Survival of Pediatric Cancer Patients

By Teresa Cameron

Final Report

**Introduction**

From 1975 to 2006, there has been a 50% decrease in all malignant childhood cancer mortality rates. This reduction is due to advances in treatment and detection of cancer. Even with these advances, an estimated 1,960 deaths from cancer will occur annually among children and adolescents aged birth to 19 years. The overall 5-year survival rate for childhood cancers is approximately 80%, but it remains the second leading cause of death in children in developed countries.

**Description of the Problem**

The incidence rate of pediatric cancer in the United States has increased slightly at an annual rate of 0.6% since 1975. Figure 1 below shows the trends in pediatric cancer incidence rate by site from 1975 to 2010.

Chart, line chart

Description automatically generated**Figure 1**

The sudden increase in diagnoses of Brain and CNS cancers in the 1980s is thought to be due to the introduction of magnetic resonance imaging (MRI) and stereotactic biopsies that allowed for more accurate diagnosis.

Figure 2 below shows the trends in pediatric cancer mortality rates by site. As you can see, mortality has decreased.

Chart, line chart

Description automatically generated**Figure 2**

Although cancer rates are slightly increasing while mortality rates have been decreasing rapidly, there are still factors that influence survival that could be improved.

**Data**

The data used for this project is from the SEER (Surveillance, Epidemiology, and End Results) Program. The program began in 1973 and provides information regarding cancer statistics to help reduce the cancer burden among the U.S. population. The data was obtained from population-based registries covering approximately 48% of the United States population. Figure 3 shows the number and percentages of people represented in the data by state and race.

Table

Description automatically generated**Figure 3**

**Table from https://seer.cancer.gov/registries/data\_p.pdf**

There are a total of 263 variables available for analysis. Figure 4 shows the number of variables for each category of data.

Graphical user interface, text, application, table

Description automatically generated **Figure 4**

For this analysis, the variables analyzed were age, sex, year of diagnosis, race / origin, the county’s median household income and population where the patient lived, cause of death, and survival months.

This data was obtained using SEER\*Stat. This program allows a user to access the database and filter the data by variables. The variables chosen were those listed above with an age range of 0-17 years old and the behavior of the tumor as malignant. The query produced 69,263 rows of data with year diagnosed between the years 2000 – 2018. These years were generated because this is when the race variables that are most comprehensive were documented.

**Methods for Data Analysis**

**Approach**

The methods I used for analyzing this data was to compare the total counts and rates for survival with each variable. If a variable appeared to have a possible relationship with survival, a z-test was done to compare the survival rate among the different categories.

**Justification**

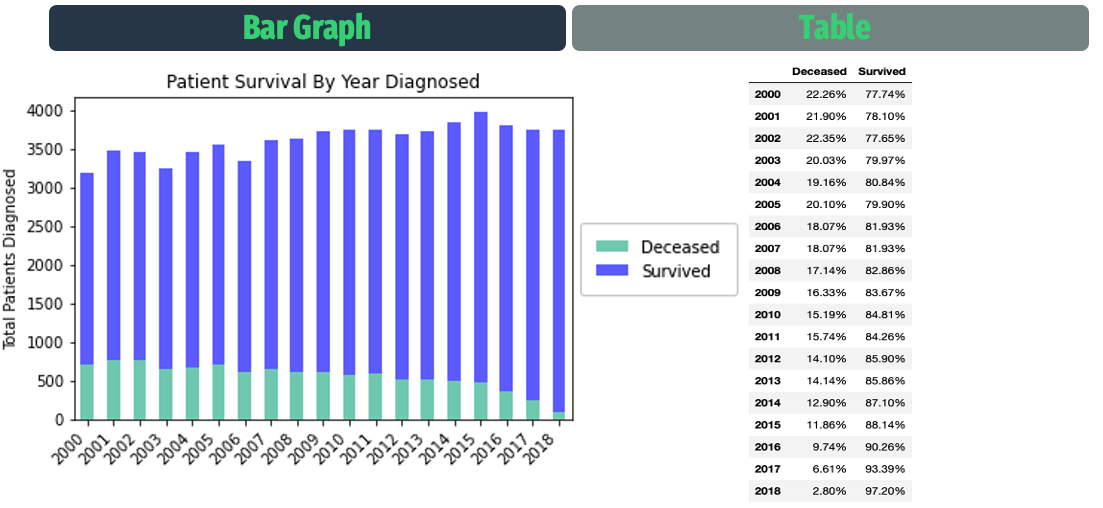
By examining both the total count and percentage, it can be determined which variables affect the greatest amount of people, but also detect differences in the variable categories.

