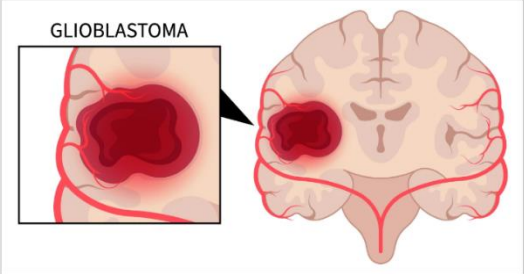
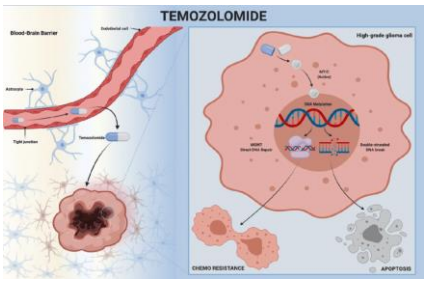
BCS410. Practical Class 1 & Homework 1 [Due: March 14]

Caution: You must submit the code you wrote and used to solve each problem. When submitting, name the folder with your student ID and submit each problem's code file as 'ProblemX_sol' (e.g., 'Problem1_sol'). You are free to use Python libraries. However, you are not allowed to ask AI to generate the entire code for you. If detected, you will receive a score of zero. You must provide your answers in a descriptive format for each problem. For example, if you are investigating whether there is a difference in the expression levels of a specific gene between two groups, you should clearly state the statistical hypothesis you formulated and specify the test you conducted. Additionally, you must accurately describe which values from the test results were used to draw your conclusion.



Background: Glioblastoma (GBM) is an aggressive and malignant brain tumor that arises from astrocytes, a type of glial cell in the central nervous system. It is classified as Grade IV glioma by the World Health Organization due to its rapid growth, high invasiveness, and poor prognosis. One of the major challenges in GBM treatment is the blood-brain barrier (BBB), which restricts the entry of most chemotherapeutic agents, limiting the effectiveness of chemotherapy. Only a few drugs can successfully cross the BBB. A key chemotherapeutic successfully crossing the BBB is Temozolomide (TMZ), an oral alkylating drug. It is the standard-of-care chemotherapy for GBM and is often administered alongside radiotherapy as part of the Stupp protocol. TMZ achieves nearly equal concentrations in plasma and cerebrospinal fluid, with a brain penetration ratio of approximately 20–40%.



Previous preclinical studies showed that a predictive biomarker for TMZ efficacy in GBM is O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation status. Patients with methylated MGMT respond better to TMZ, while those with unmethylated MGMT often exhibit poor response and resistance (see supplementary information below for details). Recently, strategies to inhibit MGMT (e.g., MGMT inhibitors like O6-benzylguanine) have been explored to enhance TMZ efficacy.

Preclinical study design: In this study, we aim to develop a novel drug that effectively inhibits MGMT, thereby enhancing the efficacy of TMZ in GBM treatment. Previous drug candidate screenings identified three potential MGMT inhibitors: *Mi1*, *Mi2*, and *Mi3*. To evaluate their effectiveness, we conducted a preclinical study using a GBM xenograft mouse model, ensuring that the initial tumor sizes were consistent across all subjects. Specifically, GBM xenograft mice were treated with either *placebo*, *Mi1*, *Mi2*, or *Mi3*, followed by TMZ administration. Fourteen days after treatment, the mice were sacrificed, and MGMT expression levels and tumor sizes were measured. Please see data "GBM_MGMT_expression_levels_data.csv" and "GBM_tumor_sizes_data.csv".

Problem 1: Please visually compare the MGMT expression levels of the four groups *placebo+TMZ, Mi1+TMZ, Mi2+TMZ, Mi3+TMZ* using a box-whisker plot. Then, perform a one-way analysis of variance (ANOVA) to test whether at least one group had a statistically significantly different MGMT expression level. Finally, conduct a multiple comparison test using Tukey's honestly significant difference (HSD) test to identify which group showed the lowest MGMT expression level.

Problem 2: Perform a two-way ANOVA with replication (factors: drug and age) to test whether tumor size is

statistically different depending on treatment approach and age. Finally, conduct a multiple comparison test using Tukey's HSD test to identify which group exhibited the smallest tumor growth. Assume that the data are normally distributed and have equal variance.

