Analysis for Radiation Exposure and Risk of Death from Leukemia and Other Cancers

Roland Moore, Jerri Schorr, Nicholas Vasquez
Statistics Consultants
12/16/2020

Background

In this analysis, we are interested in studying the risk of cancer a person may face from a lifetime of repeated exposure to radiation; specifically from multiple computed tomography (CT) scans. CT Scanning may lead to mutations in a person's DNA which can develop into tumors. One CT scan is approximately equivalent to one year of exposure to radiation from natural and artificial environmental sources.

Objective

The goal of this analysis is to establish a causal relationship between radiation exposure and the risk of death from leukemia and other cancers. The statistical consultants will analyze the data from the hiroshima.csv dataset, develop an appropriate statistical model, and explain the relationship.

Statistical Methods

Model 1, with leukemia deaths as the response variable is:

$$leuk = \beta_0 + \beta_1 * radgrp + \beta_2 * agegrp$$

Model 2, with other cancer deaths as the response variable:

$$cancer = \beta_0 + \beta_1 * radgrp + \beta_2 * agegrp$$

The consultants used the methods of a Poisson Distribution to find how radiation and age affected the death count per 1,000 people, using R to fit the data to an exponential model befitting a poisson distribution. From this model, we will be able to run a random trial and approximate data points and beta coefficients.

Results

Average Deaths per 1,000 People

	Rad 1	Rad 2	Rad 3	Rad 4	Rad 5	Rad 6	Avg Δ
Age 1	0.2295	.4491	.8847	1.7499	3.4754	6.9309	1.3403
Age 2	.2447	.4802	.9461	1.8718	3.7183	7.4171	1.4345
Age 3	.2626	.5154	1.0157	2.0098	3.9934	7.9677	1.5410
Age 4	.2829	.5553	1.0945	2.1663	4.3054	8.5920	1.6618
Age 5	.3058	.6005	1.1840	2.3439	4.6594	9.3006	1.7990
Age 6	.3320	.6519	1.2856	2.5457	5.0617	10.1061	1.9548
Age 7	.3617	.7104	1.4013	2.7755	5.5198	11.0233	2.1323
Avg Δ	0.0220	.0436	.0861	.1709	.3407	.6821	

Figure 1

Figure 1: We see the results of this code in the form of a table of values. Each row represents an age group, from 1 to 7, and each column is a radiation group, from 1 to 6. The table also contains the average change per row and column. The average (delta) column shows us the average increase from each column or row to the next. We can see a difference of .2196 between radiation 2 and radiation 1 for age group 1. There is a 3.4555 difference between radiation 6 and radiation 5 for age group 1. When we take the average of all of these values in age group 1 we get an average increase in deaths of 1.3403.

Radiation increasing tends to result in far more deaths than the age increasing, which could mean that deadly amounts of radiation, while more harmful for older people, are still especially harmful to everyone. This can be seen by the significantly lower Δ value for fixed rad amounts. For a fixed age, each increase in rad results in over 1 more expected death per 1,000 people from that age group, but in each rad group, age only increases average death count by less than 0.7 per 1,000 people. It is interesting to note that for fixed rad, the average Δ as age changes tends to double from one rad group to the next. This does show that while dangerous for all groups, higher radiation levels are more dangerous for older groups.

Beta Coefficients For Leukemia Model

	Mean	Median	95% CI
β_0: Y intercept	-9.180	-9.179	(-9.8928, -8.4739)
β_1: RADGRP	0.6844	0.6839	(0.5601, 0.8112)
β_2: AGEGRP	0.0753	0.0758	(-0.0449, 1.9596)

Figure 2

Figure 2 gives us information about the Beta Coefficients. The important notes are that the y-intercept is below 0, meaning that for every thousand people with no radiation or age, there would be a death count of 0. We are 95% confident that the coefficient for RADGRP, Beta 1, is between 0.5566 and 0.8082, with a mean of 0.6844. This means that if you were to take the difference of the natural logs of any two results from the table with fixed age and a difference of 1 rad group, you would get 0.6844. For example, for rad level 2 and age group 4, 0.5553 out of every thousand people die, and for rad level 3, 1.0945 people/thousand die. If you took the differences of the natural logs:

$$\beta_1 = ln(1.0945) - ln(0.5553)$$

$$\beta_1 = 0.6785$$

0.6785 is very close to 0.6844, and rests right in the middle of the interval we found. Doing the same calculation for Beta 2 by keeping rad the same and changing age gets us a similar result:

$$\beta_2 = ln(0.3617) - ln(0.3320)$$

$$\beta_2 = 0.0857$$

This value falls right in the range of the 95% interval for Beta 2, and is fairly close to the mean value of 0.0753.

Beta Coefficients For Other Cancer Model

	Mean	Median	95% CI
β_0: Y intercept	-7.595	-7.595	(-7.8558, -7.3297)
β_1: RADGRP	0.0250	0.0249	(-0.0212, 0.0707)
β_2: AGEGRP	0.7106	0.7106	(0.6693, 0.7521)

Figure 3

Figure 3 shows us that age is a larger factor in deaths by other cancers than radiation is. This information helps us to see that the relationship between radiation and leukemia is in fact strong and valid.

