR elevated at 2.5, 10, 20, 40, 60, 80, 100, 200, 300, and 400 mGy, with LCL95% > 1, except for method 2 in B<sup>+</sup> where P3d is preferred, with LCL95% ≥ 1. In all cases Deviance < dfres. The proportion of Deviance explained





**Table 3** Coverage by P2e and P2d of true dose response

Name	ERR	$\beta$	$\sigma$	$\tau$	$\gamma$	Form	Mean (lhe)	Mean (lhd)
N1	$g(d, \beta, \sigma, \tau)$	0	0			null	95.18%	94.90%
L1	$g(d, \beta, \sigma, \tau)$	0.002	0			linear	81.46%	94.56%
H1	$f(d, \beta, \sigma, \tau)$	0.004	-0.2	0.05		hormesis	75.08%	69.86%
P1	$f(d, \beta, \sigma, \tau)$	0	0.2	0.05		plateau	93.36%	69.68%
C1	$f(d, \beta, \sigma, \tau)$	0.0012	0.08	0.5		two phase	90.88%	50.08%
C2	$f(d, \beta, \sigma, \tau)$	0.015	0.75	0.2		two phase	94.80%	13.98%
R1	$g(d, \beta, \sigma, \tau)$	0.002	0.01	0.08		two phase	93.64%	68.14%
R2	$g(d, \beta, \sigma, \tau)$	0.002	0.01	0.04		two phase	93.98%	66.10%
T1	$h(d, \gamma)$				50	threshold	88.40%	87.66%
T2	$h(d, \gamma)$				25	threshold	86.88%	86.90%

All Solid cancers, fitting over B<sup>-</sup>

ERR: true value as specified by function

$$f(d, \beta, \sigma, \tau) = \beta \cdot d + \sigma \cdot (1 - \exp(-\tau \cdot d))$$

$$g(d, \beta, \sigma, \tau) = \beta \cdot d + \sigma \cdot d \cdot \exp(-\tau \cdot d)$$

$$h(d, \gamma) = \text{ifelse}(d < \gamma, 0, 1/5 \cdot ((d - \gamma)/(100 - \gamma))^3)$$

Mean (lhe or lhd): mean coverage at 10, 20... 100 mGy of 95% CI for P2e (P2d) see also Appendix B

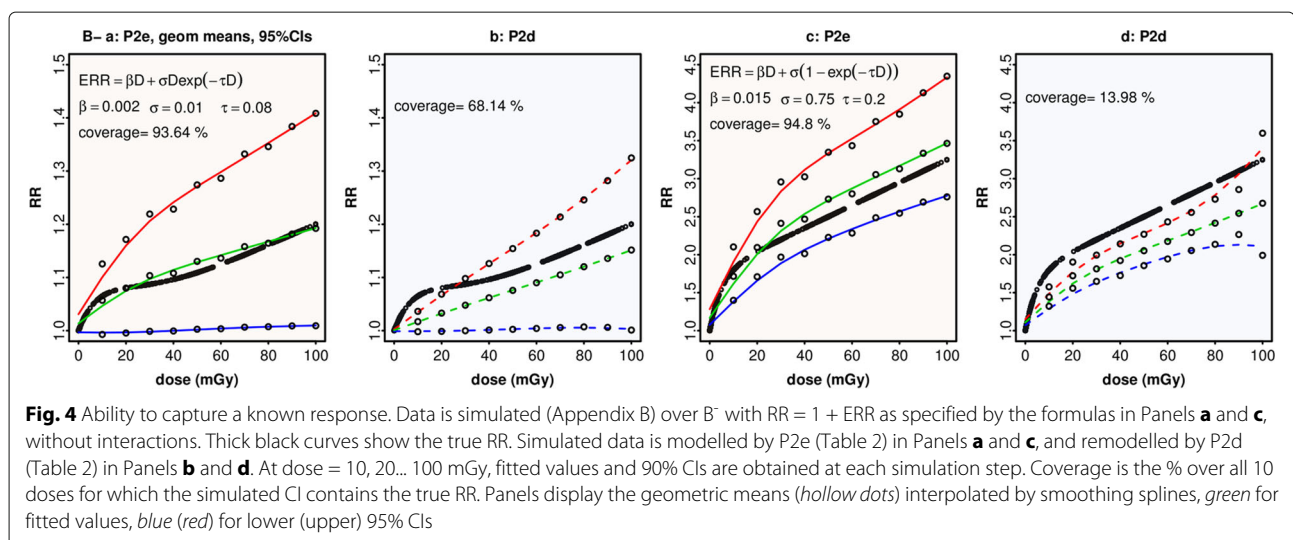
varies from 46.1% in C<sup>-</sup> to 68.1% in A<sup>+</sup>. The attributable fraction varies from 1.4% (method 2 in B<sup>+</sup>) to 9.4% in A<sup>-</sup>.

Additional file 1: Table S2 also shows the percentage of cases and py in D<sub>0</sub> for which RR > 1, LCL95% > 1, or UCL95% < 1; likewise for cells in D<sub>0.5</sub>, S<sub>0.5</sub>, T<sub>0.5</sub>, or S<sub>0.5</sub> ∩ T<sub>0.5</sub> (see “Methods”). For each of these preferred models, cell and dose ranges, RR > 1 holds for over 80% of cases and py, and UCL95% < 1 never occurs. Except with method 2 in B<sup>+</sup>, LCL95% > 1 holds for over 70% of cases in T<sub>0.5</sub>, over 75% of cases in S<sub>0.5</sub>, and over 90% of cases in S<sub>0.5</sub> ∩ T<sub>0.5</sub>.

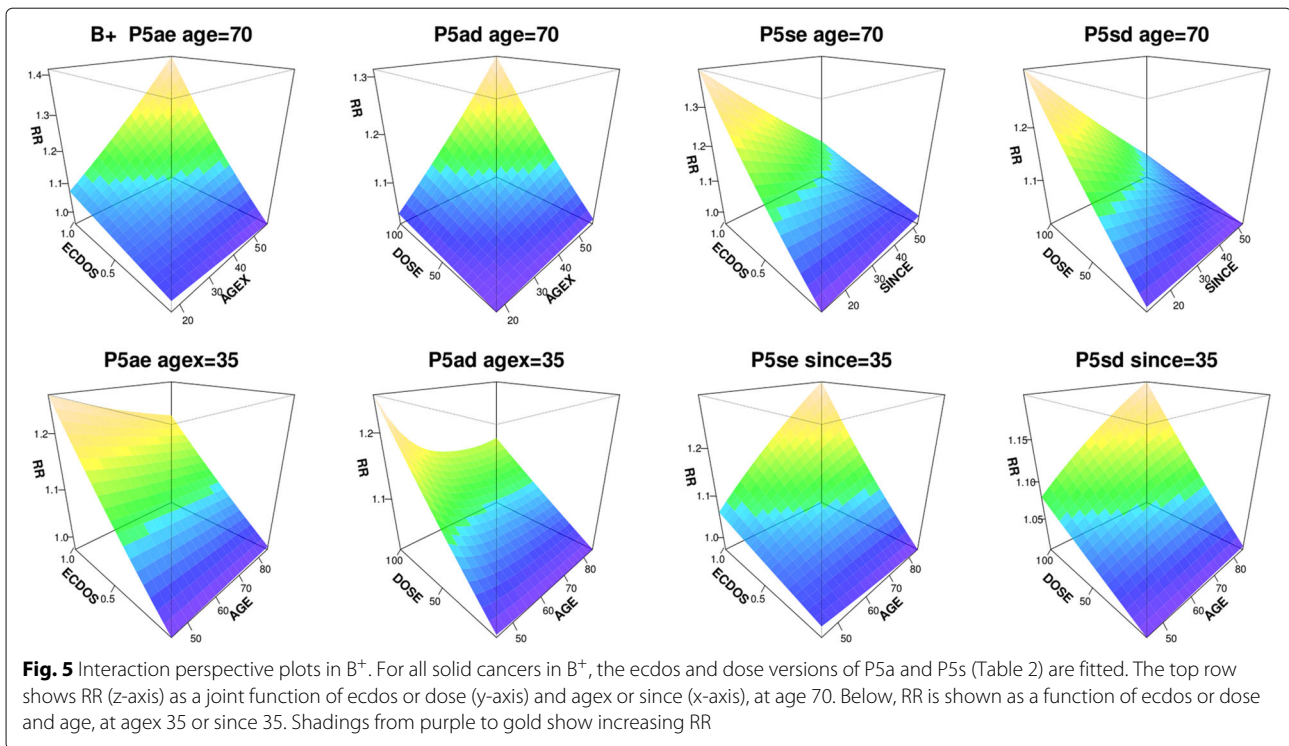
Significant elevation of RR persists over a wide range of age and since (or agex), as shown for all models in B<sup>+</sup> in Fig. 7. With the ecdos models, RR is significantly raised for regions containing age > 55, agex > 35, or since < 35. With the dose models, the regions of significance

are similar, though smaller. For the preferred (method 1) Poisson models in A<sup>+</sup>, B<sup>+</sup>, and C<sup>+</sup>, Fig. 8 shows the dose response, the dependence of RR on since and age, and the regions of significance.

The dose models themselves give non-linear responses over C<sup>+</sup> (see Fig. 3). With P2d, RR rises approximately linearly below 150 mGy, before tapering. Anova comparison with P1d, in which dose appears as a log-linear unsmoothed term, gives p = 0.015. Below 100 mGy, the Excess Relative Risk (ERR = RR - 1) is about 70% higher with P2d than with P1d. With P5sd RR tapers above 150 mGy as with P2d. The term ti(since) has e.d.f. = 3.658 and p = 0.037, and anova(P3d, P5sd) = 0.008 while anova(P4sd, P5sd) = 0.07. Thus P5sd is decidedly non-linear in dose, whose interaction with logage is borderline, but since has significant impact.