Clinical study design: In the preclinical study above, we found that *Mi1* is the most promising drug candidate. Next, we aimed to investigate its efficacy in humans. To achieve this, we conducted a prospective cohort study. Specifically, we collected data from 1000 patients with GBM (see 'GBM_patient_data.csv'). The dataset includes: (1) basic demographic information; (2) treatments that patients underwent (i.e., *TMZ*-untreated vs *TMZ*-treated alone vs. *Mi1+TMZ*); (3) clinical information, such as Isocitrate Dehydrogenase (IDH) mutation status; and (4) treatment outcomes, specifically overall survival (OS), which refers to the length of time from the initiation of treatment until the patient's death. Our first goal is to confirm whether TMZ administration improves survival outcomes (Problems 3,4, and 5). Next, we will assess the efficacy of *Mi1+TMZ* administration (Problem 6).

Problem 3: To confirm the efficacy of TMZ administration in GBM patients, we will compare the survival probability between the following two groups: (1) patients who were not treated with TMZ and (2) patients who were treated with TMZ alone. Before performing a survival analysis, we need to assess whether there is an imbalance in potential confounders between the two groups. To do this, examine whether the patient characteristics (i.e., age, sex, MGMT promoter methylation status, and IDH mutation) did not differ statistically significantly between the two groups. For categorical variables, use the Chi-square test. For continuous variables, use either a t-test or a Mann-Whitney test, depending on the data distribution. Note that when performing a t-test, it is also necessary to check for variance equivalence.

Problem 4: Please examine whether TMZ was effective in lengthening the OS of patients with GBM by performing a Kaplan-Meier analysis on the two groups in Problem 3. In addition to plotting the Kaplan-Meier curves, please also report the p-value to demonstrate the statistical significance of the difference in OS between the two groups.

Problem 5: Please examine whether the efficacy of TMZ on OS remains consistent even after adjusting for confounding factors using a Cox proportional hazards model. Specifically, plot the survival curves of the Cox proportional hazards model for cases where TMZ is either untreated or treated alone. Additionally, investigate the effects of other covariates on overall survival outcomes, including their hazard ratios and p-values (i.e., age, sex, MGMT promoter methylation status, and IDH mutation).

Problem 6: Please assess whether *Mi1* was effective in lengthening the OS of GBM patients with unmethylated MGMT when treated with TMZ by performing a Cox proportional hazards regression, as you did in Problems 3, 4, and 5. Hint: First, clearly define the two patient subgroups being compared.

Supplementary information

1. Role of MGMT in DNA Repair: MGMT is a DNA repair enzyme that removes alkyl groups from the O6 position of guanine, which is a critical site of alkylation by alkylating agents like TMZ. This repair mechanism helps maintain genomic stability and prevents mutations.

2. Mechanism of TMZ Cytotoxicity: TMZ is an alkylating chemotherapy drug that works by adding methyl groups to DNA, primarily at: O6-guanine (O6-MeG), N7-guanine (N7-MeG), and N3-adenine (N3-MeA). Among

these, O6-MeG lesions are the most cytotoxic because they lead to mismatch repair (MMR)-mediated apoptosis. If MGMT is active, it removes the O6-methyl group, thus preventing apoptosis and conferring resistance to TMZ.

3. MGMT Promoter Methylation and TMZ Sensitivity: If the MGMT gene promoter is methylated, MGMT expression is silenced, and cells cannot repair O6-MeG lesions, making tumors more sensitive to TMZ. If the MGMT promoter is unmethylated, MGMT is actively expressed, and it repairs TMZ-induced damage, leading to chemoresistance.

4. Impact of IDH Mutations on Tumor Biology: IDH1/2 mutations lead to a gain-of-function enzyme activity that produces D-2-hydroxyglutarate (D-2HG), an oncometabolite. D-2HG affects DNA and histone methylation, leading to widespread epigenetic dysregulation, including the CpG island methylator phenotype (G-CIMP). This alters gene expression and cellular differentiation, contributing to tumor initiation and growth. Therefore, IDH-mutant GBM patients tend to have a better prognosis than IDH-wildtype GBM patients, i.e., IDH mutations are associated with longer overall survival.