A Retrospective Statistical Analysis of Second Primary Cancers in the Delmarva Peninsula, U.S.A.

Marvin Li[1] and Veera Holdai[2]

[1] Bennett Middle School, Fruitland, Maryland.

[2] Department of Mathematics and Computer Science, Salisbury University, Salisbury, Maryland.

Summary

A significant percentage of cancer survivors develop a second primary cancer. Using data of deceased patients provided by the Peninsula Regional Medical Center, a retrospective statistical analysis was conducted to investigate whether the type of the first cancer affects the occurrence time and type of the second primary cancer. Cancer patients were stratified according to the first cancer type, and the time elapsed between the first and second cancer diagnosis was examined to see if there are statistically significant differences. Histograms of second cancer occurrence times for lung and bronchus, breast and melanoma-skin cancers are strongly skewed right while those for prostate and urinary bladder cancers have lower right skewness. Both one-way analysis of variance test and Kruskal-Wallis test show a p-value below the significance level of 0.05, confirming that the first cancer type affects the second cancer occurrence time. For each first cancer type, the most common types of the second primary cancer are identified. About 40% of breast cancer and lung and bronchus cancer survivors develop a second primary cancer of the same type. 25% of melanoma-skin cancer survivors develop the same cancer again and 17% of them develop a second melanoma cancer. In contrast, less than 1% of prostate cancer survivors develop a second prostate cancer and only 14% of urinary bladder cancer survivors develop the same cancer again. Lung and bronchus cancer is one of the most prevalent second cancer types regardless of the first cancer type.

Introduction

The cancer survival rate has doubled since the 1970's due to earlier detection and better treatment methods and techniques (1, 2). The 5-year relative survival rate among all cancer patients is 66% (3). This, in combination with the growing number of people diagnosed with cancer due to population growth, results in a larger number of cancer survivors (4). However, a significant percentage of cancer survivors develop a second primary cancer within a few years (5). In general, cancer patients are more likely to develop another second primary cancer compared to the general population's chance of developing a first primary cancer (6). The number of patients with multiple primary cancers is growing rapidly, accounting for 16% of cancer cases reported to the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) Program in the U.S. (5). Moreover, second cancers are a leading cause of mortality among several populations of long-term survivors (7).

Among other things, second cancers may be the result of lifestyle choices, genetics, environmental exposures, and effects of previous treatments (2, 5, 8). For cancer survivors, quantifying the risks of second malignancies has important implications for screening and prevention strategies (9, 10). Identifying the second cancer, which has an elevated likelihood of occurrence after a particular first cancer, can shed light on possible shared etiologies and mechanisms of carcinogenesis (11, 12).

There have been a number of studies on second primary cancers, mostly focusing on the standardized incidence ratios (SIRs). The SEER Program of the National Cancer Institute in the U.S. provided data on over 2 million cancer patients over a follow-up period between 1973 and 2000. It was found that cancer survivors had a 14% higher risk of developing a new malignancy than would have been expected in the general SEER population. A total of 185,407 new primary cancers were observed compared with 162,602 expected (6). A retrospective cohort study on first cancer survivors in Queensland, Australia stratified SIRs of patients by age, sex, type of first primary cancer and type of second primary cancer etc. (4). Both males and females across all age groups had a statistically significant higher chance of developing a second primary cancer relative to the SIRs in the general population. The larger SIRs among the cancer survivors were attributed to factors such as similar etiologies, genetics, and the effects of treatment.

A Japanese study tracked patients who had been diagnosed with a first primary cancer and registered with the Nagasaki Prefecture Cancer Registry between 1985 and 2007 (13). The SIRs were stratified according to the site and the years after diagnosis of the first primary cancer. Some specific relationships were observed between the sites with risk factors in common, such as smoking, drinking, and hormone status. Similar large-scale studies have been conducted in Italy (14) and France (15). A recent systematic review summarized population-based studies of multiple primary cancers recorded in Medline/PubMed and Embase databases from inception to August 2016 (16). For all first cancer sites combined, a higher rate of multiple primary cancers was reported in more recent calendar periods as compared to earlier calendar periods in four of the six relevant studies. The SIRs ranged from 1.14 in the early 1980s to 1.21–1.46 in the late 1990s in the USA and Australia. The two studies from Italy (14) and France (15) showed no significant difference in SIRs across time periods 1978–2010 and 1989–2004.