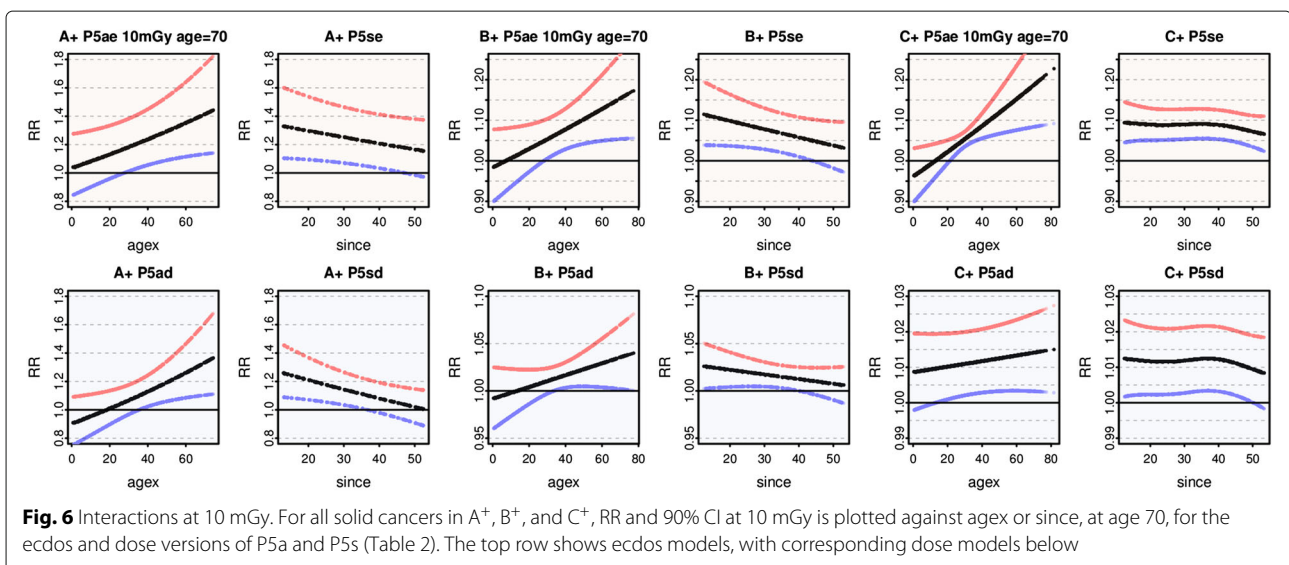
**Fig. 4** Ability to capture a known response. Data is simulated (Appendix B) over B<sup>-</sup> with RR = 1 + ERR as specified by the formulas in Panels **a** and **c**, without interactions. Thick black curves show the true RR. Simulated data is modelled by P2e (Table 2) in Panels **a** and **c**, and remodelled by P2d (Table 2) in Panels **b** and **d**. At dose = 10, 20... 100 mGy, fitted values and 90% CIs are obtained at each simulation step. Coverage is the % over all 10 doses for which the simulated CI contains the true RR. Panels display the geometric means (*hollow dots*) interpolated by smoothing splines, *green* for fitted values, *blue* (*red*) for lower (upper) 95% CIs



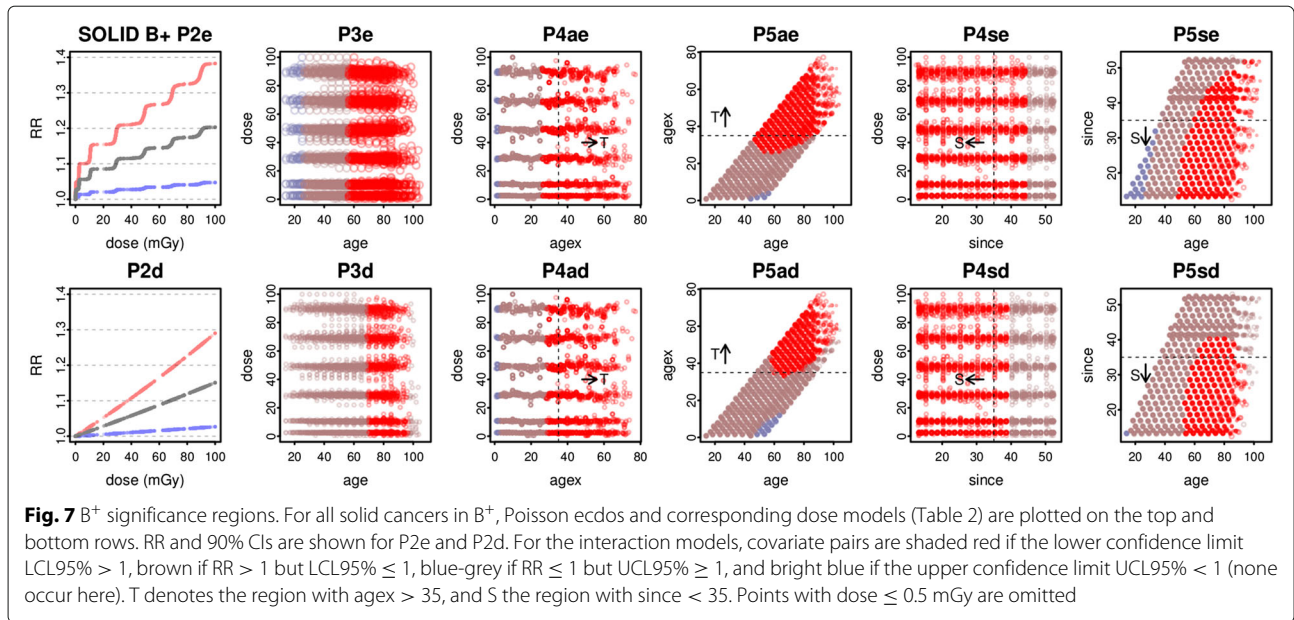
The dose models also show RR raised in A<sup>+</sup>. For P1d and P2d, in both ranges  $RR|_{10\text{mGy}} = 1.09 (0.99, 1.19)$ . P4sd is preferred amongst dose models by methods 2 and 3. At since 35, in A<sup>-</sup>  $RR|_{10\text{mGy}} = 1.11 (1.00, 1.22)$  with significant elevation when since < 35. In A<sup>+</sup>  $RR|_{10\text{mGy}} = 1.09 (0.99, 1.19)$  with significant elevation when since < 32. In A<sup>+</sup> P5sd is preferred amongst dose models by method 1. At age 70, since 35 in A<sup>-</sup>  $RR|_{10\text{mGy}} = 1.11 (1.00, 1.23)$  whilst in A<sup>+</sup>  $RR|_{10\text{mGy}} = 1.11 (1.01, 1.22)$  and RR is significantly raised for 37.2% (28.2%) of cases (py) in D<sub>0.5</sub>, 75.0% (42.3%) in S<sub>0.5</sub>, and 68.4% (79.6%) in T<sub>0.5</sub>.

In general, point estimates and CIs are primarily affected by the dose range (less by the inclusion of NIC) and the choice of dose or ecdds.

The stability of these results was tested, first by using thin plate regression splines. RR and CIs were similar to those shown in Additional file 1: Table S2, but cubic regression splines gave lower ML scores. If exceptional cells detected visually by Cook's Distance (CD) are deleted, RR values and 90% CIs were slightly altered from those in Additional file 1: Table S2, but still significantly raised (borderline for B<sup>+</sup> method 2). Deletion affected





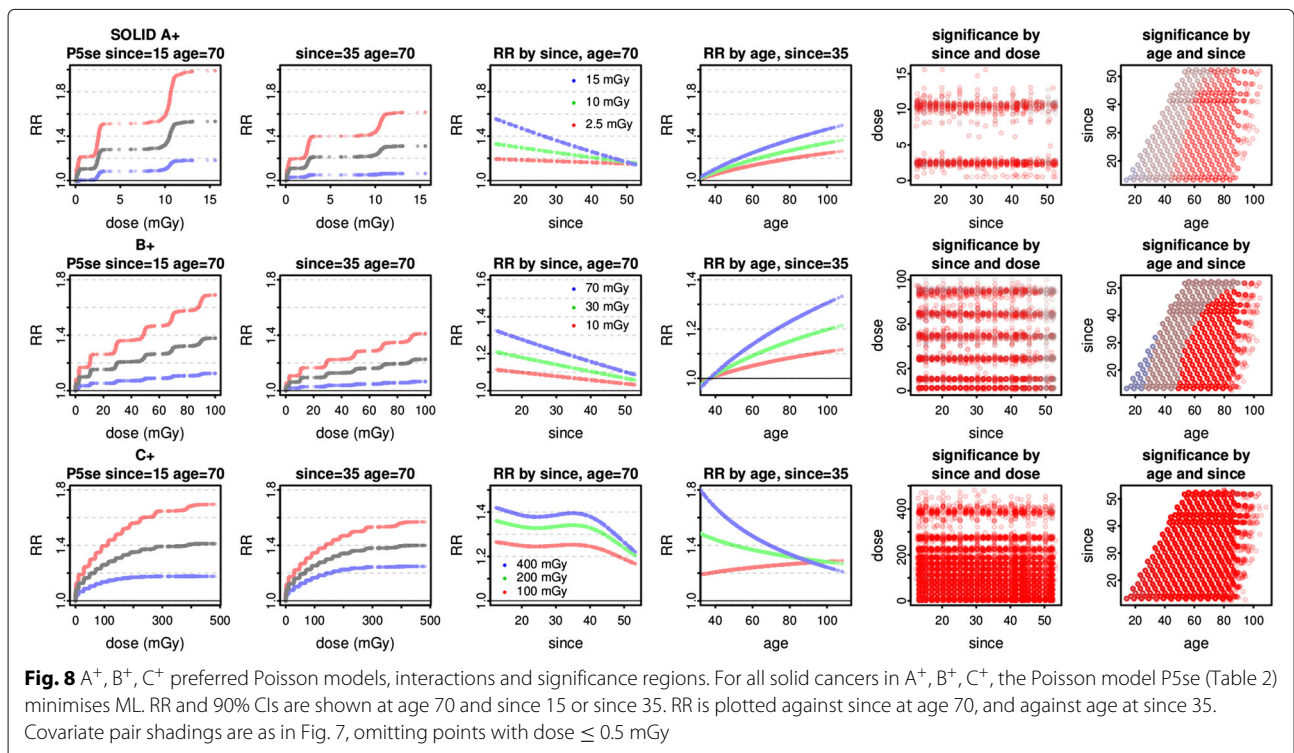


the choice of preferred models only for  $C^-$ , where P5ae was preferred by all methods, and  $C^+$ , where P5se was preferred by all methods. When preferred models were refitted after excluding cells with the highest 5% of CD values, RR was raised with at least borderline significance across each dose range.

