**Do sex, age, and whether the tumor is IDH1 mutated impact the survival times of people with glioblastoma multiforme and, if so, how?**

Introduction:

Glioblastoma multiforme (GBM) is the most common and aggressive malignant primary brain tumor. More than 40% of malignant primary brain tumors are GBM and few risk factors have been identified that may lead to this form of cancer. Less than 5% of patients survive more than 5 years after diagnosis (Kanderi and Gupta, 2021). Considering the aggressiveness of this cancer, it is important to identify some factors that may have a positive or negative effect on survival time since ensuing investigations of the biological mechanisms behind these correlations may be beneficial for the development of future treatments. Moreover, prediction of outcomes for individual patients is not yet possible despite many clinical trials (Walid, 2008).

Literature Review:

Previous studies have found several factors possibly associated with longer survival times of GBM patients. Trifiletti et al. suggest that the female sex is associated with longer survival whilst both McLendon and Halperin and Chandler et al. do not note such a relationship (2017) (2003) (1993). Trifiletti et al. studied 27,000 GBM cases using the National Cancer Database. The other two studies were conducted on much smaller scales, 766 and 449 patients respectively. However, Trifiletti et al. did not review the clinical histories of all patients in detail; McLendon and Halperin note that a detailed review of GBM patients’ clinical histories is important because the review allows the researchers to identify patients whose GBM developed from other types of cancer, which may disqualify some patients considered GBM long-term survivors (2003). Although this project will not take into consideration the clinical histories of the cases studies either, it will nevertheless attempt to assess whether the female sex is associated with longer survival times to contribute more evidence to the existing ambiguous evidence.

Additionally, Yang et al. found a correlation between longer survival and IDH1 mutated GBMs, meaning that the gene IDH1 is one of the tumor’s affected genes (2019). A study by Schiffgens et al. found a statistically significant effect of IDH1 mutated tumors on survival only for male patients, not for females, however, it analysed only 4 males and 5 females with IDH1 (2016). Given this evidence, we will evaluate previous findings about the impact of IDH1 mutated tumors in general and the possibly differing impact of IDH1 mutated tumors on survival depending on sex.

Lastly, Tamimi and Juweid state that young age at the time of diagnosis is a predictor of longer survival In GBM patients (2017). However, the strength of this relationship is not discussed. Thus, we seek to establish how strong the relationship between age and survival time is.

Research Question:

Considering these findings and limitations of the aforementioned studies, this project seeks to answer four sub-questions that, together, answer the main research question.

* Investigation 1: Are there statistically significant differences in survival time between men and women with GBM?
* Investigation 2: Are there statistically significant differences in survival time between people with and without IDH1 mutated tumors?
* Investigation 3: Does the impact of an IDH1 mutated tumor differ depending on sex?
  + Investigation 3.0: Are there statistically significant differences in survival time between men and women with IDH1 mutated tumors?
  + Investigation 3.1: Are there statistically significant differences in survival time between men with and without IDH1 mutated tumors?
  + Investigation 3.2: Are there statistically significant differences in survival time between women with and without IDH1 mutated tumors?
* Investigation 4: How much of the variance in survival times can be explained by age at diagnosis?

The Genomics Data Commons (GDC) Data Portal is used to obtain the data for this study (National Cancer Institute, 2021). It provides data on GBM cases including the affected genes and demographics.