Many previous studies have focused on estimating the cancer incidence rates among first cancer survivors but have not directly addressed the question of the time elapsed between the first and second cancer diagnoses. This research analyzes this occurrence time of the second primary cancer and examines if there are any connections between the second and first cancers. We hypothesize that the second cancer occurrence times depend on the type of the first cancer. We also hypothesize that the second primary cancer type may be related to the first cancer type for certain types of the first cancer.

The data analyzed in this study were obtained from the Richard A. Henson Cancer Institute of the Peninsula Regional Medical Center. It is located in the middle of the Delmarva Peninsula on the East Coast of the United States, serving patients in southern Delaware as well as the eastern shore of Maryland and eastern shore of Virginia. The Henson Institute collected cancer patients' data between 2000 and 2014 and provided them to the authors through a joint research program with Salisbury University.

Results

A total of 15,390 cancer cases were reported, most of which (89%) were single cancer cases. 10.3%, or 1625 patients, were diagnosed with two primary cancers (**Figure 1**). In the dataset of patients with two primary cancers, there were 688 males (54.4%) and 577 females (45.6%) (**Table 1**). Thirty-two (2.5%) of the patients were under 40 years old, 714 (56.4%) between 40 and 70 years old, and 519 (41%) greater than 70 years old.

To test the hypothesis that the second cancer occurrence

times depend on the type of the first cancer, we conducted a series of statistical analyses. First, we compared statistical features of the second cancer occurrence times for different types of the first cancer using box-whisker plots, histograms, and confidence intervals. Second, a one-way analysis of variance (ANOVA) test was conducted to determine whether the means of the second cancer occurrence times differ among the types of the first cancer. The ANOVA test is extended by a post-hoc Tukey Honestly Significant Difference (HSD) multiple comparisons of means analysis to determine which first cancer type group means differ from each other. The Tukey HSD test assumes that observations are independent among the groups and that the data
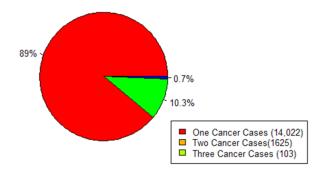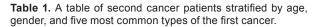

Figure 1. Percentage and the total number of patients with one, two, or three primary cancers reported to the Richard A. Henson Cancer Institute between 2000 - 2014.

form a normal distribution. Since the occurrence times may not have a normal distribution, a non-parametric Kruskal-Wallis test was conducted to verify the ANOVA test result. The Kruskal-Wallis test does not require a normal distribution of the residuals, but assumes that the groups are independent. Third, the most common types of the second primary cancers were determined for each type of the first cancer, and their potential connections to the first cancer type were examined.

As shown in **Figure 2**, the five most prevalent types of first cancers in our sample of patients diagnosed with two primary cancers are breast cancer (14.5%), lung and bronchus cancer (12.3%), prostate cancer (11.2%), urinary bladder cancer (7.3%) and melanoma-skin cancer (7%). Carcinomas in-situ (cancer cells that line the internal organs) account for 6.6% of cancers, melanoma cancer accounts for 5.1%, and kidney and renal pelvis cancers account for 5% within the two-cancer patients. All other cancer types have percentages less than 5% and sum to approximately 31%. In order to have adequate sample size, our statistical analyses focused on the patients with two cancers whose first cancer type is one of the five most common types listed above.