The combined impact of sampling variability and log-normal dosimetry errors was simulated (see Appendix B)

using P5se and P5sd in  $A^-$ , shown in Additional file 1: Figure S4. Geometric mean fitted values and 90% CIs are below the values derived from the original data, but still raised with borderline significance; while bootstrap-t [26] and stretched CIs derived from the simulated data remain significantly raised.

Models were fitted with RBE as the function of neutron dose  $D_n$  shown in [25] (Fig. 4 and Eq. (2)). This raises



the dose assigned, as  $RBE > 50$  in the ranges considered here (for which  $D_n < 0.003$  Gy). Results are shown in Additional file 1: Table S3. In all dose ranges and with all three methods of Poisson selection, RR is elevated with at least borderline significance. In 15 out of 18 cases, and for 6 of 6 with method 1, the same models are selected, and RR and CIs are slightly lower but still comparable to those in Additional file 1: Table S2.

Excluding genital or, separately, respiratory cancers, the remaining solid cancers were modelled over  $A^+$ ,  $B^+$ , and  $C^+$ , with results in Additional file 1: Table S4. RR estimates are mainly similar to those in Additional file 1: Table S2. The heightened significance when  $age_x > 35$  or  $since < 35$  generally survives these exclusions.

For urban residents in A, B, or C, the preferred Poisson models show RR elevated with at least borderline significance, and a tapering response at higher doses (see Additional file 1: Table S5). With method 3, RR is significantly elevated in each range.

In each dose range, the factor variables  $fc$  and  $fd$  have significant interaction, lowering the fitted values for rural Nagasaki residents. Modelling separately by city (Additional file 1: Table S6), for Hiroshima, models with  $since$  are preferred for every dose range and selection method. None show RR significantly or borderline raised at rounded mid age and  $since$  over  $A^{++}$  or  $B^{++}$ , but over  $C^{++}$  RR is significantly raised (method 1) or borderline (at 10 mGy, methods 2 and 3) at age 70 and  $since > 35$ , and significantly raised (all methods) for high percentages of cases and  $py$ .

For Nagasaki, the preferred models show RR significantly raised in  $A^{++}$  and  $B^-$  (and for method 1 in  $B^+$ ) at rounded mid age,  $since$ , or  $age_x$ , and for high percentages of cases and  $py$ . Similar results are obtained over  $C^+$  with method 1, though exposure at younger ages appears more significant. Figure 9 shows the response by city at the respective rounded mid age and  $since$  or  $age_x$  for the preferred models (method 1) in  $A^+$ ,  $B^+$ , and  $C^+$ , and the variation of significance with covariates.

Specific sites were analyzed by quasipoisson models with the Pearson estimate of scale  $\hat{\phi}$ . Additional file 1: Table S7 gives details for Q2 in  $B^+$  at 21 sites, using scale estimates from Q2e and Q2d with identical values of  $min.sp$ . The ML score of Q2d at  $\hat{\phi}_{Q2d}$  is compared with that of Q2e refitted at  $\hat{\phi}_{Q2d}$ , and vice-versa. The same covariate was preferred whichever optimised scale was used. The Q2 response curves and preferred covariates are shown in Fig. 10. Q2e curves are smoothed for display.

Several sites, including the uterus, show non-linear responses. For all solid, genital, respiratory, gallbladder, thyroid, and  $cns$ , the response is significantly elevated for both models in all cells in  $D_0$ . With the preferred covariate, RR is also significantly elevated throughout  $D_0$  for the uterus and corpus, with weaker evidence for squamous,

cervix, and rectum. For lung cancers, using the preferred covariate RR is elevated with borderline significance for over 80% of  $py$  in  $D_0$ .

These 12 sites, along with the breast and liver which are radiosensitive at higher doses but show (non-significant) negative response in Fig. 10, were analysed with interaction models. In Additional file 1: Table S8, method 4 is used for each site and dose range, to choose the preferred quasipoisson model. Poisson selection by methods 1, 2, and 3 is also shown. Using the preferred quasipoisson models, RR and 90% CIs are shown along with the percentages of cases and  $py$  in  $D_0$ ,  $D_{0.5}$ ,  $S_{0.5}$ ,  $T_{0.5}$ , and  $S_{0.5} \cap T_{0.5}$  for which RR is elevated, significantly elevated, or significantly lowered.

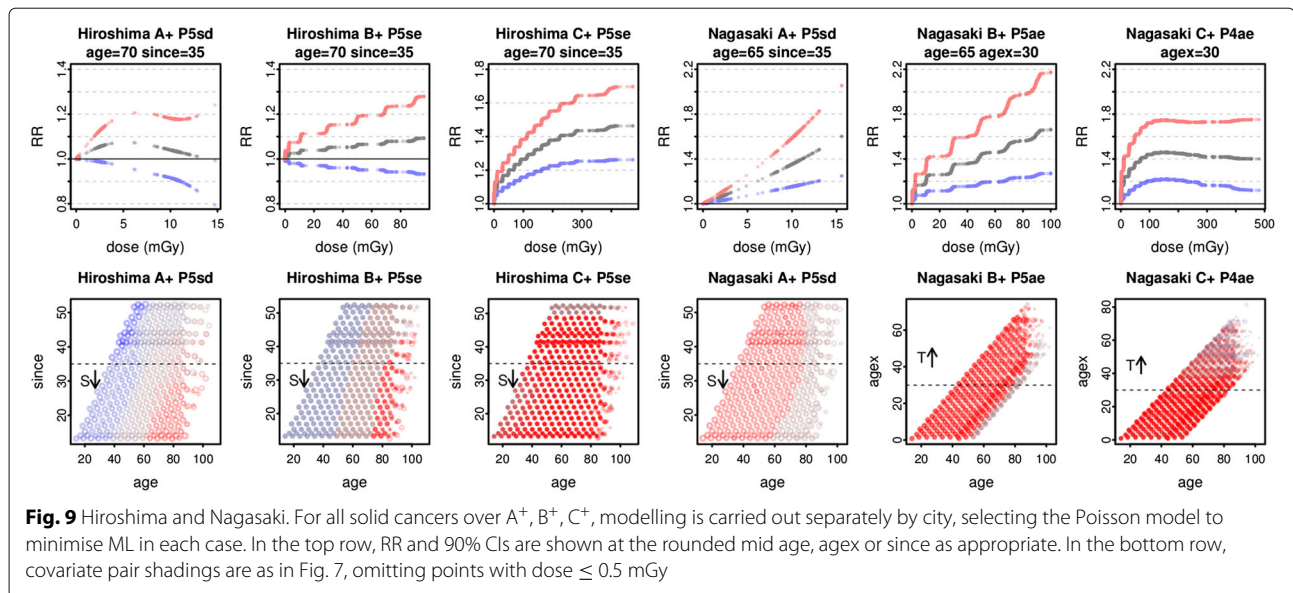
RR is significantly elevated across the dose range for all solid cancers in  $A^-$ ,  $A^+$  (borderline),  $B^{++}$ , and  $C^{++}$ . Squamous cancers show significantly elevated RR in  $A^-$ ,  $B^-$ ,  $B^+$  (borderline 80 mGy), and  $C^{++}$ . The corpus has significant, highly elevated RR in  $A^{++}$  and  $B^{++}$ . Thyroid RR is significantly elevated in  $A^{++}$ ,  $B^{++}$ , and  $C^{++}$ .

Respiratory RR is significantly lowered for 37.2% of  $py$  in  $A^-$  (all with  $age_x < 18$ ), 13.7% of  $py$  in  $A^+$  (all with  $since > 44$ ) and 11.9% of  $py$  in  $C^-$  (all with  $age_x < 8$ ); it is significantly raised in  $S_{0.5}$  and  $T_{0.5}$  in  $B^{++}$ , and in  $D_{0.5}$ ,  $S_{0.5}$  and  $T_{0.5}$  in  $C^{++}$ . For the breast, RR is significantly lowered only for 10.9% of  $py$  in  $B^-$ . There is little significant elevation in  $A^{++}$  or  $B^{++}$ , but RR is significantly raised in  $D_{0.5}$ ,  $S_{0.5}$  and  $T_{0.5}$  in  $C^-$ , and for all  $py$  in  $C^+$ . For the liver, RR is significantly lowered for 24.8% of  $py$  in  $A^-$ , 26.2% in  $A^+$ , 8.7% in  $B^-$ , 5.6% in  $B^+$ , 10.1% in  $C^-$ , and 2.2% in  $C^+$ .

Additional file 1: Figures S5–S10, for respiratory, uterus, and thyroid, correspond with Figs. 7 and 8 for all solid cancers. For the uterus, RR is a U-shaped function of age. For the thyroid, RR peaks around age 60 and decreases with  $age_x$ .

Whilst the preferred models vary over the sites, dose ranges, and selection methods, ecdos models are chosen in 215 out of 336 cases, with a similar proportion for each selection method. Models with  $since$  are chosen in 194 cases and models with  $age_x$  are preferred in 80. Among the 24 models for all solid cancers, ecdos is chosen in 22,  $since$  in 18, and  $age_x$  only once. Overall, there is a clear preference for ecdos and  $since$ , and models with both covariates are preferred in 140 of 336 cases in total.