Presentation of Data:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Location** | | **Spread** | | | | | | | | |
|  | Mean | Median | Standard Deviation | Max | Min | Range | Upper Quartile | Lower Quartile | IQR | Upper Tukey Fence | Lower Tukey Fence |
| **Dataset 1: Ages at Diagnosis of 593 People and Survival Times of 492 People with GBM** | | | | | | | | | | | |
| **Age at Diagnosis**  **(Years)** | 57.8 | 59 | 14.4 | 89 | 10 | 79 | 68 | 50 | 18 | 95 | 23 |
| **Survival (Days)** | 504 | 382 | 538 | 3881 | 3 | 3878 | 608 | 172 | 436 | 1261 | -481 |
| **Sample 1: Survival Times of 308 Men with GBM** | | | | | | | | | | | |
| **Survival (Days)** | 486 | 383 | 491 | 3524 | 3 | 3521 | 589 | 178 | 411 | 1205 | -439 |
| **Sample 2: Survival Times of 184 Women with GBM** | | | | | | | | | | | |
| **Survival (Days)** | 535 | 375 | 609 | 3881 | 6 | 3875 | 636 | 165 | 471 | 1342 | -541 |
| **Sample 3: Survival Times of 492 People (Whose Tumors Are Not IDH1 Mutant)** | | | | | | | | | | | |
| **Survival (Days)** | 490 | 377 | 510 | 3667 | 3 | 3664 | 596 | 167 | 429 | 1240 | -477 |
| **Sample 4: Survival Times of 14 People (Whose Tumors Are IDH1 Mutant)** | | | | | | | | | | | |
| **Survival (Days)** | 985 | 647 | 1065 | 3881 | 112 | 3769 | 988 | 501 | 487 | 1719 | -230 |
| **Sample 5: Ages at Diagnosis of 11 Women and Survival Times of 5 Women (Whose Tumors Are IDH1 Mutant)** | | | | | | | | | | | |
| **Survival (Days)** | 1846 | 1179 | 1434 | 3881 | 498 | 3383 | 2791 | 880 | 1911 | 5658 | -1987 |
| **Sample 6: Survival Times of 9 Males (Whose Tumors Are IDH1 Mutant)** | | | | | | | | | | | |
| **Survival (Days)** | 507 | 541 | 306 | 1024 | 112 | 912 | 691 | 203 | 488 | 1423 | -529 |

Table 1: Summary Statistics of All Datasets and Samples

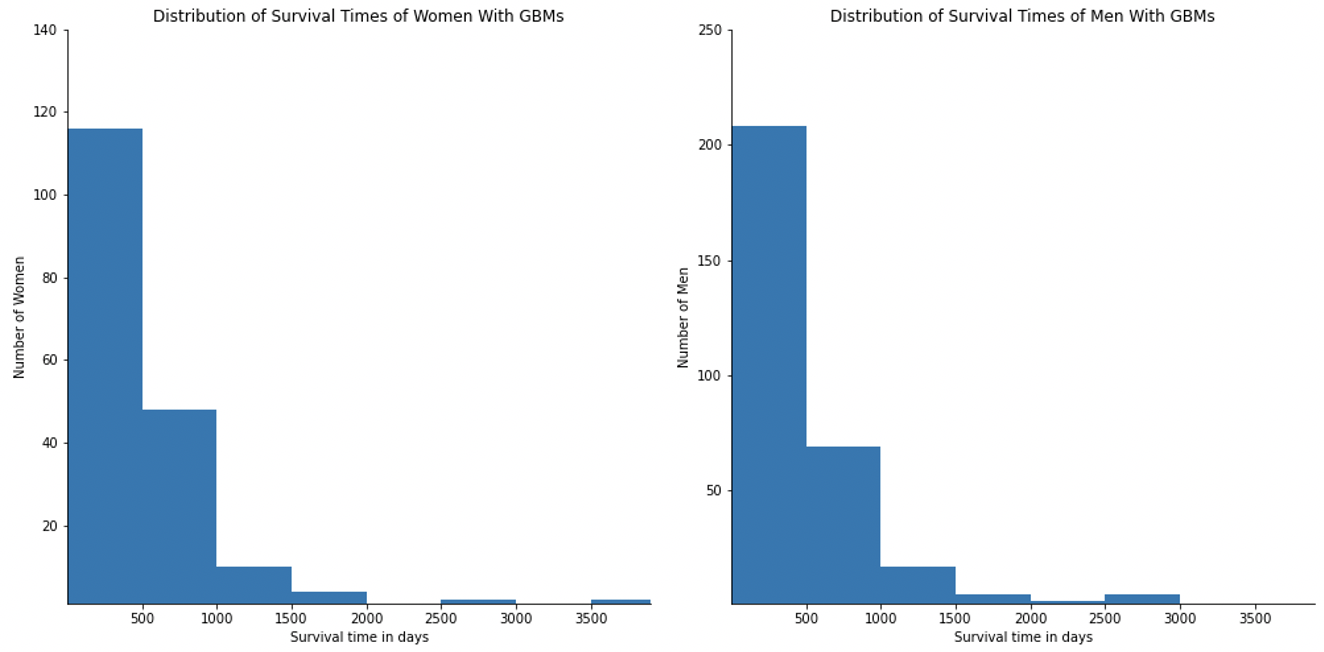


Figure 1: Histograms Showing The Distributions Of Samples 1 (left) and 2 (right)