The median values of the second cancer occurrence

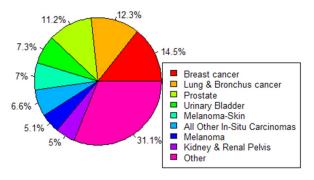|	Total	Age at first diagnosis (<40 year old)	Age at first diagnosis (>= 40 years old & < 70 year old)	Age at first diagnosis (>= 70 years old)
Total	1,265	32	714	519
Male	688	5	391	292
Female	577	27	323	227
Breast cancer	183	6	116	61
Lung & Bronchus cancer	155	0	83	72
Prostate cancer	142	0	88	53
Urinary Bladder cancer	92	0	42	50
Melanoma -Skin cancer	88	10	51	27

Table 1. A table of second cancer patients stratified by age, gender, and five most common types of the first cancer.

Figure 2. Most prevalent types of first cancer among the patients with two primary cancers.

times are shown in **Figure 3**: breast cancer, lung and bronchus cancer, melanoma-skin cancer, urinary bladder cancer, and prostate cancer. Lung and bronchus and breast cancers not only have similar median values (20.5 and 23.5 months, respectively) but also have similar inter-quartile ranges (3.2, 57.1) months and (3.1, 60.5) months respectively. The other three first cancer types have larger median values as well as larger inter-quartile ranges. Melanoma-skin cancer has a median second cancer occurrence time of 32.4 months and the largest inter-quartile range of 69.5 months (6.6, 76.5). Urinary bladder cancer has a median second cancer occurrence time of 45.3 months and a smaller inter-quartile range of 62.8 (11.8, 75.8) months. The median value of the second cancer occurrence for prostate cancer is 57.2 months and lies almost mid-way between the upper and lower limit of the inter-quartile range.

Figure 4 shows the histograms of the second cancer occurrence times when the patients are categorized according to the type of their first cancer. Importantly, none of the occurrence times display a normal

distribution. The histograms for lung and bronchus cancer, breast cancer and melanoma-skin cancer are all skewed right (positive skewness), indicating that a high percentage of the patients are diagnosed with the second primary cancer within the first 32.9 months. The degree of skewness is different among these three cancer types: lung and bronchus cancer (0.83), breast cancer (0.69), and melanoma-skin cancer (0.71). When the first cancer is lung and bronchus cancer, nearly half (46%) of the patients develop the second cancer within 16.4 months and a very small percentage (3.9%) of patients are diagnosed with the second cancer at a date later than 98.7 months (**Figure 4B**). Although the histogram for breast cancer is also skewed right, with 41% of the second cancer cases occurring within the 16.3 months of the first cancer diagnosis, there are a comparatively larger number of outliers at longer time intervals between the two cancer diagnoses (**Figure 4A**). For example, 23% of the second cancers occur 65.6 months after the first cancer. The histogram for melanoma-skin cancer resembles a linear relationship (**Figure 4E**). The skewness for melanoma-skin cancer is less than lung and bronchus cancer and slightly larger than breast cancer. The histograms for urinary bladder cancer and prostate cancer appear to have different frequency distributions from the above three cancers (**Figure 4C-D**). The skewness is significantly lower compared to the other cancer types: urinary bladder (0.48) and prostate (0.28). When the first cancer is prostate cancer, the second cancer occurrence time displays a relatively uniform distribution from 32.8 to 114.8 months. This suggests that the second cancer diagnoses occur with a similar frequency throughout 32.8 to 114.8 months. The skewness in the prostate cancer histogram is the lowest of all five cancers. Such a
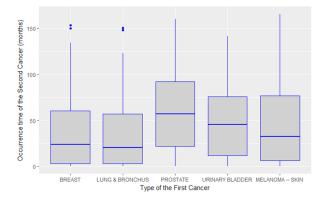