The male cohorts in  $A^+$  and  $B^+$  were fitted with Poisson models. In  $A^+$ , P5ad (method 1), P4ae (method 2) and P2e (method 3) are preferred. With P5ad at age 70 and with P4ae, the response increases with  $age_x$  but lacks borderline significance. With P2e the response is insignificant. In  $B^+$ , P5se (method 1), P4se (method 2) and P2e (method 3) are preferred. With P5se at age 70, the response decreases with  $since$  but fails to reach borderline



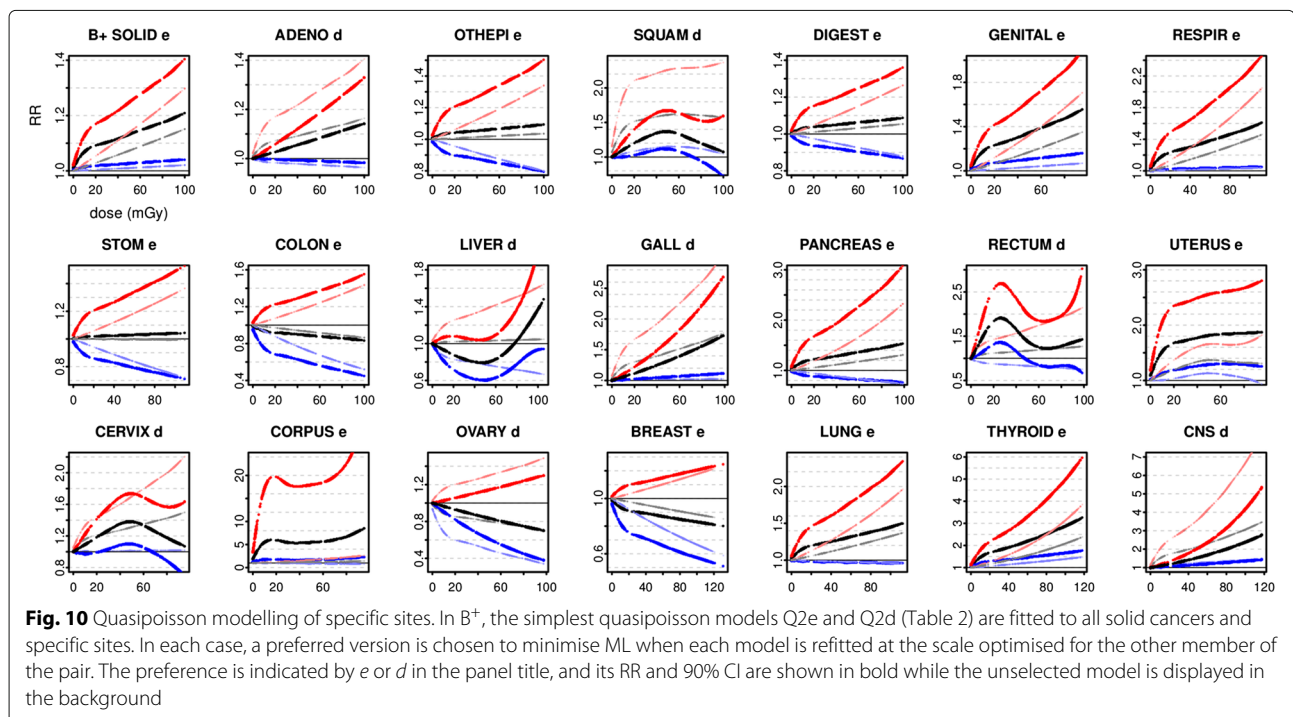
significance. With P4se, the response decreases with since and is borderline for since ≤ 20. With P2e, the result is insignificant.

The combined cohorts in A<sup>+</sup> and B<sup>+</sup> were fitted with Poisson models, controlling for gender, city, distcat and their first order interactions, obtaining a gender averaged dose response. In A<sup>+</sup>, P5sd (method 1) and P4sd (methods 2 and 3) are preferred. For P4sd at 10 mGy RR is significantly raised (90% CIs) when since ≤ 33. At since

30, RR<sub>|10 mGy</sub> = 1.09 (1.02, 1.17). Results from P5sd at age 70 are nearly identical. In B<sup>+</sup>, P5se (method 1) and P4se (methods 2 and 3) are preferred. For P4se at since 30, RR<sub>|10 mGy</sub> = 1.07 (1.01, 1.12), but significance is lost for 19 ≤ since ≤ 27.

### Discussion

Low dose radiation has significant effects at cell level including some non-linear responses [27], but cancer



risk is typically modelled as a linear function of dose and the risk below 100 mGy is still under debate. The body's multiple defences may filter out most abnormalities and the dose-response for disease could be linear.

In analysing the Japanese A-bomb survivors incidence or mortality, the Lifespan studies and other reports [1, 3, 28, 29] use a form of Poisson regression in which  $ERR = RR - 1$  is modelled as  $\beta D$ , where  $\beta$  may be modified by sex, age and age-at-exposure through other parameters; alternatively,  $ERR \sim \beta D + \gamma D^2$ . If the response were known to take a specific form with the same parameter values at all doses, a model fitted to the full data could predict risk at 0.01 Gy. Otherwise, a model which estimates risk at 1 Gy may be inappropriate at lower doses, as several studies recognise [2, 8, 30].

In one approach to detecting non-linearity, numerical data can be reduced to categories before smoothing the results [1]. The outcome may depend on the cut points, and information is discarded by categorising data. The Generalized Additive Model [19, 21] is a more sophisticated method. Using GAM, covariates and interactions enter the model through unknown linear combinations of splines.

The public data released by RERF is stratified by essentially arbitrary dose categories with cutpoints 0, 5, 20, 40 mGy ... as well as categories of age and age-at-exposure, city, sex and location. However the data also contains py-weighted mean values of dose, age, age-at-exposure, and calendar year (giving time-since-exposure). The analysis uses these values, whose distribution, particularly for dose, is far from uniform. Stratification causes clumping of mean dose, but the individual exposures are also concentrated at low doses. As Preston observed [1] "One aspect of the LSS cohort that is often neglected is the highly skewed dose distribution and the large number of survivors who received low-dose, albeit at a high dose rate, exposure."

To compensate for stratification and skew, dose was converted to its empirical cumulative distribution, a monotone (invertible) transformation. The resulting variable *ecdos* is approximately uniformly distributed. Models using *ecdos* gave better fit for all solid cancers and many specific sites. This transformation may still be relevant with the individual data as dose remains skewed, though no longer clumped.

If the true dose-response is linear then modelling with dose is preferable. Whilst there is little difference in detecting a threshold, *ecdos* models are better able to capture alternatives, including plateau, hormesis, and a variety of curves with rapid initial rise and a smaller increase at higher doses. One such curve (Fig. 4 panels c and d) is loosely based on *in vivo* data for  $\gamma$ -H2AX Foci (see [31] (Fig. 1)).

The models have several sources of non-linearity, including the log link, the relation between *ecdos* and dose, and the splines. In fact, the fitted models are often log-linear in *ecdos* as a main effect. However, the interaction of *ecdos* or dose with *logage* and *since* or *agex* can introduce further non-linearities.

For all solid cancers, in  $A^+$ ,  $B^+$  and  $C^+$  all three Poisson selection methods yield models with RR raised across the dose range at the rounded mid values of covariates, with at least borderline significance. Results are stable when deleting exceptional cells, simulating random errors in dosimetry, adjusting RBE, using different splines, or excluding genital or respiratory cancers. For all solid cancers, results are similar with Poisson or quasipoisson models, which explain around 68% of the deviance in  $A^+$ , 58% in  $B^+$ , and 50% in  $C^+$ .

For all solid cancers in  $A^+$ , all the dose or *ecdos* Poisson models show effects which are at least borderline ( $0.98 \leq LCL95\% \leq 1$ ) or significant ( $LCL95\% > 1$ ) at 2.5 and 10 mGy, at the rounded mid values of covariates. The preferred models with *ecdos* refine the results. Significance persists over a wide range of covariates.

With the Nagasaki data, the Poisson model chosen to minimise ML shows RR elevated across the cohort in  $A^+$ ,  $B^+$ , and  $C^+$ ; and in  $B^+$   $ERR|_{10\text{ mGy}}$  is more than twice the value for the combined cohort. Using a very different approach, Sasaki [7] found an exceptional dose response for Nagasaki (male and female).

For all solid cancers, risk estimates are similar with or without the Not In City data. NIC cells are not distinguished by the urban/rural status of the individual members' normal residence, which makes a significant contribution to the model. Even without including NIC, extremely low non-zero doses arise from distant rural residents. However, including NIC gives more cases and greater statistical power.

For all solid cancers and most individual sites, time-since-exposure was a better predictor than age-at-exposure. The fitted dependence of risk on *since* is not governed by a latency threshold, and for all solid cancers, RR decreases with *since* at low doses. Risk could rise with *since* < 12.4 years, the minimum value in the cohort.

Shuryak [32] attributes childhood radiation risk to initiation when cells are proliferating, while older adult risk arises from progression of previously acquired mutations. For all solid cancers, RR increases with *agex* at low doses, suggesting that for women, low dose radiation is acting primarily via progression.

The risks found here are surprising, as linear extrapolation from [1] predicts much lower values at 10 mGy. Preston's Table 10 shows female  $ERR = 0.58/\text{Gy}$  with

90% CI (0.43 , 0.69) at age 70 and agex 30 for solid cancers in the 0 – 4 Gy (+ NIC) dose range (denoted here as J<sup>+</sup>). Adjusting for age-at-exposure, at agex 35  $ERR|_{1\text{ Gy}} = 0.53$  (0.37 , 0.67). Linear extrapolation gives  $ERR|_{10\text{ mGy}} = 0.0053$  (0.0037 , 0.0067). Over B<sup>+</sup>, the ecodos model P5se is preferred by method 1; at age 70 and since 35  $RR|_{10\text{ mGy}} = 1.068$  (1.018 , 1.120), with  $ERR|_{10\text{ mGy}}$  12.8 times the extrapolated value. As Table 4 shows, in A<sup>+</sup>, B<sup>+</sup>, and C<sup>+</sup>, the four selection methods choose ecodos models in 22 out of 24 cases, 90% CIs for  $ERR|_{10\text{ mGy}}$  exclude 0 in those cases and are borderline in 2 cases, and in 23 cases the resulting estimates of

$ERR|_{10\text{ mGy}}$  are 9.6 – 45.7 times the prediction from linear extrapolation.

However, the GAM Poisson models give estimates comparable with [1] when fitted over J<sup>+</sup> and evaluated at 1 Gy. Over this range, the preferred model (method 1, 2 or 3) is P5sd. At 1 Gy, age 70 and since 35,  $RR = 1.80$  (1.60 , 2.03) while for P5ad at agex 35  $RR = 1.56$  (1.40 , 1.73), close to Preston's estimate and CI. The dose response in J<sup>+</sup> for P5ae, P5ad, P5se, or P5sd is roughly comparable with the gender-averaged ERR shown in [1] (Fig. 3, using smoothed category estimates from the 0 – 2 Gy data). Similar results are obtained over J<sup>-</sup> (excluding NIC).