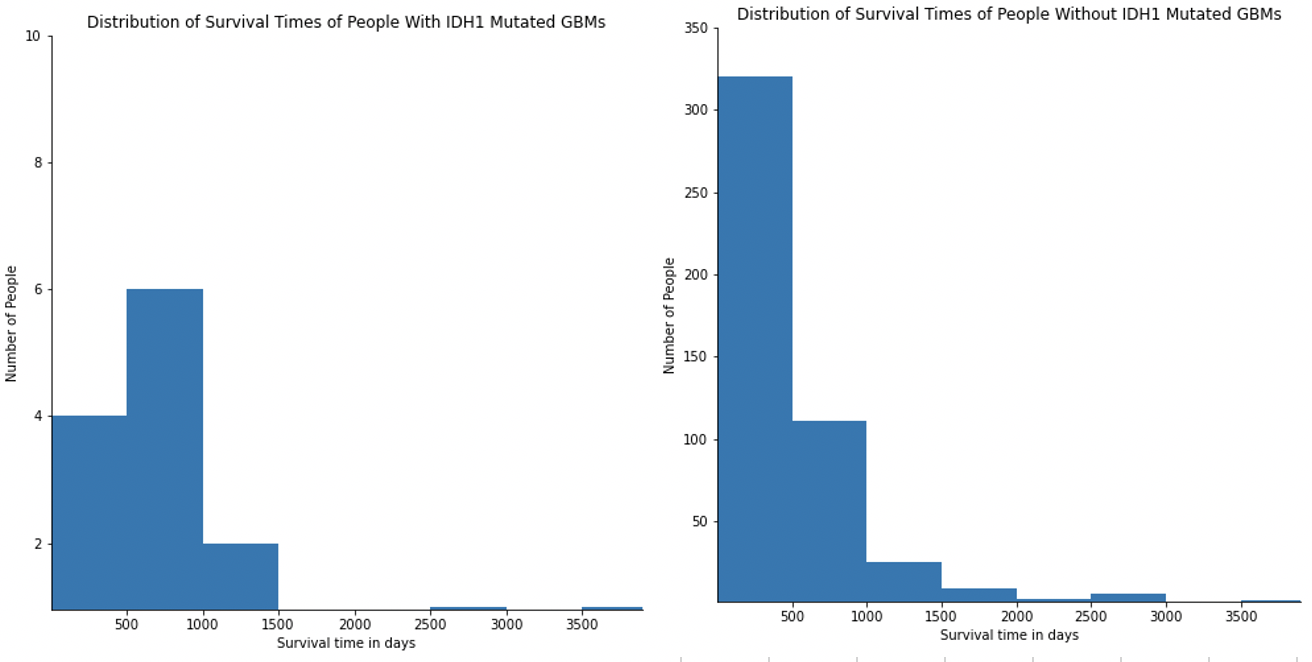


Figure 2: Histograms Showing The Distributions Of Samples 3 (right) and 4 (left)