Figure 3. Box and whisker plots of the second primary cancer occurrence times for five major types of the first cancer. The inner box displays the interquartile (25th to 75th quartile), and the whisker lines extending from the interquartile represent the full range of values (at significance level of 0.05). The dots display the outliers.
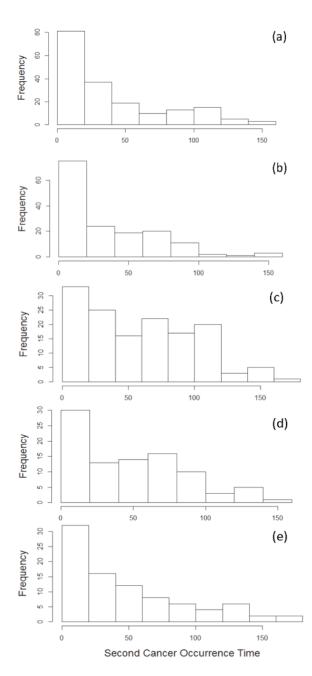
Figure 4. Histogram of the second primary cancer occurrence time (in months) for five major first cancer types. (a) Breast Cancer; (b) Lung and Bronchus Cancer; (c) Prostate Cancer; (d) Urinary Bladder Cancer; (e) Melanoma-Skin Cancer.

dramatic difference in the histogram between the two groups (lung and bronchus, breast and melanoma-skin cancers versus prostate and urinary bladder cancers) and the large range of skewness values (0.28-0.83) raises a question about possible connections between the first and second primary cancers. A large right skewness may indicate a connection between the first and second cancers whereas a uniform frequency

distribution is a sign of weak connection between the two primary cancers. This connection was further explored when we compared the types of the first and second primary cancers later.

The confidence interval provides a visualization of the range of the true mean with a 95% confidence. Breast and lung and bronchus cancers have similar sample means around 36.2 months, melanoma-skin and urinary bladder cancers have similar sample means around 48.0 months, and prostate cancer has a sample mean around 58.7 months (**Figure 5**). However, the confidence intervals show significant overlaps: the only exception is between lung and bronchus cancer and prostate cancer. An ANOVA test was conducted to determine if the means of the second cancer occurrence times differed among the types of first cancer. The ANOVA test is superior to pairwise t-test because it limits Type I error in multiple sample groups (17).

The five sample groups are the second cancer occurrence times stratified according to the first cancer type: breast cancer (its mean denoted as μ_1), lung and bronchus cancer (μ_2), prostate cancer (μ_3), urinary bladder cancer (μ_4), or melanoma-skin cancer (μ_5). Our null hypothesis stated that there would be no differences between the groups that were tested, namely that the mean occurrence times would be the same ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$). The alternative hypothesis stated that there would be a difference between the five groups. **Table 1** summarizes the ANOVA test result of the second cancer occurrence times. The F-statistic, the ratio of the between-group variance to the within-group variance, is equal to 8.61 (degrees of freedom: 4). The p-value for this test was $p = 8.9e-07$. At a significance level of $\alpha = 0.05$, we rejected the null hypothesis. Since a majority of the histograms are skewed and do not follow a normal distribution, a non-parametric Kruskal-Wallis test was performed. It confirms the ANOVA result, with $p = 2.1e-$
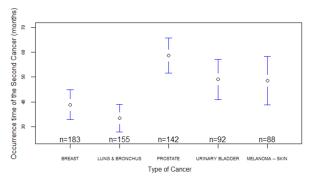

Figure 5. The 95% confidence interval of the second primary cancer occurrence time showing the sample mean (open circles), and the upper and lower bounds of the real mean with a 95% confidence. The data are grouped according to the types of the first cancer. The sample size n is marked along the horizontal axis.

07 at a significance level of α = 0.05. Therefore, there are statistically significant differences in the mean of the second cancer occurrence times among the five groups.