**Table 4** Estimates of ERR at 10 mGy, age 70, agex 35, since 35

Sex	Range	Method	Model	Scale	Covariates	$ERR _{10\text{ mGy}}$ , Female 90% CI, age 70, agex 35, since 35	Ratio
MF	0–4 Gy		LinExt from [1]	1	s,c,d,dose,age,agex	0.0053 (0.0037, 0.0067)	1.0
MF	0–0.5 Gy (urban)		WtMovAv from [2]	1	s,dose,age,agex	0.02 <sup>a</sup> (estimate)	3.8
F	A <sup>-</sup>	1	P5se	1	c,d,ecdos,age,since	0.242 (0.057, 0.459)	45.7
F	A <sup>-</sup>	2	P4se	1	c,d,ecdos,since	0.225 (0.045, 0.437)	42.5
F	A <sup>-</sup>	3	P4se	1	c,d,ecdos,since	0.225 (0.045, 0.437)	42.5
F	A <sup>-</sup>	4	Q4se	0.901	c,d,ecdos,since	0.228 (0.055, 0.430)	43.0
F	A <sup>+</sup>	1	P5se	1	c,d,ecdos,age,since	0.230 (0.054, 0.436)	43.4
F	A <sup>+</sup>	2	P5se	1	c,d,ecdos,age,since	0.230 (0.054, 0.436)	43.4
F	A <sup>+</sup>	3	P4se	1	c,d,ecdos,since	0.202 (0.032, 0.401)	38.1
F	A <sup>+</sup>	4	Q4sd	0.868	c,d,dose,since	0.088 (-0.001, 0.185)	16.6
F	B <sup>-</sup>	1	P5se	1	c,d,ecdos,age,since	0.088 (0.017, 0.164)	16.6
F	B <sup>-</sup>	2	P4se	1	c,d,ecdos,since	0.085 (0.015, 0.160)	16.0
F	B <sup>-</sup>	3	P4se	1	c,d,ecdos,since	0.085 (0.015, 0.160)	16.0
F	B <sup>-</sup>	4	Q4se	1.141	c,d,ecdos,since	0.077 (0.011, 0.148)	14.5
F	B <sup>+</sup>	1	P5se	1	c,d,ecdos,age,since	0.068 (0.021, 0.117)	12.8
F	B <sup>+</sup>	2	P3d	1	c,d,dose,age	0.013 (0.000, 0.026)	2.5
F	B <sup>+</sup>	3	P2e	1	c,d,ecdos	0.061 (0.015, 0.110)	11.5
F	B <sup>+</sup>	4	Q2e	1.102	c,d,ecdos	0.061 (0.013, 0.112)	11.5
F	C <sup>-</sup>	1	P5se	1	c,d,ecdos,age,since	0.064 (0.038, 0.090)	12.1
F	C <sup>-</sup>	2	P5ae	1	c,d,ecdos,age,agex	0.068 (0.045, 0.091)	12.8
F	C <sup>-</sup>	3	P3e	1	c,d,ecdos,age	0.055 (0.034, 0.076)	10.4
F	C <sup>-</sup>	4	Q3e	1.079	c,d,ecdos,age	0.051 (0.034, 0.068)	9.6
F	C <sup>+</sup>	1	P5se	1	c,d,ecdos,age,since	0.091 (0.055, 0.128)	17.2
F	C <sup>+</sup>	2	P4se	1	c,d,ecdos,since	0.057 (0.040, 0.075)	10.8
F	C <sup>+</sup>	3	P4se	1	c,d,ecdos,since	0.057 (0.040, 0.075)	10.8
F	C <sup>+</sup>	4	Q4se	1.066	c,d,ecdos,since	0.056 (0.039, 0.073)	10.6

*Method* Model Selection Method

*Model* for P5se etc. see Table 2

*LinExt* Linear Extrapolation

*WtMovAv* Weighted Moving Averages

*Covariates* s=sex, c=city, d=distcat

*Ratio* ERR estimate from model / ERR estimate by Linear Extrapolation from [1]

<sup>a</sup>estimate from [2] Fig. 1



At 1 Gy, the interaction of ecdos or dose with logage and agex from fitting P5a over  $J^+$  is consistent with [1], whose Fig. 4 is comparable with Fig. 11 here. If agex or since is fixed,  $RR|_{1\text{ Gy}}$  decreases with age. Preston's Fig. 6 (using smoothed category estimates) is comparable with the curve for 1000 mGy in Fig. 12 panel b (P5ad at age 70). At age 70,  $RR|_{1\text{ Gy}}$  initially decreases with agex but rises when agex > 40. Using P5sd, with better fit, at age 70  $RR|_{1\text{ Gy}}$  peaks at since 35. Similar curves arise at age 50, but with higher values of RR.

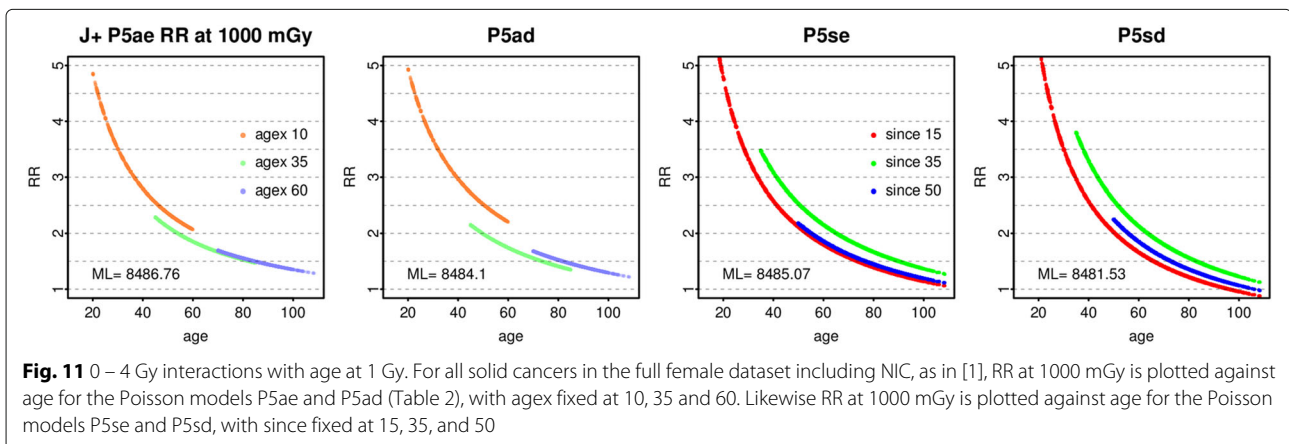
However, the interactions at lower doses (see Figs. 5, 6, and 8) conflict with the main model in [1], which has RR decreasing with agex. At low doses, RR generally decreases with age when agex is fixed, but increases with agex when age is fixed. Likewise, above 100 mGy in  $C^+$  RR decreases with age when since is fixed, but increases with age in  $A^+$  and  $B^+$ . At age 70, RR decreases with since in  $A^+$  and  $B^+$  and when since > 35 in  $C^+$ . In all the low dose ranges considered here, RR increases with agex. Even in  $J^+$ , the curve for 100 mGy does not decrease with agex (see Fig. 12 panel b) Thus, the decrease of RR with agex (when agex < 40) in [1] is specific to the higher dose range and the response at 500 mGy or above.

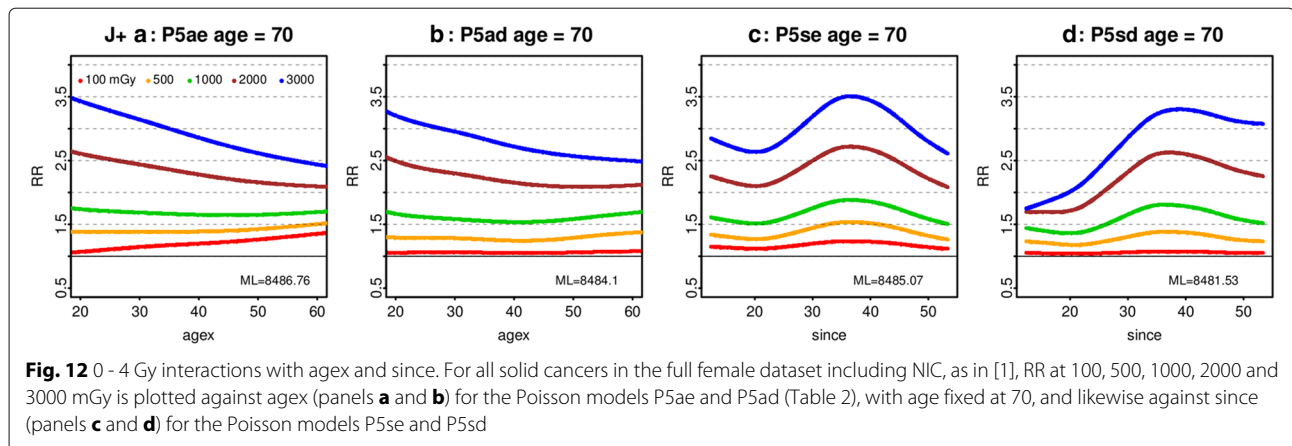
Pierce and Preston [2] applied a simplified linear model to the 0 – 500 mSv urban incidence data for both sexes, finding a higher response than linear extrapolation would predict. Their Fig. 1 shows a smoothed curve obtained as a weighted moving average of category estimates. The sex-averaged RR at agex 30 and 100 mSv  $\sim 1.09$ , from which  $RR|_{\text{female}, 100\text{mSv}} \sim 1.12$  as the fitted model has  $ERR|_{\text{female}} \sim 2 * ERR|_{\text{male}}$ . At agex 35,  $RR|_{\text{female}, 100\text{mSv}} \sim 1.10$ . Figure 1 suggests that  $RR|_{\text{female}, 10\text{mSv}} \sim 1.02$ . By comparison, the preferred model (method 1) for the urban subcohort of  $C^-$  is P5se, and at age 70, since 35,  $RR|_{100\text{ mGy}} = 1.14$  (1.09, 1.18), while  $RR|_{10\text{ mGy}} = 1.03$  (1.02, 1.04). The interaction with since is very weak, and the preferred model with method 2 or 3 is P3e, for which at age 70  $RR|_{100\text{ mGy}} = 1.14$  (1.09, 1.19), as with the preferred quasi model Q3e. Models with either dose or ecdos show RR rising rapidly below

150 mGy before tapering, evidence of non-linearity in the urban data below 500 mGy.

