*Ibrutinib and bleeding REB*

Mina Dehghani, Cyrus Hsia, Shona Philip, Martha Louzada, Taylor Dear, Lenicio Siqueira, Anthony Quint

Methods: Statistical Analysis

Table of Contents

[Primary and Secondary Objectives (Outlined in the Protocol) 3](#_Toc74252405)

[Additional Statistical Tests (Not outlined in the protocol) 4](#_Toc74252406)

[Data visualization plots 6](#_Toc74252407)

[How to use R 7](#_Toc74252408)

# **Primary and Secondary Objectives (Outlined in the Protocol)**

*Primary objective*: the primary objective of this study is to investigate the risk of major bleeding associated with thrombocytopenia in patients on ibrutinib for treatment of CLL. We will be comparing rate of major bleeding in patients who have platelet counts 50 or less with patients who have platelet count more than 50.

Notes:

* Thrombocytopenia: low platelet counts

Statistical Analysis:

* Dataset consists of individuals who meet the inclusion and exclusion criteria, i.e., are taking ibrutinib for treatment of CLL
* Establish two categorical variables:

1. Bleeding severity - Major bleeding (defined as bleeding events with ISTH score 3 or 4) and non-Major bleeding (defined as bleeding events with ISTH score 1 or 2)

* Question: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?). If not, could set up 3 categories: major bleeding, minor bleeding, no bleeding?
* Answer:

1. Platelet count (Thrombocytopenia risk) – High platelet count (> 50) or low platelet count (< 50).

* Create a contingency table containing the frequency of each pair of categorical variables. Apply a fisher exact or chi-square test to establish significance between the variables. Can also calculate effect size if there is a statistical correlation present

*Secondary objectives:*

1. To investigate the risk of non-major bleeding associated with significant thrombocytopenia in patients on ibrutinib for treatment of CLL.

Statistical Analysis:

* This is taken into consideration in the primary objective

2.To investigate the risk of bleeding associated with anticoagulation in patients on ibrutinib.

Notes:

* Anti-coagulation (Y/N) is given in the excel data
* Question: same as before: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?)
* Answer:

Statistical tests:

* Contingency table with anticoagulation (yes or no) vs. Risk of bleeding (major vs. minor)
* Chi square or fisher exact to establish significance

3.To investigate the risk of bleeding associated with antiplatelets in patients on ibrutinib.

Notes:

* Antiplatelets are medicines that stop cells in the blood (platelets) from sticking together and forming a clot.
* Some patients are also taking extra antiplatelets on top on ibrutinib (such as aspirin)
* Continency table: antiplatelets (yes or no) vs. risk of bleeding (major vs. minor)
* Chi square or fisher exact to establish significance
* Question: same as before: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?)
* Answer:

4.To investigate the risk of bleeding associated with invasive procedures in patients on ibrutinib.

Notes:

* Question: Which is the column that indicates whether they had an invasive procedure?
* Answer:
* Question: same as before: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?)
* Answer:
* Contingency table: invasive procedure (yes or no) vs. risk of bleeding (major vs. minor)
* Chi square or fisher exact to establish significance

# **Additional Statistical Tests (Not outlined in the protocol)**

Some General Notes:

* Only using CTAE, it is 1 to 5
* Major bleeding classified as 3 and above, as well as any major bleeding in body, like in the brain, or anything that results in death
* Perform univariate and multivariate analysis

1.INR vs. risk of bleeding

Notes:

* INR: test that can measure some of the clotting ability of the blood. Can only get affected if on certain blood thinners. Anything above 1.5 would put you at a risk of bleeding
* Question: Do we care to make a column with INR data?
* Answer:

Statistical tests:

* Contingency table: INR (yes or no) vs. bleeding (major vs. minor)
* Chi square or fisher exact to establish significance
* Question: same as before: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?)
* Answer:

2.Survival Analysis

Notes:

* Idea: I can create a survival curve using duration of ibrutinib (in days) data. The survival curve can analyze how long an individual survives while on ibrutinib, i.e., the event of interest is death, while all other “duration of ibrutinib” values would be censored. This is because the current “duration on ibrutinib” column refers to time of death after starting ibrutinib, or last follow-up (which-ever comes first). Alternatively, I could also create a survival curve where the event of interest is a major bleeding event, i.e., when does a major bleeding event occur while on ibrutinib. However, to follow up with this second idea, I would need an extra “duration of ibrutinib” column with days corresponding to either the occurrence of a major bleeding event, or last follow-up (whichever comes first).
* Insert comments here:

Statistical tests:

* Kaplar Meier curves. Can create a curve with no stratification factors. Can also create Kaplar Meier curves with the stratification factors indicated in the primary and secondary objectives (for ex. >50 or <50 platelets).
* Can also establish a multivariate proportional cox test with all these stratification factors to view their influence on survival in this multifactorial approach. Calculates their hazard ratios too.