To determine if there are differences in the mean occurrence times of the second cancer between any pair in the five groups, we conduct pairwise tests using the post-hoc Tukey HSD multiple comparisons of means algorithm (18). The *p*-value is less than the significance level of α = 0.05 for the following pairs: prostate and breast cancers (p=0.00014); prostate and lung and bronchus cancers (p=0.0000011); and urinary bladder

cancer cases are lumped together in a category called "others." There are clear differences in the distribution of the second cancer types among the five pie charts. A significant percentage of patients with breast cancer or lung and bronchus cancer as a first cancer develop a second cancer of the same respective type (41% for breast cancer and 38% for lung and bronchus cancer, respectively.). When the first cancer type is melanoma-skin cancer, 25% of the patients develop the same cancer again and 17% develop melanoma cancer as the second cancer. 14% of urinary and bladder cancer
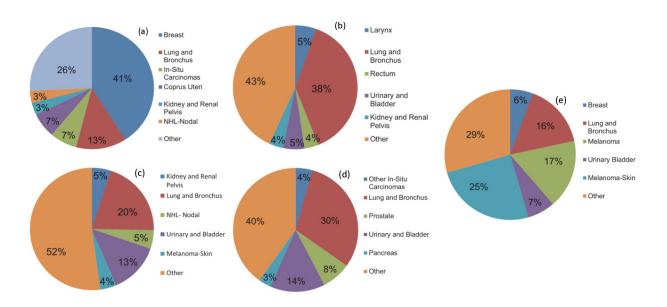

Figure 6. Proportion of second cancer types for different types of the first cancer. (a) Breast Cancer; (b) Lung and Bronchus Cancer; (c) Prostate Cancer; (d) Urinary-Bladder Cancer; (e) Melanoma Skin.

and lung and bronchus (p=0.03). These cancer type pairs have statistically significant differences (**Table 2**). The pair of melanoma-skin and lung and bronchus cancers has a *p*-value of 0.043. Other pairs have p-values exceeding α = 0.05, suggesting no significant differences in the mean occurrence times of second cancer.

Results from the post-hoc Tukey analyses are visualized in the box and whiskers plots in **Figure 3** and confidence interval plots in **Figure 5**. For example, breast cancer and lung and bronchus cancer have similar sample means while urinary bladder cancer and melanoma-skin cancer have similar sample means.

These analyses have focused on the second cancer occurrence times. We then identified the most common types of the second primary cancers for a given type of the first cancer. For each first cancer type, the pie charts show the percentages of most common types of second cancers (**Figure 6**). All cancer types with < 3% of the

patients develop the same cancer again. In contrast, very few patients with a first cancer of prostate cancer develop a second prostate cancer (<1%). Other than developing the same cancer again, Lung and bronchus does not have a strong correlation with the development of a particular second cancer . None of the remaining cancer types exceed 5%. However, 13% of breast cancer survivors developed lung and bronchus cancer as the second cancer. In fact, lung and bronchus cancer was one of the most prevalent second cancer types when the first cancer was prostate cancer (20% of the cases), urinary bladder (30% of the cases), and melanoma-skin cancer (16% of the cases).

Discussion

Our statistical analyses have shown that there are significant differences in the second cancer occurrence times for the five most common types of first cancers. The occurrence times for the diagnosis of a second cancer

	Difference in Means	*p*-value
ANOVA (BREAST, LUNG & BRONCHUS, PROSTATE, URINARY BLADDER, MELANOMA-SKIN)		0.00000089
LUNG & BRONCHUS versus BREAST	-5.4518515	0.7337271
MELANOMA–SKIN versus BREAST	9.6554872	0.3551718
PROSTATE versus BREAST	19.8305459	0.0001418
URINARY BLADDER versus BREAST	10.1556684	0.2883205
MELANOMA–SKIN versus LUNG & BRONCHUS	15.1073387	0.0432680
PROSTATE versus LUNG & BRONCHUS	25.2823974	0.0000011
URINARY BLADDER versus LUNG & BRONCHUS	15.6075199	0.0295804
PROSTATE versus MELANOMA–SKIN	10.1750587	0.3472560
URINARY BLADDER versus MELANOMA – SKIN	0.5001812	0.9999894
URINARY BLADDER versus PROSTATE	-9.6748775	0.3857791