Furukawa and co-workers [8] analysed the combined male and female incidence data below 400 mGy with models including Linear Non Threshold, cubic smoothing spline, and a Bayesian semiparametric piecewise linear model, with dose responses at age 70, agex 30 shown in their Fig. 1. The Bayesian and spline models gave similar responses at low doses. Whilst comparisons with the Bayesian model are beyond the scope of this paper, the spline model is roughly similar to P5ad, if dose knots are set at 0, 5, 20, 40, 60, 80, 100, 125, 150, 175, 200, 250, and 300 mGy. Furukawa analysed the individual data, but results using P5ad (with these knots) applied to the public data below 400 mGy are not significantly different. Figure 13 (panel a) shows the gender averaged dose response for P5ad (knots). The spline estimate in [8] for  $ERR|_{100\text{ mGy}} = 0.03$  (CIs are not shown for this model) whilst for P5ad (knots)  $ERR|_{100\text{ mGy}} = 0.07$  (0.02, 0.11)<sub>95%</sub>. The difference might result from the interaction structures in [8] which follow those in [1]. However, as Fig. 13 shows, the male response (panel b) lacks 95% significance below  $\sim 300$  mGy, but (panel c) female  $ERR|_{100\text{ mGy}} = 0.11$  (0.04, 0.19). Panel d shows the ecdos model P5ae with knots corresponding to those chosen for P5ad (knots). Now  $ERR|_{100\text{ mGy}} = 0.17$  (0.10, 0.24). With the female data, P5ae (knots) is preferred to P5ad (knots) by ML score and AIC. At 10 mGy with 95% CIs, P5ad (knots) gives  $ERR|_{10\text{ mGy}} = 0.011$  (-0.001, 0.023) whilst for P5ae (knots)  $ERR|_{10\text{ mGy}} = 0.052$  (0.031, 0.074).

The values listed in Table 4 are obtained from preferred models. P5ad is never selected, and P5ae is chosen only once. Table 4 largely reflects the predominant selection of ecdos models. Nevertheless, if P5ad and P5ae (with their default knot locations) are fitted over  $A^+$  and evaluated at 10 mGy, age 70 and agex 35, for P5ad RR = 1.10 with 90% CI (1.00, 1.21) while with P5ae RR = 1.21 (1.04, 1.41). P5ae is again preferred to P5ad by ML score and AIC. The default knots for P5ad in  $A^+$  are 0, 0.12, 0.33, 2.30, 2.57,





9.69, 10.55, 15.58, giving a much richer structure than 0, 5, 20. The estimate at 10 mGy when fitting P5ad over A<sup>+</sup> is much higher than the estimate when fitting P5ad over C<sup>+</sup>.

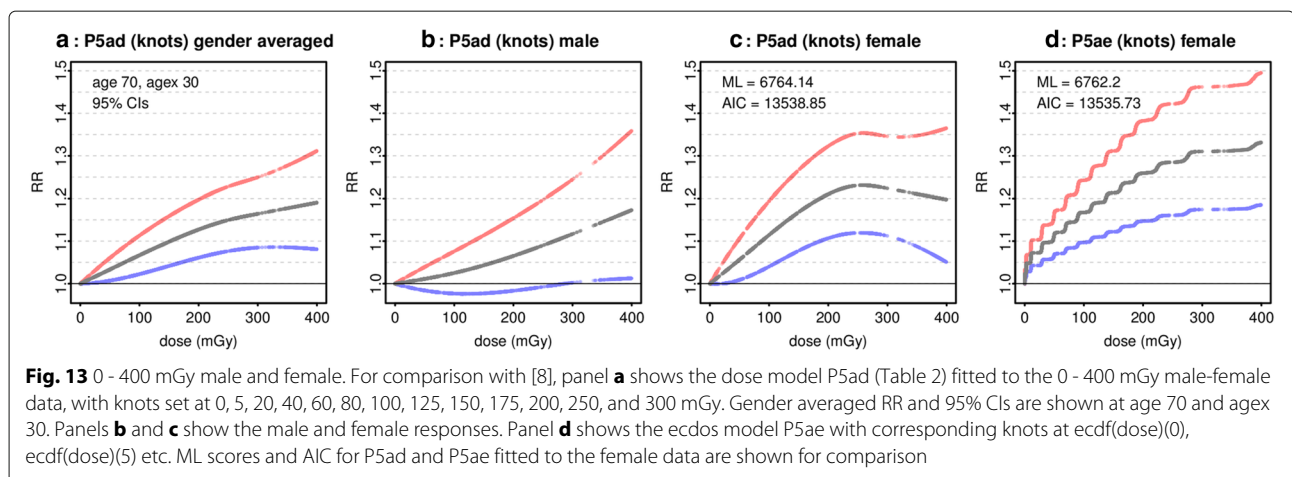
Preston and co-workers [33] analysed the 1958–1998 solid cancer incidence data for exposure in utero or early childhood (0 – 5 years), including NIC. For women exposed in early childhood, at attained age 50  $ERR|_{1Sv} = 2.2 (1.3, 3.4)_{95\%}$  (see [33] Table 5). Linear extrapolation is comparable with results from P5sd over C<sub>i</sub><sup>+</sup> (age<sub>x</sub> ≤ 5, dose ≤ 500 mGy) at age 50 and since 45, for which  $RR|_{10\text{ mGy}} = 1.03 (1.01, 1.05)_{95\%}$ . With the preferred Poisson model (method 1) P5se,  $RR|_{10\text{ mGy}} = 1.20 (1.06, 1.35)_{95\%}$ . Methods 2 or 3 select P2e, giving  $RR|_{10\text{ mGy}} = 1.14 (1.05, 1.24)_{95\%}$ , while method 4 chooses Q2e giving  $RR|_{10\text{ mGy}} = 1.14 (1.06, 1.22)_{95\%}$ .

For site-specific risks, some studies of other cohorts have found RR significantly raised after low dose exposure to low LET radiation. Some evidence from the literature on the thyroid, uterus and uterine corpus, is compared with results from quasipoisson models applied to appropriate subcohorts here.

Direct evidence on thyroid cancer risk arises from children given medical irradiation, residents affected by Chernobyl, nuclear workers, radiotherapy patients, and the A-bomb survivors. Except as indicated, these studies applied a linear ERR model.

Additional file 1: Table S9 shows some results from [1, 6, 7, 11, 13, 15, 34–37]. In each case, an appropriate subcohort of the female incidence data was analysed for comparison, using the simplest quasipoisson dose model Q2d and the preferred quasi model for the subcohort, chosen to approximate the dose and age range in the study, or with mean dose close to that at which risk was reported.

In most cases, results from the subcohorts using either model are compatible with the reported estimates, as CIs overlap. There is striking agreement with the *tinea capitis* studies. Sadetzki [34] used a linear model which, at the mean dose, gives  $RR|_{93\text{ mGy, female}} = 2.96 (2.07, 4.31)_{95\%}$ . Fitting the quasi models over C<sup>+</sup> restricted to age<sub>x</sub> < 16, Q2d gives  $RR|_{93\text{ mGy}} = 2.30 (1.36, 3.89)_{95\%}$ , while the preferred model Q2e gives  $RR|_{93\text{ mGy}} = 2.95 (2.05, 4.24)_{95\%}$ . The earlier study by Ron [35] included



the estimate  $RR|_{10\text{ mGy}} = 1.27 (1.15, 1.42)_{95\%}$ . Fitting the quasi models over  $B^+$  restricted to  $\text{agex} < 16$ , for which the mean py-weighted dose = 10.7 mGy, Q2d gives  $RR|_{10\text{ mGy}} = 1.18 (1.06, 1.31)_{95\%}$ , and at the mean values of  $\text{agex}$  and  $\text{age}$  in [35] the preferred model Q5ad gives  $RR|_{10\text{ mGy, agex 7.4, age 37.6}} = 1.28 (1.07, 1.53)_{95\%}$ . These close agreements contrast with the differences seen in the pooled study [11], where the *tinea capitis* estimates appeared exceptionally high. However, the pooled estimates are compatible with GAM modelling on the appropriate subcohorts.

For Chernobyl survivors with  $\text{agex} < 18$ , Little [6] considered a linear-exponential model with low-dose asymptote  $ERR = \alpha D$ , estimating  $\alpha = 9.72 (2.67, 94.31)_{95\%}$ . At the geometric mean dose as reported in [38], this gives an asymptotic  $RR|_{190\text{ mGy}} = 2.85 (1.51, 18.92)_{95\%}$ . Fitting the quasi models over  $C^+$  restricted to  $\text{agex} < 18$ , Q2d gives  $RR|_{190\text{ mGy}} = 2.62 (1.50, 4.57)_{95\%}$ . At the mean attained age from [39], the preferred model Q3e gives  $RR|_{190\text{ mGy, age 23.7}} = 4.27 (1.30, 14.02)_{95\%}$ .

Sasaki [7] reanalysed the incidence data in [1], with a very different non-linear model. For the thyroid, from his Fig. 3(i)  $ERR|_{0.1\text{ mSv, female}} \sim 0.35$ , though no CI or interactions are shown. Results from  $C^+$  appear compatible. The occupational study by Sont [13] did not estimate RR for female thyroid cancer, but found a Standardized Incidence Ratio = 1.42 (1.19, 1.69)<sub>90%</sub>. This was comparable with internal estimates of SIR (see "Methods"), with CIs by non-parametric bootstrap.

The results from GAM models applied to relevant subcohorts are compatible with [6, 11, 13, 15, 34, 35]. They are higher than extrapolations from [1, 36, 37], which concerned higher dose ranges. However they appear compatible with [7].

Several studies found RR decreasing at higher doses. In the radiotherapy studies, doses were high and the downturn was only apparent above 6 Gy [5] or 10 Gy [36], whilst for Chernobyl survivors risk tapered above 1 Gy [6]. Using the no-interaction dose model Q2d, RR tapers above 400 mGy.

Unlike the thyroid, the uterus is not generally recognised as radiosensitive, although [1] notes an excess of cancers of the uterine corpus at young ages. Occupational studies [40–42] found significant excess mortality from uterine cancers in nuclear workers with a radiation record. The mortality rate increased with dose, but with few cases, mainly endometrial (corpus), ERR/Sv was not reported.