A z-test was chosen to determine correlation because the sample size would be over 30.

**Results**

The first analysis was done on the count and percentages of patients who survived by year diagnosed. Figure 5 shows a bar graph with the total patients by year and the table has the percentages.

**Summary of Findings**

**Figure 5**

From the figure, we can confirm the findings mentioned above that the percentage of children who survive after being diagnosed with a malignant tumor has increased over the years. It is common practice to look at a five-year survival period and the data was most recently updated November 2021. For patient’s diagnosed in 2016, the percent who survived is 90.26%.

The second analysis was to examine the site of the cancer and number of patients who survived. Figure 6 shows the bar graph and table.

Graphical user interface, text, application

Description automatically generated**Figure 6**

The most common type of pediatric cancer is acute lymphocytic leukemia (ALL) followed by brain cancer. From the graph and table, it can be determined although most of the patients in the data are diagnosed with ALL, more patients do not survive brain cancer. The survival percentage is 90% for ALL while brain has 73%.

Figure 7 shows the total deaths from 2000 to 2018 by Cancer type. This graph gives a better understanding of just how many more patients do not survive brain cancer as it is almost double ALL.

Chart, bar chart

Description automatically generated

**Figure 7**

The data was next analyzed to determine which cancers had the highest and lowest percentages of patients who survived. Figures 8 and 9 are the graphs and tables showing this information.

Graphical user interface, text, application

Description automatically generated**Figure 8**

**Graphical user interface, text, application

Description automatically generated Figure 9**

Cancer of the Esophagus had no survivors and the rest of the cancer sites on this graph and table are associated with the colon. These cancers are rare in the dataset as there were three patients who had cancer of the esophagus. There were a few cancers that had 100% of the patients survive.

The next variable that was analyzed was age. Figure 10 is a histogram showing the distribution of the patient’s age by survival.

Chart, histogram

Description automatically generated

**Figure 10**

The distribution shows most of the patients were diagnosed when they were either younger or older. However, there is not much of a difference in the survival by age relative to the number of patients.

The age of the patient seems to have a relationship with the site of the tumor. Figure 11 is a boxplot displaying the median and distribution of the patient’s age by site of tumor.

Chart, box and whisker chart

Description automatically generated**Figure 11**

The median age for Leukemia patients is lower than others, but the distribution is skewed to the right indicating there are still many patients who are older, but the ages are more spread. Brain and soft tissue cancers have a median age of 7 and have even spreads in both directions. Hodgkin – Nodal and Bones and Joints have a higher median age of around 13 and 14, but there are a few younger patients who are outliers.

Figures 12 and 13 examine the patients’ county median household income and population size by survival.

Graphical user interface, application

Description automatically generated**Figure 12**

Graphical user interface, text

Description automatically generated**Figure 13**

The percent survival from the county’s median household income shows the highest survival rate is from households in the $65,000 – 69,000 range while the lowest is the $40,000 – 44,999. The difference is 2.76% suggesting there is not much of a difference. Plus, the percent difference between the groups is not consistent has it increases or decreases by range.

The counties with a population of 250,000 to one million had the highest survival percentage of 85% while the lowest was Alaska or Hawaii at 83%. The difference is 2.08% indicating there may not be a difference for county population. Although the last category indicated unknown, missing data, there was a category for unknown that was removed for this analysis.