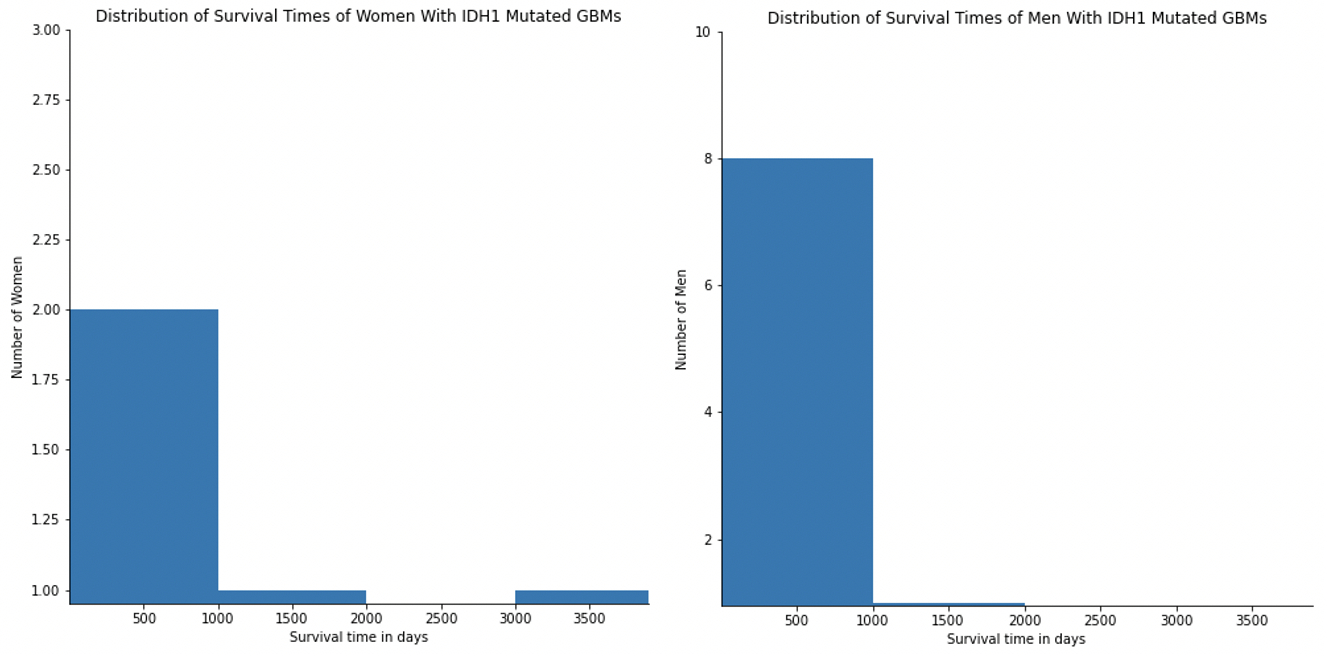


Figure 3: Histograms Showing The Distributions Of Samples 5 (left) and 6 (right)

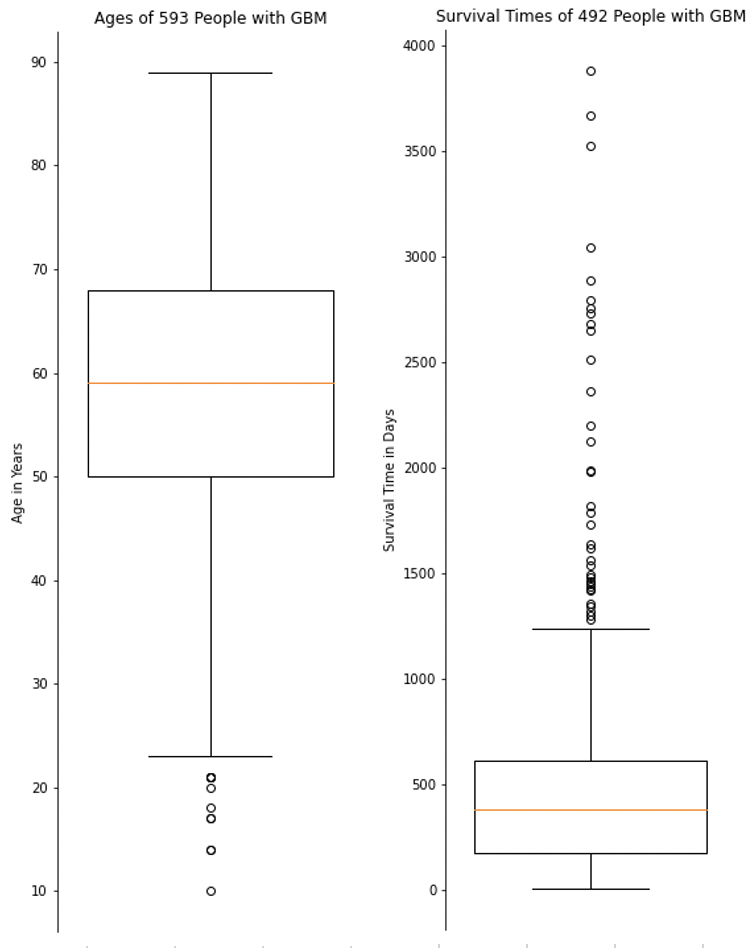


Figure 4: Boxplots of The Ages and Survival Times of People With GBM (Dataset 1)

