

# BIOSTAT 702: Exercise 1.2

Table 1: Describing The Study Participants

August 14, 2025

## Contents

How to Do This Exercise	1
Grading Rubric	1
Resources	2
Question 1 (4 points)	2
Question 2 (4 points)	2
Question 3 (4 points)	3
Question 4 (4 points)	4
Question 5 (4 points)	6

## How to Do This Exercise

We recommend that you read this entire document prior to answering any of the questions. If anything is unclear please ask for help from the instructors or TAs before getting started. You are also allowed to ask for help from the instructors or TAs while you are working on the assignment. You may collaborate with your classmates on this assignment—in fact, we encourage this—and use any technology resources available to you, including Internet searches, generative AI tools, etc. However, if you collaborate with others on this assignment please be aware that *you must submit answers to the questions written in your own words. This means that you should not quote phrases from other sources, including AI tools, even with proper attribution.* Although quoting with proper attribution is good scholarly practice, it will be considered failure to follow the instructions for this assignment and you will be asked to revise and resubmit your answer. In this eventuality, points may be deducted in accordance with the grading rubric for this assignment as described below. Finally, you do not need to cite sources that you used to answer the questions for this assignment.

## Grading Rubric

The assignment is worth 20 points (4 points per question). The points for each question are awarded as follows: 3 points for answering all parts of the question and following directions, and 1 point for a correct answer. Partial credit may be awarded at the instructor's discretion.

## Resources

The following resources on Canvas / the internet will be helpful for answering the questions for this exercise.

1. The article by Douglas Altman, Comparability of Randomised Groups (The Statistician (1985) 34, pp. 125-136). The PDF is available on Canvas and the article can be viewed online [here](#).
2. The CONSORT 2010 statement available [here](#).
3. The CONSORT 'explanation and elaboration' paper. The PDF is available on Canvas and the article can be viewed online [here](#).
4. The manuscript by Jabbour titled *Ponatinib vs Imatinib in Frontline Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia: A Randomized Clinical Trial*. The PDF is available on Canvas and the article can also be viewed online [here](#).
5. The dataset for this exercise, simulated\_clinical\_trial\_data.csv.

## Question 1 (4 points)

Complete the first 12 items of the CONSORT checklist for the trial reported by Jabbour (you can ignore items 13-25). How closely do the authors adhere to the CONSORT guidelines?

See the attached checklist with items 1-12 completed.

## Question 2 (4 points)

1. Write a brief description of the study by Jabbour in your own words, including the following:
  - (a) Research objectives
  - (b) Patient population recruited for the study
  - (c) Study design
  - (d) Primary outcome
  - (e) Result for the primary outcome

This is a Phase 3 trial comparing Ponatinib to Imatinib (standard care) in adults with Philadelphia chromosome positive acute lymphoblastic leukemia who don't have serious cardiovascular disease. The study was conducted at 77 centers worldwide and enrolled 245 patients, randomized 2:1 (164 on Ponatinib and 81 on Imatinib). The primary outcome was remission without minimal residual disease at the completion of the first 3 cycles of treatment. The intention to treat analysis included 213 patients in whom Philadelphia chromosome status could be verified (154 on Ponatinib and 78 on Imatinib). A total of 34.4% of patients on Ponatinib had remission whereas only 16.7% on Imatinib had remission (risk difference: 0.18, 95% CI: 0.06-0.29, p-value=0.002) suggesting that Ponatinib is superior to the standard care.

2. Answer the following questions about Table 1 in the article by Jabbour.
  - (a) Which variables are continuous? Age
  - (b) Which variables are unordered categorical variables? BCR:ABL1 dominant isoform is a multi-level categorical variable. The others are binary: Sex, CNS disease, Cardiovascular comorbidities (these are two binary variables), Hypertension, Diabetes, Obesity, Dyslipidemia, and History of Smoking.