The final demographic variable that was examined was race. Figure 14 shows the rate of patients in this data who survived based on their race.

Graphical user interface, application

Description automatically generated**Figure 14**

The highest survival percentage of patients are those in the category of Non-Hispanic White at 86%. The lowest are Non-Hispanic Black patients at just under 80%. The closest survival percentage category to Non-Hispanic White category is Hispanic patients at 84%. The difference is 2.58%.

Since this category seems to have a larger percent difference, a two-sample hypothesis test was done on this variable to determine statistical significance.

Samples from each category of race was selected and the survival rates were calculated. The rates from the samples were analyzed for approximate normal distribution with histograms. The z-score and p-values were calculated to compare the means of each sample to the sample mean of Non-Hispanic White patients.

The null hypothesis is the race of patients does not have an effect on survival rate and the alternative hypothesis is the race of patients affected their survival rate.

Figure 15 shows the distribution of the sample survival rates and the means of the samples are shown in Figure 16. The z-scores and p-values are found in Figure 17.

**Shape, polygon

Description automatically generatedFigure 15**

|  |  |
| --- | --- |
| Means of Samples | |
| Non-Hispanic White | 87% |
| Hispanic | 83% |
| American Indian / Alaska Native | 82% |
|  |
| Asian or Pacific Islander | 84% |  |
|  |
| Non-Hispanic Black | 79% |  |

**Figure 16**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Z-Scores | |  | P-Values | |
| Hispanic | 4.79 |  | Hispanic | 1.65E-06 |
| American Indian / Alaska Native | 5.21 |  | American Indian / Alaska Native | 1.89E-07 |
|  |
| Asian or Pacific Islander | 3.53 |  | Asian or Pacific Islander | 0.00041 |
|  |
| Non-Hispanic Black | 8.08 |  | Non-Hispanic Black | 6.40E-16 |

**Figure 17**

The z-scores indicate how many standard deviations a value is from the mean while the p-value measures the probability that the measured effect is due to randomness. The p-values are all lower than the alpha of 0.05 indicating we can reject the null hypothesis and accept the alternative hypothesis that the race of patients affected their survival rate.

**Discussion**

Although we can conclude from the hypothesis test that race affected the survival rate in the samples, we cannot establish causation. This also does not indicate the correlation can be found in the population. Additional tests need to be done to make this correlation.

Additional areas of focus would be on performing the same tests but standardizing the other variables. For example, choosing only patients with the same site of tumor, from counties that have the same median household income range and population. It would also be interesting to compare the tumor morphology and treatments given to these patients. These variables are more complex, and it would take additional time to fully comprehend the role they play in survival.

**Conclusion**

Although there has been much progress in the survival rate of pediatric patients, there is still room for improvement. By examining the SEER data, researchers can gain a better understanding where these improvements should be focused. By examining the survival of patients according to tumor site, advances should be focused on better treatment and detection of brain cancer. A patient’s age can help direct doctors what are the most likely areas a tumor might be located. Resources also can be allocated to assist patients with certain demographics so they may have the same chance of survival as those in other categories.

Children who are diagnosed with cancer deserve to have every avenue explored so they are given the best chance to survive.

References

Kaatsch P. Epidemiology of childhood cancer. Cancer Treat Rev. 2010 Jun;36(4):277-85. doi: 10.1016/j.ctrv.2010.02.003. Epub 2010 Mar 15. PMID: 20231056.

Kinsey SE, Picton SV. Childhood cancer in high resource settings. Cancer Epidemiol. 2021 Apr;71(Pt B):101838. doi: 10.1016/j.canep.2020.101838. Epub 2021 Jan 22. PMID: 33495096.

Ward, E., DeSantis, C., Robbins, A., Kohler, B. and Jemal, A. (2014), Childhood and adolescent cancer statistics, 2014. CA A Cancer Journal for Clinicians, 64: 83-103. <https://doi.org/10.3322/caac.21219>