

Quantitative Analysis of Physical data: Assignment 3

Rebekah Hawkins and Madhurima Nath

Department of Physics, Virginia Tech, Blacksburg, VA, USA 24061

(Dated: March 2, 2017)

The effect of low doses of ionizing radiation on the mutation of DNA and onset of cancer remains largely unknown. The NRC report on the health effects of such low dose radiation included a study of solid cancer tumors in Japanese atomic bomb survivors. The data obtained from this study is used to perform maximum likelihood analyses for two different models. The linear model assumes that the onset of cancer in humans is directly proportional to the dosage of radiation exposure. The threshold model is introduced, which takes into account the efficiency of the body to defend against mutations up to a certain threshold dose beyond which cancer is linearly dependent on the additional radiation exposure. The goodness of fit is measured for both models and the effects on policy consequences are discussed.

INTRODUCTION

It is well known that there exists a correlation between cancer and radiation, due to radiation causing mutations in DNA, which can lead to cancer if not corrected. Because of this, radiation has become a major public health concern to be studied and regulated. The harmfulness of ionizing radiation at high doses is made plain by the effects on the exposed tissue shortly after exposure, but it is not as clear how much of an effect low doses (defined as up to 0.1 Sv) of ionizing radiation have on the cancer risk of an individual. If we assume that the rate of mutations in the DNA of an exposed tissue is directly proportional to the dose of ionizing radiation the tissue received, then it seems reasonable to conclude that without any corrections to the DNA, the probability of getting cancer due to this radiation would also be proportional to the radiation dose. However, the body has many layers of defense against DNA mutation, so the nature of the body's efficiency of correcting mutations is what would determine the relationship between dosage of ionizing radiation and probability of cancer developing as a result of the radiation.

Two models, in particular, for the body's efficiency at defending against cancer-causing mutations would have significant consequences for policies regulating low-dose radiation. The first assumes that the efficiency does not depend of the radiation dose, which means that the radiation dose and the probability of getting cancer from it is directly proportional. We will call this the linear model. The second assumes that the body is very efficient unless it is overwhelmed with mutations, which means that the probability of cancer from that radiation is negligible up to a certain threshold dose, after which it becomes directly proportional to the dose received in excess to the threshold dose. We will call this the threshold model. If in fact the threshold model is the better fit, there would be no reason to regulate low-dose radiation, since it would not increase the likelihood of an individual getting cancer.

In 2006, the National Research Council (NRC) of National Academies released a report on health effects of low dose, low-LET (linear energy transfer) ionizing radiation

[1]. This report included a study of the rates of solid cancer tumors in Japanese atomic bomb survivors, the data of which is shown in figure 1. This population includes people of both sexes and all ages, in addition to a wide range of doses estimated for individual subjects and the whole body exposure. They fit the data with two curves, one linear and one quadratic, and they determined that the linear model was consistent with the data, though the quadratic curve gave a better fit. They stated that the large number of survivors at doses of < 0.5 Sv and the analysis of the dose-response behavior helps to establish that linear functions are a reasonable approximation to represent solid cancer risks.

While the NRC did not find the data to be inconsistent with the linear model, we would like to know if the data is also consistent with the threshold model and which one provides a better fit. In this paper, we will fit both a linear model and a threshold model to the data, determining the maximum likelihood estimate for each parameter, and compare the goodness-of-fit for each model.

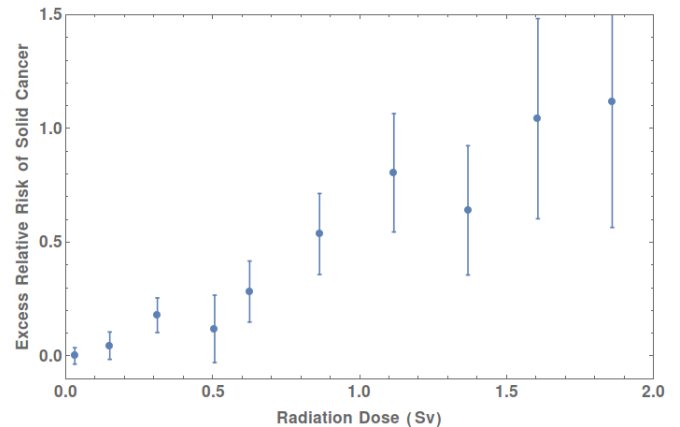


FIG. 1: National Research Council's data for the rate of solid cancer tumors amongst Japanese atomic bomb survivors who experienced different doses of radiation. The cancer rate is reported as relative to (and in excess of) cancer rates for those who were not exposed to significant radiation[1].