Table 2. ANOVA test on the second primary cancer occurrence time for five major types of the first cancer (first row). We performed Tukey HSD Post-hoc tests on pair-wise comparisons of two cancer types. The first column represents possible pairs of the first cancer types. The second column represents the calculated difference in the means of the cancer pairs. The third column displays the calculated *p*-value of cancer pairs.

break into three groups: lung and bronchus and breast cancers with shortest times; melanoma-skin and urinary bladder cancers in the middle, and prostate cancer with longest times. In addition to these differences in the mean occurrence time, the time distribution of the second cancer occurrence times also show marked differences: the histogram is strongly skewed right when the first cancer is lung and bronchus, breast, melanoma-skin but has low skewness for urinary bladder and prostate cancers. Based on these analyses and significance testing using ANOVA and Kruskal-Wallis tests, we conclude that there is a statistical correlation between first cancer types and second cancer occurrence times.

Additionally, we identified the most common types of the second cancer for each of the five major types of the first cancer. Lung and bronchus, breast and melanoma-skin cancers have a large chance of developing a second cancer of the same type, suggesting a possible link between the first and second primary cancers. However, prostate and urinary bladder cancers have a near zero chance of developing a second cancer of the same type, indicating a weak link between the two primary cancers. Urinary bladder cancer is the second most common type of the second primary cancer when the first cancer is urinary bladder cancer or prostate cancer.

Our findings show that the first cancer type affects the occurrence time and the type of the second cancer and may suggest similar etiologies and genetic makeup between the first and second primary cancers (5). A Japanese study (13) also found strong site relationships between the first and second primary cancers. Previous studies also suggested other possible causes of second primary cancer, such as the age of the first cancer diagnosis (19), behavior and lifestyle choices (15,20), environment (2,12) and treatments used for the first cancer (5,21). The prevalence of lung and bronchus cancers in the second primary cancer is likely related to the high percentage of smokers on the eastern shore of Maryland (22).

A few factors may have influenced our statistical analysis. Some distributions of the second cancer occurrence times stratified by the first cancer type are skewed. They violate the normal distribution assumption in the ANOVA test, although the sample sizes (ranging from 88 to 183) are adequately large. The non-parametric Kruskal-Wallis test gives essentially the same results and gives confidence to the ANOVA test results. When the second primary cancers are further stratified by their type, such as those in **Figure 5**, some sample sizes fall below the widely used limit of 30. As more data becomes available at Richard A. Henson Cancer Center or a large data set is analyzed, we would be able to conduct a statistical testing of the second cancer type. One interesting extension of the current research is to develop statistical models to predict the occurrence times and types of the second primary cancers. For example, Luciana *et al.* created a decision tree to predict the second cancer type based on its various contributing factors (12). There is an overwhelming amount and variety of second cancer types, so a decision tree may not accurately predict the second cancer type without running the risk of overfitting. Sophisticated statistical

tools such as machine learning may be exploited to predict the occurrence times and types of the second primary cancers. Another venue for future research is to make use of these statistical results to design improved screening and prevention strategies for first cancer survivors (10).

Methods

Our retrospective analysis is based on a non-random and large sample of deceased cancer patients from the Peninsula Regional Medical Center (**Table 1**; 22). No Independent Review Board approval was needed. Primary cancers were found through histologic diagnoses, although some second cancers occurred a few months after the first cancer diagnosis. The last date for reporting a first cancer was 12/17/2012 and the last date for reporting a second cancer was 9/8/2014.

Dataset Organization

For each patient, the date and age at the first cancer diagnosis as well as the site of the first cancer were recorded. In addition, the following data of the second cancer was recorded: the date and age of the second primary cancer diagnosis and the site of the second primary cancer. The days and months between the first and second primary cancer diagnoses were also reported. Our analyses focus on the occurrence time of the second primary cancer (i.e. the months between the first and second primary cancer diagnoses) and the site of the second primary cancer. Patients' names were removed by Peninsula Regional Medical Center administration to protect their privacy.