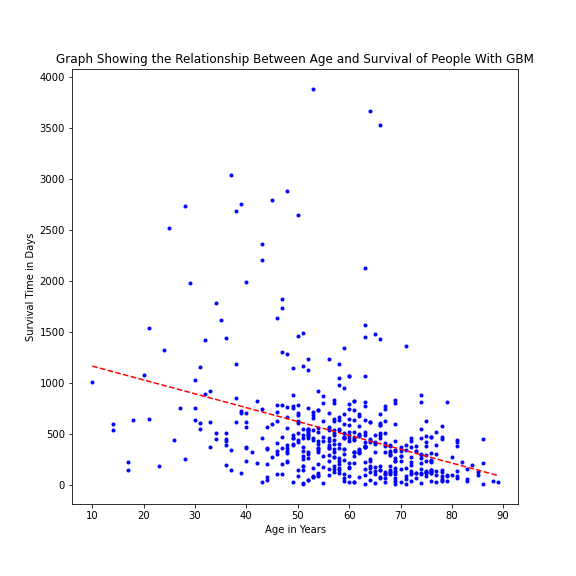


Figure 5: Linear Regression Curve (Dataset 1)

Methodology:

Investigations 1, 2, and 3

Figure 4 demonstrates that the Dataset 1 contains many outliers (43 in total). Table 1 shows that all samples except for Samples 5 and 6 contain at least one outlier as their maximum values lie outside of the upper Tukey fences. No outliers will be omitted as part of this investigation because they represent a substantial portion of the datasets and are likely valid datapoints. Because the T-Test uses the means of two samples for comparison, its statistical power can be compromised due to the outliers. Additionally, the T-Test requires that the two samples it is comparing are either large or normally distributed, neither of which is the case for Samples 4, 5 and 6; Figures 2 and 3 demonstrate the skewness of all six samples. The Mann-Whitney test assesses whether there is a statistically significant difference between two samples by determining if it is equally likely that a random element from the first sample is greater than a random element from the second sample and vice versa (Hart, 2001). Additionally, the Mann-Whitney test’s statistical power is not strongly affected by outliers (Zimmerman, 1994). There are several requirements for the usage of Mann-Whitney: The shapes of the distributions of the dependent variables of both datasets must be similar, the samples must be independent, and the dependent variables must be continuous (Karadimitriou and Marshall, n.d.). These requirements are largely fulfilled in the investigations made in this study except for in investigation 3. Figures 2 and 3 show that the histograms of Samples 5 and 6 do not match each others’ shapes nor any of the other samples’ shapes.

The Common Language Effect Size (CLES) is used to assess the extent to which two samples differ. It is the likelihood that a random element from the first sample is greater than a random element from the second (Oxford Reference, 2021). Thus, it can be calculated by comparing each element from the first sample with each element from the second sample and measuring which proportion of elements from the first are greater than those of the second. LeCroy and Krysik argue that knowing the effect size is important as, even if the differences between two samples are not statistically significant (perhaps due to small sample sizes), a high effect size points to a potentially significant relationship and thus can motivate further studies (2007). The CLES does not show how much greater elements in one sample are than those in another; the difference between the two medians of two samples can do that in a way that is less impacted by outliers than the difference between means. Hence, the median difference is also calculated.

Investigation 4

Linear regression is used to give an indication of the strength of the relationship between age at diagnosis and survival time. However, since no outliers are omitted, the linear regression’s statistical power is compromised. Thus, to evaluate whether the regression’s assessment of the strength of the relationship is realistic, Spearman’s correlation coefficient is calculated as it is robust to outliers (Croux and Dehon, 2010).