METHODS

For a given model fitting a function $f(x_i, \boldsymbol{\lambda})$ to the set of data points $\{(x_1, y_1), \dots, (x_n, y_n)\}$ using function parameters $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_d\}$, the log-likelihood function $\log[L(\boldsymbol{\lambda})]$ is given by

$$2\chi^2 = -\log[L(\boldsymbol{\lambda})] = \frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(x_i, \boldsymbol{\lambda}))^2}{\sigma_i^2} \quad (1)$$

where σ_i is the standard deviation of data point i and χ^2 is what we will refer to as the log-likelihood in the rest of the paper. Thus, the probability density of $\mathbf{y} = \{y_i\}$ given the parameters $\boldsymbol{\lambda}$ would be

$$p(\mathbf{y}|\boldsymbol{\lambda}) = A \prod_{i=0}^n \exp \left\{ -\frac{[y_i - f(x_i, \boldsymbol{\lambda})]^2}{\sigma_i^2} \right\} \quad (2)$$

where A is a normalization constant. For the data we are analyzing, y_i is the excess relative risk (ERR) for a dose x_i and σ_i is 1/4 the total length of each error bar in figure 1. The ERR is defined to be one less than the ratio of the rate of solid cancer tumors observed among the Japanese atomic bomb survivors who experienced the specified dose of radiation from the event to that observed in the general population.

The maximum likelihood estimates (MLE) of parameters for a model are the values of the parameters for which χ^2 is a minimum. We will denote them as $\hat{\boldsymbol{\lambda}}$. These are the values of the parameters used in reporting the best fit curves for each model. The value of the minimum χ^2 is denoted as $\hat{\chi}^2$.

Once we have the values of $\hat{\boldsymbol{\lambda}}$ from the χ^2 minimization, we use them as the true values in Monte Carlo simulations to determine the range of $\boldsymbol{\lambda}$. The simulation involves drawing a random set \mathbf{y} from $p(\mathbf{y}|\hat{\boldsymbol{\lambda}})$ and obtaining the MLEs $\tilde{\boldsymbol{\lambda}}$ and value of the minimum χ^2 , $\tilde{\chi}^2$, for this set. This is done 1000 times for each model to obtain sampling distributions of $\hat{\boldsymbol{\lambda}}$ and of $\hat{\chi}^2$. The $\hat{\boldsymbol{\lambda}}$ distribution is used to determine the standard deviation σ_{λ_i} of each parameter λ_i using the equation

$$\sigma_{\lambda_i} = \sqrt{\frac{1}{1000} \sum_{n=1}^{1000} (\tilde{\lambda}_{in} - \hat{\lambda}_i)^2} \quad (3)$$

The $\hat{\chi}^2$ distribution is used to determine the goodness-of-fit (or p-value), given by the fraction of $\tilde{\chi}^2$ for which $\tilde{\chi}^2 < \hat{\chi}^2$.

Linear Model

For the linear model, there is only one parameter λ , since no excess radiation exposure implies no excess rate of cancer, which means the line needs to go through the origin. The equation of the fit curve is thus given by

$$f_i(x_i, \lambda) = \lambda x_i \quad (4)$$

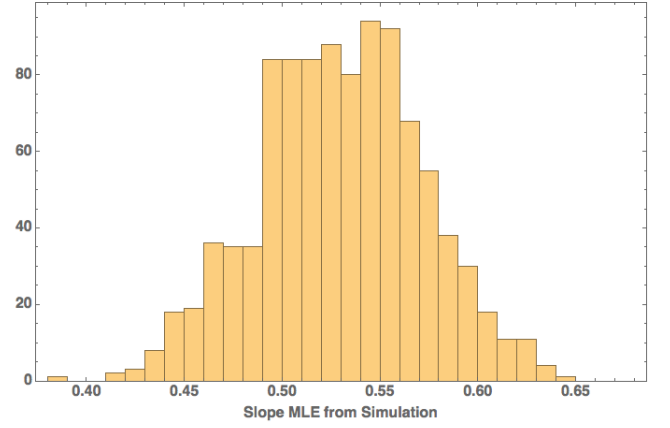


FIG. 2: Histogram of the values of λ obtained by fitting simulated data points with a linear model for 1000 simulations.