Analysing the National Registry for Radiation Workers, Muirhead [43] found that for mortality (internally calculated SMR) amongst solid cancers, only all uterine cancers showed trend with dose significant at the 5% level using a two-sided test. For incidence, there was a significant increasing trend with dose for endometrial cancer. RR was estimated with a linear model. For incidence of uterine ca

$ERR/Sv = 10.523 (0.27, 39.4)_{90\%}$ . Near the mean dose (24.9 mSv), this corresponds to  $RR|_{25\text{ mSv}} = 1.26 (1.01, 1.99)$ . For incidence of corpus ca, the Health Protection Agency [44] reported that Muirhead found  $ERR/Sv = 14.2 (1.06, 58.56)_{90\%}$ , corresponding to  $RR|_{25\text{ mSv}} = 1.36 (1.03, 2.46)$ .

For comparison, models were fitted over  $B^+$ , restricted to  $16 \leq \text{agex} < 60$ . For the uterus, Q2d gives  $RR|_{25\text{ mGy}} = 1.07 (0.99, 1.15)$ . The preferred model Q3d has  $RR|_{25\text{ mGy, 60yrs}} = 0.96 (0.75, 1.23)$ , while at age 65  $RR = 1.23 (0.98, 1.54)$ , and at age 70, 1.29 (1.04, 1.60). For the corpus, Q2d gives  $RR|_{25\text{ mGy}} = 1.08 (0.89, 1.31)$ . The preferred model Q4se gives  $RR|_{25\text{ mGy, since 30}} = 5.24 (1.22, 22.55)$ , and 4.90 (1.13, 21.24) and 4.00 (0.93, 17.14) at since 25 and 35 respectively.

Davis [16] studied solid cancer incidence in the Techa River cohort, and found significant effects only for the oesophagus and uterus, using a linear model without interactions. Women constitute 57% of the cohort and age of entry begins at 0. The mean dose was 53 mGy and 90% of doses were below 125 mGy. For the uterus, mean age at diagnosis was 57, and the result for  $ERR/100\text{mGy} = 0.21 (0.01, 0.51)_{95\%}$  corresponds to  $RR|_{50\text{ mGy}} = 1.11 (1.01, 1.26)_{95\%}$ .

For comparison, models were fitted over  $B^+$ . Q2d gives  $RR|_{50\text{ mGy}} = 1.37 (1.09, 1.72)_{95\%}$ . The preferred model Q5se gives  $RR|_{50\text{ mGy, age 60, since 15}} = 1.96 (1.12, 3.42)_{95\%}$ , and at since = 20, 25, 30, and 35, the corresponding RR and 95%CI = 1.82 (1.10, 3.02), 1.67 (1.01, 2.76), 1.64 (0.99, 2.73), and 1.70 (1.02, 2.84).

Using a linear model without effect modification, Preston [1] estimated for the uterus  $ERR/Gy = 0.10 (-0.09, 0.33)_{90\%}$ , and for the corpus  $ERR/Gy = 0.29 (-0.14, 0.95)$ . However, for women with  $\text{agex} < 20$ ,  $ERR/Gy$  for the uterus = 0.37 (0.001, 0.86) and for the corpus 1.00 (0.14, 2.40). Near the py-weighted mean dose (85.3 mGy), with  $\text{agex} < 20$  the model predicts for the uterus  $RR|_{80\text{ mGy}} = 1.03 (1.00, 1.07)_{90\%}$  and for the corpus 1.08 (1.01, 1.19).

Fitting over  $C^+$  restricted to  $\text{agex} < 20$ , for the uterus Q2d gives  $RR|_{80\text{ mGy}} = 1.79 (1.19, 2.68)_{90\%}$ . With the preferred model Q5se,  $RR|_{80\text{ mGy, age 50, since 30}} = 2.64 (1.17, 5.97)_{90\%}$ , and at since = 35, 40, 45 and 50 the corresponding RR and 90%CI = 3.07 (1.41, 6.67), 4.51 (2.05, 9.94), 3.76 (1.65, 8.59), and 2.72 (1.11, 6.62). For the corpus Q2d gives 1.03 (0.88, 1.21). For the preferred model Q4ad,  $RR|_{80\text{ mGy, agex 10}} = 0.88 (0.56, 1.37)$ , whilst at  $\text{agex}$  12, 14, 16, 18 and 20 the corresponding values are 1.13 (0.74, 1.74), 1.59 (1.05, 2.43), 2.04 (1.33, 3.11), 2.24 (1.40, 3.60), and 2.18 (0.96, 4.98), i.e. the response is higher after puberty.

I have not used another implementation of GAM or the fully Bayesian methods of MCMC. Confidence intervals, based on Wald-type CIs for the linear predictor, were validated by simulation only for Poisson modelling of all

solid cancers. Although the models explain nearly 60% of the deviance for all solid cancers (in  $B^+$ ), the figures for individual cancer sites are lower, given the smaller case numbers. Minimum smoothing parameters imposed for some sites depended on a visual judgement of the output. The public data lacks information on other exposures which may interact with radiation. The effects could be due to confounding by some other factor linked to distance from the hypocentres, although modelling did control for urban, rural, or NIC location. The modelling was not designed to test a particular hypothesis for cellular mechanism, or for the higher risks for women or Nagasaki residents.

Although increased risk with age-at-exposure (at fixed attained age) was found for many sites, as Boice observed [45] the flash doses all occurred in August 1945, so age-at-exposure is confounded with birth cohort and changes in diet, smoking habits or infection rates, may explain apparent age-at-exposure (or time-since-exposure) effects. Protracted or fractionated exposures may also have different effects.

The impact of radiation in this cohort may be affected by blast damage. Stewart and Kneale [46] investigated possible selection bias, as subjects were only eligible to enter the (mortality) cohort if they survived to 1950. Incidence data begins in 1958, with the minimum time-since-exposure exceeding 12 years.

Wing and Richardson [47] point out that the dose estimates, revised several times, also depend on “the ability to elicit accurate information on location, position and shielding [which] was affected not only by traumatization of the survivors and their domestic stigmatization but by their distrust of medical teams working under occupation forces [48].” The A-bombs annihilated Hiroshima and Nagasaki, destroying their social fabric.

For all these reasons this cohort may not predict risk elsewhere.

## Conclusion

The Generalized Additive Model avoids potential problems in fitting a mis-specified parametric model. In fact, the linear model is an inadequate description of relative risk for cancer incidence in female Japanese A-bomb survivors exposed to low dose radiation. Contrary to the interactions at high doses, for all solid cancers risk increases with age-at-exposure and decreases with time-since-exposure, suggesting that low dose radiation acts through progression of previously accumulated damage.

Time-since-exposure models were generally preferred to age-at-exposure models. Transforming dose to its empirical cumulative distribution improved the model fit for all solid cancers and many specific sites. Constructing splines in this way improves their capacity to detect

responses at very low doses. Covariate interactions were modelled with tensor product splines. Marginal Likelihood was the preferred method of smoothing parameter optimisation, and confidence intervals obtained from the Bayesian posterior covariance performed well.

The results at 10 mGy are much higher than expected. Whilst higher female cancer risks and the highly skewed distribution of dose are recognised, significantly raised Relative Risk for women exposed to low doses, not found for all solid cancers in men, is new. Most other A-bomb survivor studies consider much wider dose ranges.

Whilst the results may be specific to this cohort, some studies of other cohorts with other methods give comparable results, notably for thyroid cancer where estimates are consistent with those derived here, and for the uterus.

Direct epidemiological evidence of risks from doses around 10 mGy is a current research priority [18]. Risk estimates derived here from low dose data are much higher than those on which the ICRP has relied when setting recommended annual dose limits [4], including the annual occupational dose limit of 20 mSv. For all solid cancers, almost all estimates here of Excess Relative Risk for women at 10 mGy are 9.6 – 45.7 times the prediction by linear extrapolation from [1].

Generalized additive models applied to the low dose data for female Japanese A-bomb survivors show significant Relative Risks, particularly if the heavily skewed doses are transformed to improve their distribution. Evidence of elevated risk at low doses should contribute to higher radiation risk estimates and improve radiation protection for women.

## Additional file

**Additional file 1:** Additional files are linked from a mini-website available with the online version of this paper. They consist of 1 PDF file with the Appendices, 10 images, 1 Excel workbook with 9 tables, 4 code files, 2 data files, and 6 output files. (ZIP 3727 kb)

## Abbreviations

$\chi^2$ : Pearson Chi-squared statistic;  $\hat{\phi}$ : Pearson estimate of scale for quasipoisson model;  $A^+$ : 0 - 20 mGy including (excluding) NIC subjects;  $B^+$ : 0 - 100 mGy + NIC;  $C^+$ : 0 - 500 mGy + NIC; CD: Cook's Distance; CI: Bayesian posterior confidence interval;  $D_\delta$ : data cells with dose >  $\delta$ ; dfres: residual degrees of freedom; e.d.f.: effective degrees of freedom; ERR: Excess Relative Risk; ecdf: empirical cumulative distribution function; ecdos: empirical cumulative distribution of dose; GAM: Generalized Additive Model; GCV.Cp: Generalized Cross Validation/Mallows Cp;  $J^+$ : 0 - 4 Gy + NIC; LCL: Bayesian posterior lower confidence limit; ML: Marginal Likelihood; NIC: Not In City; REML: Restricted Maximum Likelihood; RERF: Radiation Effects Research Foundation; RR: Relative Risk;  $S_\delta$ : dose >  $\delta$  and since < 35;  $T_\delta$ : dose >  $\delta$  and agex > 35; UCL: Bayesian posterior upper confidence limit

## Acknowledgments

Thanks to Simon Wood for help with GAMs and mgcv, and to David Richardson for encouragement. Thanks to the reviewers Richard Clapp and Kyoji Furukawa for editorial comments, and for alerting me to Dr Furukawa's analysis.

The data described in this report were obtained from the Radiation Effects Research Foundation (RERF) in Hiroshima, Japan. RERF is a private foundation that is funded equally by the Japanese Ministry of Health and Welfare and the US Department of Energy through the US National Academy of Sciences. The conclusions reached in this report are mine and do not necessarily reflect the scientific judgement of RERF or its funding agencies.

#### Funding

This work was not funded.