Results:

Hypotheses:

* H0: There are no statistically significant differences between the two samples.
* H1: There are statistically significant differences between the two samples.

Significance level α = 0.05.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mann-Whitney** | **Investigation 1** | **Investigation 2** | **Investigation 3.0** | **Investigation 3.1** | **Investigation 3.2** |
| P Value | 0.378 | 0.002 | 0.063 | 0.069 | 0.002 |
| CLES | 0.508 | 0.702 | 0.806 | 0.599 | 0.862 |
| Median Difference | -8 | 270 | 638 | 158.5 | 811 |
| Result | Do not reject H0 | Reject H0 | Do not reject H0 | Do not reject H0 | Reject H0 |

Table 2: Results of Investigations 1, 2, and 3

In investigation 1, we compared Samples 1 and 2. In investigation 3.0, we compared Samples 5 and 6. In both investigations 1 and 3.0, we tested whether the female sample had longer survival times than the male sample. In investigation 2, we compared Samples 3 and 4 and in investigations 3.1 and 3.2, we compared Samples 6 and 5 with Samples 1 and 2, respectively (but we removed the IDH1 mutant cases from samples 1 and 2). In investigations 2, 3.1 and 3.2, we tested whether the sample with the IDH1 mutation has longer survival times than the sample without.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Investigation 4** | | | | | |
| **Gradient** | **Intercept** |  | **P-value (Linear Regression)** | **Spearman‘s rs** | **P-value (Spearman’s rs)** |
| -13.59 | 1297.51d | 0.13 |  | -0.42 | 0.00 |

Table 3: Regression Statistics and Spearman’s rs

The gradient means that a one-year increase in the age at diagnosis yields a decrease of ca. two weeks of survival time. The intercept means that, if a child less than one year old was to be diagnosed with GBM, it would survive for roughly 3.6 years. R2 means that 13% of the variation in the survival times can be explained by the age at diagnosis. The p-value represents the probability that the relationship between the age at diagnosis and survival time could come about randomly. The p-value is in this case much smaller than 5%, which means that there is a statistically significant relationship.

Discussion:

There are several limitations to be acknowledged within this study. First, the sample sizes are very small. Second, we did not analyze the clinical history of the cases. Third, the statistical power of Mann-Whitney is compromised in investigation 3 because one of the requirements for Mann-Whitney is not fulfilled.

We reject the claim that the female sex alone has an impact on survival time. The differences are not statistically significant. This study confirms that the IDH1 mutation has an effect on survival time; the CLES and median difference imply that a patient whose GBM is IDH1 mutant has a 70% chance that she/he will outlive a patient with a non-IDH1 mutant GBM, likely by around 270 days. However, these results must be interpreted with caution because the size of Sample 4 is quite small. Thus, the median difference may not be a strong indicator of how strong the impact of an IDH1 mutation is.

Investigation 3.0 yields that the two samples are not statistically significantly different, though by only a small margin. The CLES and the median difference, however, show a very high effect size. Taking this into consideration, it seems necessary to conduct further studies with larger sample sizes and an investigation into the possible underlying biological factors may be beneficial.

Investigations 3.1 and 3.2 are contradictory to the finding of Schiffgens et al. that the effect of IDH1 mutated tumors on survival is only statistically significant for males (2016). We find it is only significant for females. Both Schiffgens et al.’s and this study only have small data samples but our finding confirms that more research must be done into the possibly differing effects of IDH1 mutated tumors on males and females and it provides additional evidence that IDH1 mutated tumors may have a stronger effect on survival for females than males.

The results of the linear regression in investigation 4 imply that the relationship between age at diagnosis and survival time is weak and the low R2 suggests that a linear model may not be the best model for Dataset 1. Spearman’s rs confirms that the relationship is weak but statistically significant. This means that there is likely no other monotonic relationship that is significantly better than linear regression.

Conclusions:

All in all, it may be said that female sex alone likely does not have an influence on survival times of people with GBM, which contradicts previous studies. These conflicting findings warrant further investigations into the influence of female sex on GBM survival. The influence of sex on survival time in conjunction with IDH1 mutated tumors must be further explored. Additionally, the evidence for the positive influence of IDH1 on survival times has been consolidated by this study and it was found that age likely has a minor influence on survival. The latter two findings may improve the accuracy of prognostics for GBM patients.