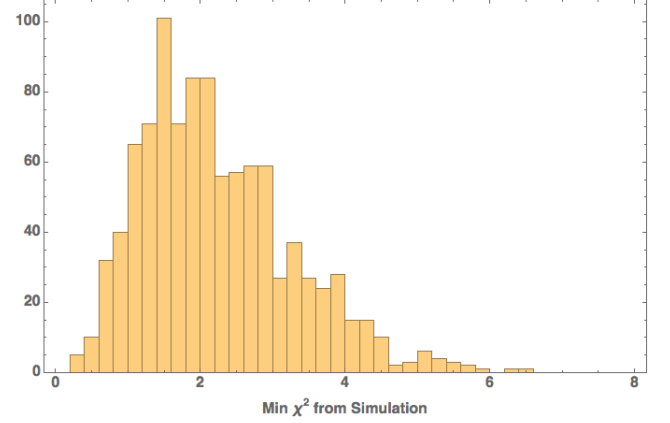


FIG. 3: Histogram of the minimum χ^2 values obtained by fitting simulated data points with a linear model for 1000 simulations.

Using the method discussed above, we obtain $\hat{\lambda} = 0.532$ and $\hat{\chi}^2 = 2.926$. Figure 2 shows a histogram of the MLE obtained from a sampling distribution for $\hat{\lambda}$. It is highly peaked around $\hat{\lambda}$ with a standard deviation of 0.04. The sampling distribution of $\hat{\chi}^2$ is shown in figure 3. The goodness-of-fit for this model's best fit curve is 0.796.

Threshold Model

For the threshold model, there are two parameters, one for the slope of the line λ_1 and another from which we derive the threshold dose λ_2 . For doses below λ_2/λ_1 , the ERR predicted by the model will be 0. This equation is given by

$$f_t(x_i, \lambda_1, \lambda_2) = \begin{cases} 0 & \text{if } x < \lambda_2/\lambda_1 \\ \lambda_1 x - \lambda_2 & \text{if } x \geq \lambda_2/\lambda_1 \end{cases} \quad (5)$$

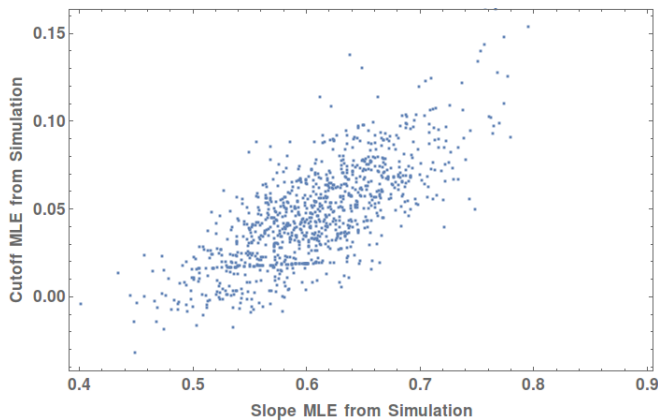


FIG. 4: Distribution of the values of λ_1 and λ_2 obtained by fitting simulated data points with a threshold model for 1000 simulations.

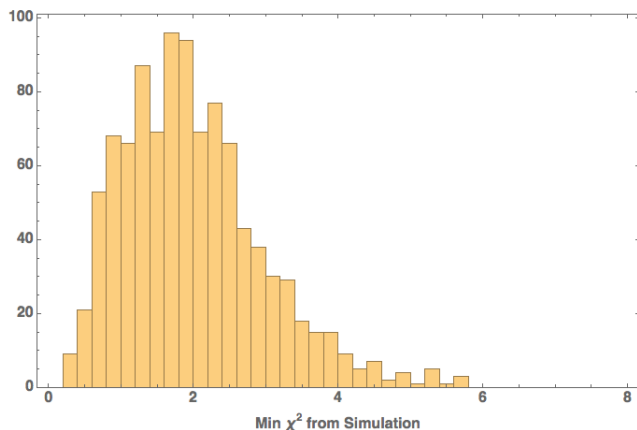


FIG. 5: Histogram of the minimum χ^2 values obtained by fitting simulated data points with a threshold model for 1000 simulations.