- (c) Which variables are ordered categorical variables? **ECOG score**
- (d) Why does the table not have an “overall” column that shows descriptive statistics for all 245 patients? **The experimental design treats the groups separately so there is no utility to considering them in a combined fashion.**
- (e) Why are there no p-values in Table 1? **P-values are not useful to identify imbalance of baseline factors as described above.**

### Question 3 (4 points)

1. Review the CONSORT diagram in Figure 1 from the article by Jabbour and write a brief summary of what you see. Describe reasons why patients were excluded (see notes under Table 1 for additional details) and what, if anything, might cause you to be concerned about generalizing results from those who were analyzed to those who were enrolled and randomized. After doing this, consider what the target population for the study is and list any concerns you have about generalizing the study results to that population.

The target population are adults recently diagnosed with Philadelphia chromosome positive (Ph+) acute lymphoblastic leukemia. These patients are typically treated with Imatinib. The source population for the study included patients visiting 77 clinic sites over a 3 year period, in an attempt to capture a sample of the target population. But the eligible subset of the source population were at least able to walk and perform self-care, if not be fully active (signified by ECOG scores of 0, 1, or 2) and didn't have serious cardiovascular disease. In that way, the eligible population might perhaps have a better general health condition than the broader target population. The trial enrolled 245 patients but only analyzed 232, the reason being that the presence of the Philadelphia chromosome wasn't verifiable in 13 patients. The primary analysis included everyone as randomized, even though many discontinued treatment before the primary outcome was measured. Overall, the minimal difference between the analyzed set and the enrolled set bolsters confidence around internal validity. The main point for the investigator to consider is whether the results are relevant for the broader target population, which likely includes much sicker patients than those who participated in the trial. One potential consequence of this is that the treatment benefit might be smaller in practice than was observed in the trial.

2. Summarize Altman's argument about not performing statistical testing on Table 1 in a randomized trial. *Note: Your instructors are aware that understanding Altman's argument on a deep level assumes familiarity with material you haven't yet covered in BIOSTAT 701. Don't worry about that for now. Simply do your best to describe Altman's argument in a non-technical fashion using what you currently know about statistics.*

A p-value is the probability of the observed difference or more extreme if  $H_0$  is true. When patients are assigned to treatments through a random process,  $H_0$  is true by construction; i.e, the process ensures that we don't expect to see differences between the groups (emphasis on "expect!"). So, it isn't sensible to interpret the p-value when you know  $H_0$  is true. Confusion arises, however, when some p-values are below 0.05, which sets off alarms for some people. But remember, if our alpha level for a test is 5% then we expect 5% of our tests to come up "positive" even when  $H_0$  is true. But the fact that p-values don't indicate imbalance doesn't mean that imbalance can't occur; clearly, it can. Expectation is again the culprit: it doesn't occur in the long run, but in any given study it might be there to some degree. The question is whether the imbalance is on a factor related to the outcome, which creates confounding. The strength of imbalance, and the strength of association with the outcome, will determine whether a factor is influential in any given study.

3. Shift your attention back to the ultra-running study for a moment. Recall that this is an observational study instead of a randomized trial, and that our objective (which we are slowly building towards) is to run a simple linear regression predicting best running time based on emotional intelligence. In the last exercise you created a table that compared the 73 participants who would be excluded from the simple linear regression with the 211 participants who would be included. The table had a very similar

structure to the typical Table 1 from a randomized trial, i.e., that it compares baseline characteristics between two independent groups. The purpose of the table was different, however. In other words, the exercise with the ultra-running study was intended to identify selection bias, i.e., systematic differences between the analysis set and the entire sample. This is contrasted with the purpose of Table 1 in a randomized trial, which is to assess imbalance in baseline factors that might result in another kind of bias called confounding, which as Altman demonstrates in his paper, can happen even in a randomized trial. This phenomenon occurs specifically when the imbalance is on factor(s) that are also associated with the outcome.

You already answered a question above about why it doesn't make sense to use p-values to assess baseline differences between randomized groups in a clinical trial. Explain why it also doesn't make sense to use p-values in the ultra-running study context, where our objective was to assess selection bias.

In a non-randomized study we don't know whether  $H_0$  is true so the p-value is at least interpretable. However, even if  $H_A$  is true then testing is also unreasonable because p-values are influenced by sample size and variability. In other words, it is possible not to reject  $H_0$  due to low power, but to still have imbalance that is meaningful.