Code Excerpt:

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

**Bibliography**

Chandler, K., Prados, M., Malec, M. and Wilson, C., 1993. Long-Term Survival in Patients with Glioblastoma Multiforme. *Neurosurgery*, [online] 32(5), pp.716-720. Available at: <https://doi.org/10.1227/00006123-199305000-00003> [Accessed 27 April 2021].

Croux, C. and Dehon, C., 2010. Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, [online] 19(4), pp.497-515. Available at: <https://doi.org/10.1007/s10260-010-0142-z> [Accessed 27 April 2021].

Hart, A., 2001. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ*, [online] 323(7309), pp.391-393. Available at: <https://doi.org/10.1136/bmj.323.7309.391> [Accessed 27 April 2021].

Kanderi, T. and Gupta, V., 2021. *Glioblastoma Multiforme*. [online] Statpearls.com. Available at: <https://www.statpearls.com/ArticleLibrary/viewarticle/22272> [Accessed 27 April 2021].

Karadimitriou, S. and Marshall, E., n.d. *Mann-Whitney U test*. [online] sheffield.ac.uk. Available at: <https://www.sheffield.ac.uk/polopoly\_fs/1.885117!/file/66\_MannWhitneyU.pdf> [Accessed 27 April 2021].

LeCroy, C. and Krysik, J., 2007. Understanding and Interpreting Effect Size Measures. *Social Work Research*, [online] 31(4), pp.243-248. Available at: <https://www.jstor.org/stable/42659906>.

McLendon, R. and Halperin, E., 2003. Is the long-term survival of patients with intracranial glioblastoma multiforme overstated?. *Cancer*, [online] 98(8), pp.1745-1748. Available at: <https://doi.org/10.1002/cncr.11666> [Accessed 27 April 2021].

National Cancer Institute, 2021. *GDC Data Portal*. [online] Portal.gdc.cancer.gov. Available at: <https://portal.gdc.cancer.gov/exploration.> [Accessed 27 April 2021].

Oxford Reference, 2021. *Common Language Effect Size Statistic*. [online] Available at: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110810104638649> [Accessed 27 April 2021].

Schiffgens, S., Wilkens, L., Brandes, A., Meier, T., Franceschi, E., Ermani, M., Hartmann, C., Sandalcioglu, I. and Dumitru, C., 2016. Sex-specific clinicopathological significance of novel (Frizzled-7) and established (MGMT, IDH1) biomarkers in glioblastoma. *Oncotarget*, [online] 7(34), pp.55169-55180. Available at: <https://doi.org/10.18632/oncotarget.10465>.

Tamimi, A. and Juweid, M., 2017. *Glioblastoma*. [online] PubMed. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK470003/> [Accessed 27 April 2021].

Trifiletti, D., Alonso, C., Grover, S., Fadul, C., Sheehan, J. and Showalter, T., 2017. Prognostic Implications of Extent of Resection in Glioblastoma: Analysis from a Large Database. *World Neurosurgery*, [online] 103, pp.330-340. Available at: <https://doi.org/10.1016/j.wneu.2017.04.035> [Accessed 27 April 2021].

Walid, M., 2008. Prognostic Factors for Long-Term Survival after Glioblastoma. *The Permanente Journal*, [online] 12(4). Available at: <https://doi.org/10.7812/TPP/08-027> [Accessed 27 April 2021].

Yang, W., Warrington, N., Taylor, S., Whitmire, P., Carrasco, E., Singleton, K., Wu, N., Lathia, J., Berens, M., Kim, A., Barnholtz-Sloan, J., Swanson, K., Luo, J. and Rubin, J., 2019. Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. *Science Translational Medicine*, [online] 11(473). Available at: <https://doi.org/10.1126/scitranslmed.aao5253> [Accessed 27 April 2021].

Zimmerman, D., 1994. A Note on the Influence of Outliers on Parametric and Nonparametric Tests. *The Journal of General Psychology*, [online] 121(4), pp.391-401. Available at: <https://doi.org/10.1080/00221309.1994.9921213>.