Using the method described previously, we obtain $\hat{\lambda}_1 = 0.606$, $\hat{\lambda}_2 = 0.045$, and $\hat{\chi}^2 = 2.173$. Figure 4 shows a scatter plot of the MLEs obtained from sampling distributions for $\hat{\lambda}_1$ and $\hat{\lambda}_2$. The oblique shape of the distribution is caused by the slope needing to be increased when the threshold is increased in order to best fit the points. The distributions have standard deviations of 0.06 and 0.03, respectively. The sampling distribution of $\hat{\chi}^2$ is shown in figure 5. The goodness-of-fit for this model's best fit curve to the data points is 0.625.

DISCUSSION

As mentioned in the introduction, the NRC report concludes that the data is consistent with the existence of a proportionate, no threshold correlation between exposure to radiation dose and cancer risk in humans. They also tested a linear quadratic model, but statistical improvement is not enough to rule out a linear model for

	Linear Model	Threshold Model
Best Fit	$f_l(x) = 0.53x$	$f_t(x) = 0.61x - 0.05$ for $x \geq 0.07$
$\hat{\chi}^2$	2.93	2.17
St. Dev.	$\sigma_\lambda = 0.04$	$\sigma_{\lambda_1} = 0.06, \sigma_{\lambda_2} = 0.03$
p-value	0.80	0.63

TABLE I: Comparison between the two fitting models used for the analyses. The standard deviation and the goodness of fit values are recorded for the linear and the threshold model.

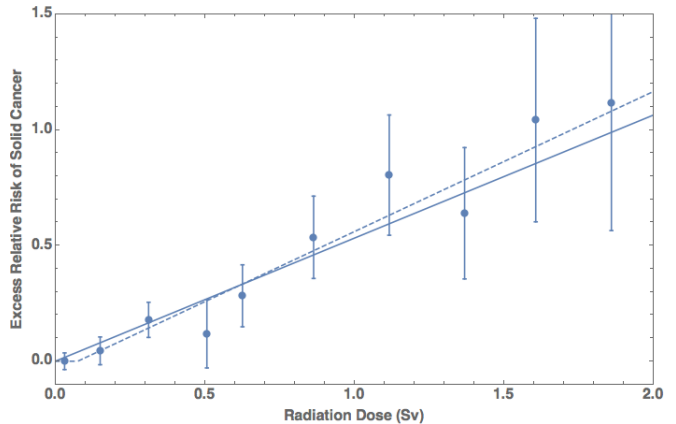


FIG. 6: The solid line is a linear fit to the data points in figure 1. The fit obtained from the delayed effects model is shown by the dotted line.

solid cancer cases [1]. The linear behavior assumes that the probability of getting cancer is directly proportional to the amount of the exposure to the radiation dose. We introduce a new model, called the threshold model, where we assume that the probability of cancer from radiation exposure only after a certain threshold value, up to which the body is very efficient at correcting mutations. The data obtained from the NRC report is analyzed using maximum likelihood estimates in a goodness-of-fit test, as described in the methods section. The results from this analysis is summarized in table I, and the best fit curves for each model are plotted with the data in figure 6.

Though the linear model depends on a single parameter λ , the slope of the fitted line, and the threshold model has two parameters, λ_1 which gives the slope the fitted line and λ_2 which is related the threshold dose, the goodness-of-fit is actually smaller for the threshold model (63%) than for the linear model (80%). However, the χ^2 for each of the best fit curves would suggest that the threshold model ($\hat{\chi}^2 = 2.17$) is a better fit than the linear model ($\hat{\chi}^2 = 2.93$).

From the functional forms of the best fit curves, we can establish that in the limit $\lambda_2 \rightarrow 0$, the function $f_t(x) \rightarrow f_l(x)$ and $\lambda_1 \rightarrow \lambda$. However, from looking at the scatter plot of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ (figure 4), we can see that the limit $\lambda_2 \rightarrow 0$ is on the fringes of the populated region of the

plot. Therefore, the threshold model appears to be a better model for the data than the linear model. Neither can be ruled out without better statistics for the data, though.

As far as regulation policies go, the threshold model's best fit curve suggests that regulation is not needed below 70 mSv. However, it is safer to base regulations off of the linear model until it is significantly discounted, since the linear model would suggest the need of more regulation.

HONOR CODE

I have neither given nor received unauthorized assistance on this assignment.

-
- [1] Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, National Research Council. *Health Risks from Exposure to Low Levels of Ionizing*

Radiation: BEIR VII Phase 2. The National Academies Press, 2006.