3.Multivariate analysis: Bleeding (major vs. minor) outcome based vs. all independent variables

Notes:

* The previous categorical tests were univariate and applied contingency tables. To analyze the effect of the independent variables on the outcome/dependent variable (bleeding) wholistically, must perform multivariate analysis.
* Independent variables: age of diagnosis, gender, platelet count, anti-coagulation, antiplatelet, invasive procedure, (maybe INR), anemia, Molecular/Cytogenetics, (maybe Prior lines of therapy), (maybe duration of Ibrutinib), (maybe PMHx bleeding risk),
* Answer:
* Question: for the independent variables that I haven’t stated that I’ll do a contingency table for (i.e., anemia etc.), should I complete a contingency table for those as well?
* Answer: I’m thinking yes. Might as well do a univariate analysis for each independent variable in relation to bleeding (major vs. minor).
* Confirm a score of zero can be part of the “minor” risk.
* Answer:

Statistical tests:

* Perform a logistic regression analysis, then a CART analysis as a sensitivity analysis.
* Question: same as before: should non- major bleeding be defined to also include patients who didn’t receive a score (i.e., didn’t have bleeding/had a score of ‘0’?). I.e., should my model have a binary outcome (major vs. minor) or have three possible outcomes (major or minor or no bleeding)
* Answer:
* Can use the logistic regression model as a predictor of outcomes for future patients, i.e., assessing their future risk of bleeding based on their present categorical and quantitative classifications.

4.Mean platelet count for those who had major or minor bleeding versus those who had no bleeding.

Notes:

* Question: Instead, could compare (major bleeding) vs. (minor or no bleeding). Instead, could also compare (major) vs. (no bleeding) and (minor) vs. (no bleeding) and (major vs. minor). Which would be best?
* Answer:
* Question: at which time point would I sample the platelet count? (Hb too, see next question)
* Answer:
* Question: in addition to mean platelet, can also repeat analysis for mean Hb, mean prior lines of therapy, mean duration of ibrutinib, mean age?
* Answer:

Statistical tests:

* Perform students t test to establish statistical significance between the means of the groups

# **Data visualization plots**

* Pie chart showing proportion of how many people experienced bleeding events (Y/N). Can also show proportions of major vs. minor vs. no bleeding events.
* Pie chart showing the relative proportions of the different types of bleedings events, if they did occur (airway, GI etc.)
* Can show a pie chart of the relative proportions of bleeding score grades which people got (0, 1, 2, 3, 4, 5)
* Can plot trend of how platelet levels change along time during treatment of ibrutinib. Can repeat for Hb.
* Pie chart showing proportion of males and females. Or a population pyramid to also show the amount of people in the respective age cohorts as well.
* Pie chart showing proportion of the cytogenic disorders?
* Pie chart showing proportion of how many prior lines of therapy people got.
* Histogram: Frequency vs. Duration of ibrutinib (with an adequate bin size chosen). Can make this a PDF and see if it follows a probability distribution.
* Pie chart showing proportion of people with PMH?
* Question: Which of these (if any at all) are valuable?

# **How to use R**

Logistics Regression:

<https://datascienceplus.com/perform-logistic-regression-in-r/>

<https://datascienceplus.com/modelling-dependence-with-copulas/> - about copulas.

<http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/>

Logistic regression is limited to only two-class classification problems. There is an extension, called *multinomial logistic regression*, for multiclass classification problem (Chapter @ref(multinomial-logistic-regression)). Note that, the most popular method, for multiclass tasks, is the *Linear Discriminant Analysis* (Chapter @ref(discriminant-analysis)).

CART:

<http://www.sthda.com/english/articles/35-statistical-machine-learning-essentials/141-cart-model-decision-tree-essentials/>