The data are organized according to the type of the first cancer. The months between the first and second primary cancer diagnoses are extracted to calculate the second cancer occurrence time for a specific type of the first cancer. Similarly, the second cancer types are stratified according to the type for the first cancer.

Data Analysis

The cancer patients' data were stored in an excel table and analyzed using the statistical programming language R (23). To describe each subset of the data on the second cancer occurrence times, we calculated the following statistics. First, the box and whiskers plots were produced in R using the ggplots2 library (24). Second, a histogram was used to study the distribution of the occurrence times of the second primary cancers. It was used to visualize and identify the distribution of the data. Skewness was used to measure the departure from the normal distribution. A positive skewness value indicates that the distribution is skewed right. A negative skewness value indicates that the distribution is skewed left. Third, the confidence interval for the second cancer

occurrence times was calculated for different types of the first cancer and graphed using the gplots library (25) in R. The confidence interval shows the range of possible values of the true mean, with a 95% confidence that the estimate is correct. The bottom value represents the lowest possible value of the true mean with 95% confidence, the middle value represents the sample mean, and the top value represents the largest possible value of the true mean with 95% confidence.

An ANOVA test was used to determine if there are statistically significant differences in the second cancer occurrence time between different types of the first cancer. ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes the t-test to more than two groups. It is useful for comparing (testing) three or more means (groups or variables) for statistical significance. The ANOVA test, which limits the Type I error, compares multiple distributions and gives an F-value. The ANOVA analysis is extended with a Tukey HSD post-hoc test (Implemented in R, function "TukeyHSD()") which finds the p-values for each pair of first cancer types from the ANOVA test. The calculated p-value is used to determine whether the pair of distributions is statistically different or not. A rank-based nonparametric Kruskal-Wallis test, created in R (function "kruskal.test()"), is also conducted to determine if there are statistically significant differences in the second cancer occurrence times among different types of the first cancers. It generalizes comparison algorithms to more than two groups. It is used to verify the results of the ANOVA test.

Pie charts were created to visualize the most prevalent types of the second primary cancer for each type of the first cancer. The pie charts visualize what percentage of values of a set is in a subset. They are used to visualize and compare the types of the second cancer for different types of the first cancer.

References

1. Hayat, M. J., N. Howlader, M. E. Reichman, and B. K. Edwards. "Cancer Statistics, Trends, and Multiple Primary Cancer Analyses from the Surveillance, Epidemiology, and End Results (SEER) Program." *The Oncologist*, vol. 12, no. 1, 2007, pp. 20-37.
2. Mariotto, A. B., J. H. Rowland, L. A.G. Ries, S. Scoppa, and E. J. Feuer. "Multiple Cancer Prevalence: A Growing Challenge in Long-term Survivorship." *Cancer Epidemiology Biomarkers & Prevention*, vol. 16, no. 3, 2007, pp. 566-571.
3. Ries L.A.G., D. Melbert, M. Krapcho, A. Mariotto,

B.A. Miller, E.J. Feuer, L. Clegg, M.J. Horner, N. Howlader, M.P. Eisner, M. Reichman, B.K. Edwards (eds). "SEER Cancer Statistics Review, 1975-2004." National Cancer Institute, http://seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER web site, 2007.

4. Youlden, D. R., and P. D. Baade. "The relative risk of second primary cancers in Queensland, Australia: a retrospective cohort study." *BMC Cancer* vol. 11, no. 1, 2011, n. pag.

5. Travis, L. B. "The Epidemiology of Second Primary Cancers." *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, no. 11, 2006, pp. 2020-2026.

6. Curtis R.,E., D.M. Freedman R.E, L.A.G. Ries, D.G. Hacker, B.K. Edwards, M.A. Tucker, J.F. Fraumeni Jr. (eds). "New Malignancies Among Cancer Survivors: SEER Cancer Registries, 1973-2000." National Cancer Institute. *NIH Pub*, n. vol., no. 05-5302, 2006, n. pp.