#### Availability of data and materials

The dataset supporting the conclusions of this article is available in the Radiation Effects Research Foundation Downloadable Data repository, as the Life Span Study Solid Cancer Incidence Data, 1958–1998 and can be downloaded from [http://www.rerf.jp/library/dl\\_e/lssinc07.html](http://www.rerf.jp/library/dl_e/lssinc07.html)

Some extracts from the dataset are included within the article's additional files.

#### Authors' contributions

GD designed the study and carried out the computations.

#### Competing interests

The authors declare that they have no competing interests.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

The study involves only freely available anonymous public data and published literature. I have not sought further Ethics Approval.

Received: 5 July 2016 Accepted: 26 October 2016

Published online: 24 November 2016

#### References

- Preston DL, Ron E, Tokuoka S, Funamoto S, Nishi N, Soda M, et al. Solid cancer incidence in atomic bomb survivors: 1958–1998. *Radiat Res.* 2007;168:1–64.
- Pierce DA, Preston DL. Radiation-related cancer risks at low doses among atomic bomb survivors. *Radiat Res.* 2000;154:178–86.
- Blue Book: EPA Radiogenic Cancer Risk Models and Projections for the U.S. Population. Washington DC: U.S. Environmental Protection Agency, Office of Radiation and Indoor Air. 2011. EPA 402-R-11-001.
- ICRP. 2007 Recommendations of the International Commission on Radiological Protection (Users Edition): ICRP Publication 103 (Users Edition). *Ann ICRP.* 2007;37(2–4).
- Ronckers CM, Sigurdson AJ, Stovall M, Smith SA, Mertens AC, Liu Y, et al. Thyroid cancer in childhood cancer survivors: a detailed evaluation of radiation dose response and its modifiers. *Radiat Res.* 2006;166:618–28.
- Little MP, Kukush AG, Masiuk SV, Shklyar S, Carroll RJ, Lubin JH, et al. Impact of uncertainties in exposure assessment on estimates of thyroid cancer risk among Ukrainian children and adolescents exposed from the Chernobyl accident. *PLoS One.* 2014;e85723:9.
- Sasaki MS, Tachibana A, Takeda S. Cancer risk at low doses of ionizing radiation: artificial neural networks inference from atomic bomb survivors. *J Radiat Res.* 2014;55:391–406.
- Furukawa K, Misumi M, Cologne J, Cullings H. A bayesian semiparametric model for radiation dose-response estimation. *Risk Anal.* 2016;36:1211–23.
- Stewart AS, Webb J, Giles D, Hewitt D. Malignant disease in childhood and diagnostic irradiation in utero. *Lancet.* 1956;268(6940):447.
- Wing S, Shy CM, Wood JL, Wolf S, Cragle DL, Frome EL. Mortality among workers at Oak Ridge National Laboratory: evidence of radiation effects in follow-up through 1984. *JAMA.* 1991;265:1397–402.
- Ron E, Lubin JH, Shore RE, Mabuchi K, Modan B, Pottern LM, et al. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res.* 1995;141:259–77.
- Richardson DB, Wing S. Greater sensitivity to ionizing radiation at older age: follow-up of workers at Oak Ridge National Laboratory through 1990. *Int J Epidemiol.* 1999;28:428–36.
- Sont WN, Zielinski JM, Ashmore JP, Jiang H, Krewski D, Fair ME, et al. First analysis of cancer incidence and occupational radiation exposure based on the National Dose Registry of Canada. *Am J Epidemiol.* 2001;153:309–18.
- Cardis E, Vrijheid M, Blettner M, Gilbert E, Hakama M, Hill C, et al. Risk of cancer after low doses of ionizing radiation: retrospective cohort study in 15 countries. *BMJ.* 2005;331(7508):77.
- Kopecky KJ, Stepanenko V, Rivkind N, Voillequé P, Onstad L, Shakhtarin V, et al. Childhood thyroid cancer, radiation dose from Chernobyl, and dose uncertainties in Bryansk Oblast, Russia: a population-based case-control study. *Radiat Res.* 2006;166:367–74.
- Davis FG, Krestinina LY, Preston D, Epifanova S, Degteva M, Akleyev A. Solid cancer incidence in the Techa River incidence cohort: 1956–2007. *Radiat Res.* 2015;184:56–65.
- Brenner DJ, Doll R, Goodhead DT, Hall EJ, Land CE, Little JB, et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. *Proc Natl Acad Sci U S A.* 2003;100:13761–6.
- NAS Nuclear and Radiation Studies Board. Planning towards the BEIR VIII report. 2014. <http://dels.nas.edu/Past-Events/Planning-Towards-BEIR-VIII-Report/AUTO-0-14-84-B?bname=nrsb>. Accessed 20 Oct 2016.
- Hastie T, Tibshirani R. Generalized additive models (with discussion). *Statist Sci.* 1986;1:297–318.
- R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013. <http://www.R-project.org>. Accessed 25 Sept 2013.
- Wood SN. Generalized Additive Models: an introduction with R. Boca Raton, London, New York: Chapman & Hall/CRC: Taylor & Francis Group; 2006.
- Life Span Study Solid Cancer Incidence Data, 1958–1998. Hiroshima, Japan: RERF; 2007. [http://www.rerf.jp/library/dl\\_e/lssinc07.html](http://www.rerf.jp/library/dl_e/lssinc07.html). Accessed 20 Feb 2010.
- Radiation Effects Research Foundation (RERF). A Cooperative Japan-US Research Foundation. <http://www.rerf.jp>. Accessed 20 Oct 2016.
- Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol.* 2011;73:3–36.
- Sasaki MS, Nomura T, Ejima Y, Utsumi H, Endo S, Saito I, et al. Experimental derivation of relative biological effectiveness of A-bomb neutrons in Hiroshima and Nagasaki and implications for risk assessment. *Radiat Res.* 2008;170:101–17.
- DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Statist Sci.* 1996;11:189–228.
- UNSCEAR. Report to the General Assembly, with scientific annexes. New York: United Nations Scientific Committee on the Effects of Atomic Radiation. 2012;2012.
- Preston DL, Shimizu Y, Pierce DA, Suyama A, Mabuchi K. Studies of mortality of atomic bomb survivors. Report 13: solid cancer and noncancer disease mortality: 1950–1997. *Radiat Res.* 2003;160:381–407.
- National Research Council. Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation. Health risks from exposure to low levels of ionizing radiation: BEIR VII Phase 2. Washington DC: National Academies Press; 2006.
- Nakashima E. Radiation dose response estimation with on low dose range using restricted cubic splines: application to all solid cancer mortality data, 1950–2003, in atomic bomb survivors. *Health Phys.* 2015;109:15–24.
- Beels L, Bacher K, De Wolf D, Werbrouck J, Thierens H. gamma-H2AX foci as a biomarker for patient X-ray exposure in pediatric cardiac catheterization: are we underestimating radiation risks? *Circulation.* 2009;120:1903–9.
- Shuryak I, Sachs RK, Brenner DJ. Cancer risks after radiation exposure in middle age. *J Natl Cancer Inst.* 2010;102:1628–36.
- Preston DL, Cullings H, Suyama A, Funamoto S, Nishi N, Soda M, et al. Solid cancer incidence in atomic bomb survivors exposed in utero or as young children. *J Natl Cancer Inst.* 2008;100:428–36.
- Sadetzki S, Chetrit A, Lubina A, Stovall M, Novikov I. Risk of thyroid cancer after childhood exposure to ionizing radiation for tinea capitis. *J Clin Endocrinol Metab.* 2006;91:4798–804.
- Ron E, Modan B, Preston D, Alfandary E, Stovall M, Boice JDJ. Thyroid neoplasia following low-dose radiation in childhood. *Radiat Res.* 1989;120:516–31.
- Bhatti P, Veiga LHS, Ronckers CM, Sigurdson AJ, Stovall M, Smith SA, et al. Risk of second primary thyroid cancer after radiotherapy for a childhood cancer in a large cohort study: an update from the childhood cancer survivor study. *Radiat Res.* 2010;174:741–52.



37. Richardson DB. Exposure to ionizing radiation in adulthood and thyroid cancer incidence. *Epidemiology*. 2009;20:181–7.
38. Likhtarov I, Kovgan L, Masiuk S, Talerko M, Chepurny M, Ivanova O, et al. Thyroid cancer study among Ukrainian children exposed to radiation after the Chernobyl accident: improved estimates of the thyroid doses to the cohort members. *Health Phys*. 2014;106:370–96.
39. Tronko MD, Howe GR, Bogdanova TI, Bouville AC, Epstein OV, Brill AB, et al. A cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: thyroid cancer in Ukraine detected during first screening. *J Natl Cancer Inst*. 2006;98:897–903.
40. Fraser P, Carpenter L, Maconochie N, Higgins C, Booth M, Beral V. Cancer mortality and morbidity in employees of the United Kingdom Atomic Energy Authority, 1946–86. *Br J Cancer*. 1993;67:615–24.
41. Carpenter L, Higgins C, Douglas A, Fraser P, Beral V, Smith P. Combined analysis of mortality in three United Kingdom nuclear industry workforces, 1946–1988. *Radiat Res*. 1994;138:224–38.
42. Atkinson WD, Law DV, Bromley KJ, Inskip HM. Mortality of employees of the United Kingdom Atomic Energy Authority, 1946–97. *Occup Environ Med*. 2004;61:577–85.
43. Muirhead CR, O'Hagan JA, Haylock RG, Phillipson MA, Willcock T, Berridge GL, et al. Mortality and cancer incidence following occupational radiation exposure: third analysis of the National Registry for Radiation Workers. *Br J Cancer*. 2009;100:206–12.
44. Independent Advisory Group on Ionising Radiation. Risk of Solid Cancers following Radiation Exposure: Estimates for the UK Population. Health Protection Agency: Chilton; 2011.
45. Boice, JD Jr. Models, models everywhere—is there a fit for lifetime risks. *J Natl Cancer Inst*. 2010;102:1606–9.
46. Stewart AM, Kneale GW. A-bomb survivors: factors that may lead to a re-assessment of the radiation hazard. *Int J Epidemiol*. 2000;29:708–14.
47. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241119/pdf/ehp01110-a00739.pdf>.
48. Lindee MS. *Suffering made real: American science and the survivors of Hiroshima*. Chicago: University of Chicago Press; 1994.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