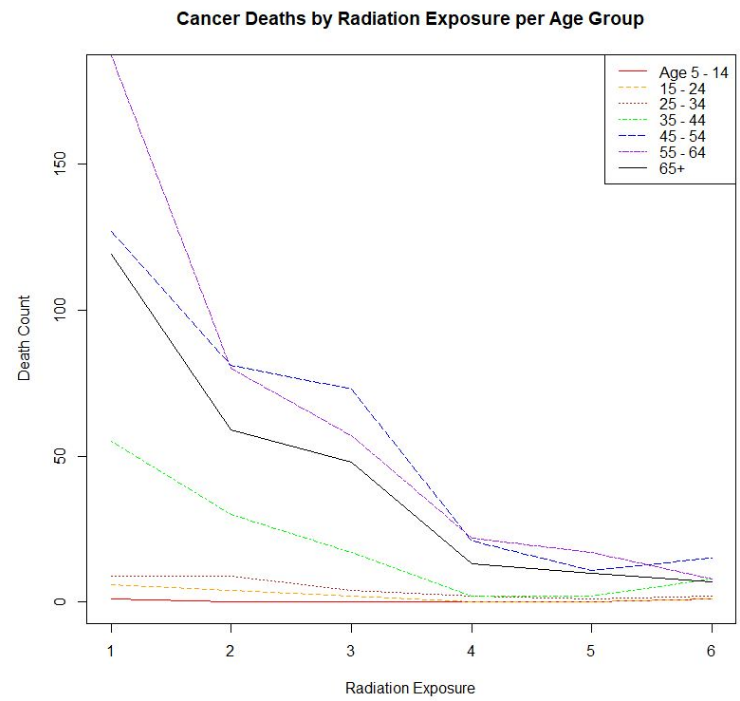

Figure 4

Figure 4 shows the relationship between radiation and death from cancer. Each line represents a different age group. We can see that age causes more deaths by cancer than radiation does. The purple, blue, and black lines represent the 45 and older population. These people had high cancer death rates even though they only had a small amount of radiation exposure. Figure 5 (below) shows us the relationship between leukemia death and radiation exposure. The lines represent the same age groups as they do in Figure 4.

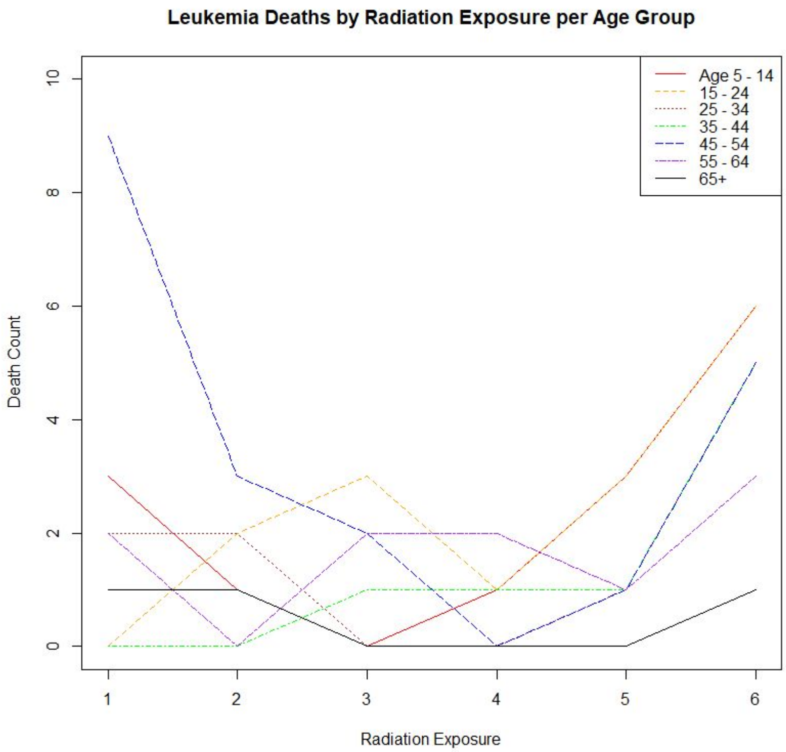

Figure 5

Discussion

Our analysis shows that there is a positive relationship between radiation exposure and death from leukemia. If we have two people in the same age group, we will see an increase in deaths by 0.6844 when we increase the radiation exposure group by one unit. We were able to validate that radiation exposure does not have a strong relationship with all cancers by creating a model with other cancers as our response variable and radiation exposure and age as our explanatory variables. The one limitation we may have is that it is possible for a more complex model to fit the data better, but for simplicity and variable interpretation purposes, we chose the above models for our analysis. These results are significant because we were able to find a positive relationship with radiation exposure and risk of death from leukemia. This means there may be a positive relationship between repeated CT scans and risk of death due to leukemia.