7. Soerjomataram, I., and J.W. Coebergh. "Epidemiology of Multiple Primary Cancers." *Methods in Molecular Biology Cancer Epidemiology*, vol. 475, n. no., 2009, pp. 85-105.

8. Simard, E. P., El. M. Ward, R. Siegel, and A. Jemal. "Cancers with increasing incidence trends in the United States: 1999 through 2008." *CA: A Cancer Journal for Clinicians*, vol. 62, no. 2, 2012, pp. 118-28.

9. Ng, A.K., and L.B. Travis. "Subsequent Malignant Neoplasms in Cancer Survivors." The Cancer Journal, vol. 14, no. 6, 2008, pp. 429-434.

10. Corkum M, J.A. Hayden, G. Kephart, R. Urquhart, C. Schlievert, G. Porter. "Screening for new primary cancers in cancer survivors compared to noncancer controls: a systematic review and meta-analysis." *J Cancer Surviv*, vol. 7, no. 3, 2013, pp. 455–463.

11. Hemminki K, Boffetta P. "Multiple primary cancers as clues to environmental and heritable causes of cancer and mechanisms of carcinogenesis." *IARC Sci Publ*, n. vol., no. 157, 2004 pp. 289-297.

12. Luciani, A., G. Ascione, D. Marussi, S. Oldani, S. Caldiera, S. Bozzoni, C. Codecà, S. Zonato, D. Ferrari, and P. Foa. "Clinical analysis of multiple primary malignancies in the elderly." *Medical Oncology*, vol. 26, no. 1, 2009, pp. 27-31.

13. Utada, M., Y. Ohno, M. Hori, and M. Soda. "Incidence of multiple primary cancers and interval between first and second primary cancers." *Cancer Science*, vol. 105, no. 7, 2014, pp. 890-896.

14. AIRTUM Working Group. "Italian cancer figures, report 2013: Multiple tumours." *Epidemiol Prev*, vol. 37, no. 4-5, 2013, pp. 100-152.

15. Jegu, J., M. Colonna, L. Daubisse-Marliac, Tretarre B, Ganry O, Guizard AV, Bara S, Troussard X, Bouvier V, Woronoff AS, et al. "The effect of patient characteristics on second primary cancer risk in France." *BMC Cancer*, vol. 14, no. 94, 2014, n. pp.

16. Ye, Y., A. L. Neil, K. E. Wills, and A. J. Venn. "Temporal trends in the risk of developing multiple primary cancers: a systematic review." *BMC Cancer*, vol. 16, no. 1, 2016, n. pp.

17. Analysis of variance Wikipedia. *Wikimedia Foundation*, Web. 22 Dec. 2016.

18. Tukey's_range_test Wikipedia. *Wikimedia Foundation*, Web. 30 Dec. 2016.

19. Tsukuma, H., Fujimoto I, Hanai A, Hiyama T, Kitagawa T, Kinoshita N. "Incidence of second primary cancers in Osaka residents, Japan, with special reference to cumulative and relative risks." *Jpn J Cancer Res*, vol. 85, no. 4, 1994, pp. 339-345.

20. Fendrich, V., J. Waldmann, D. K. Bartsch, K. Schlosser, M. Rothmund, and B. Gerdes. "Multiple primary malignancies in patients with sporadic pancreatic endocrine tumors." *Journal of Surgical Oncology*, vol. 97, no. 7, 2008, pp. 592-595.

21. Newhauser, W.D., and M. Durante. "Assessing the risk of second malignancies after modern radiotherapy." *Nature Reviews Cancer*, vol, 11, no. 6, 2011, pp. 438-448.

22. Richard A. Henson Cancer Institute. 2014. Annual Cancer report. Peninsula Regional Medical Center, 20 pp. Web.

23. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org. Computing programming language used.

24. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. 2009. Library for building plots.

25. Warnes, G.R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz and B. Venables (2016). gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. https://CRAN.R-project.org/package=gplots. Library for building plots.