## Question 4 (4 points)

Suppose that you were the statistician for this trial and were asked to create Table 1 from the raw data. Generate a Table 1 using the variables described below, stratifying by Group, and including a p-value column as well as a standardized mean difference column. *Note: The table won't exactly match the format of Table 1 in the article, and the summary statistics from the included dataset won't exactly match the numbers from Table 1 in the article either. Also note that we are asking you to put p-values in this table so that you can answer some of the questions we are asking in this exercise, not because p-values are appropriate in a table like this.*

The following is a description of the variables in the data file.

Age -	age of the patient at randomization, in years
Gender -	0=Male, 1=Female
ECOG -	Eastern Cooperative Oncology Group Performance Status
CNS -	Central Nervous System Disease, 0=No, 1=Yes
BCR -	BCR:ABL1 Isoform 0=p210, 1=p190
CV1 -	>=1 Cardiovascular Comorbidity, 0=No, 1=Yes
CV2 -	>=2 Cardiovascular Comorbidity, 0=No, 1=Yes
Hypertension -	0=No, 1=Yes
Diabetes -	0=No, 1=Yes
Obesity -	0=No, 1=Yes
Dyslipidemia -	0=No, 1=Yes
Smoking -	History of Smoking, 0=No, 1=Yes
Treatment -	randomized treatment group

```
# Load the data
data <- read.csv(
  here::here("Datasets/SimulatedTrial/simulated_clinical_trial_data.csv")
)

# Convert categorical variables to factors
data$Gender <- factor(data$Gender, labels = c("Male", "Female"))
data$ECOG <- factor(data$ECOG, labels = c("0", "1", "2"))
data$CNS <- factor(data$CNS, labels = c("No", "Yes"))
data$BCR <- factor(data$BCR, labels = c("p210", "p190"))
```

```

data$CV1 <- factor(data$CV1, labels = c("No", "Yes"))
data$CV2 <- factor(data$CV2, labels = c("No", "Yes"))
data$Hypertension <- factor(data$Hypertension, labels = c("No", "Yes"))
data$Diabetes <- factor(data$Diabetes, labels = c("No", "Yes"))
data$Obesity <- factor(data$Obesity, labels = c("No", "Yes"))
data$Dyslipidemia <- factor(data$Dyslipidemia, labels = c("No", "Yes"))
data$Smoking <- factor(data$Smoking, labels = c("No", "Yes"))

# Reorder the levels of Treatment to put Ponatinib first
data$Treatment <- factor(data$Treatment, levels = c("Ponatinib", "Imatinib"))

# Create labels for the variables
library(labelled)
var_label(data$Age) <- "Age (years)"
var_label(data$Gender) <- "Gender"
var_label(data$ECOG) <- "ECOG Performance Status"
var_label(data$CNS) <- "CNS/Extramedullary disease"
var_label(data$BCR) <- "BCR::ABL1 Isoform"
var_label(data$CV1) <- ">=1 Cardiovascular Comorbidity"
var_label(data$CV2) <- ">=2 Cardiovascular Comorbidities"
var_label(data$Hypertension) <- "Hypertension"
var_label(data$Diabetes) <- "Diabetes"
var_label(data$Obesity) <- "Obesity"
var_label(data$Dyslipidemia) <- "Dyslipidemia"
var_label(data$Smoking) <- "History of Smoking"

library(tableone)

```

## Warning: package 'tableone' was built under R version 4.4.1

```

vars <- c("Age", "Gender", "ECOG", "CNS", "BCR", "CV1", "CV2", "Hypertension",
          "Diabetes", "Obesity", "Dyslipidemia", "Smoking")
comparisonTable <- CreateTableOne(vars = vars, strata = "Treatment",
                                  data = data)
kableone(print(comparisonTable, smd=TRUE, varLabels=TRUE,
               showAllLevels=TRUE, printToggle = FALSE))

```

	level	Ponatinib	Imatinib	p	test	SMD
n		164	81			
Age (years) (mean (SD))		53.75 (20.24)	55.05 (18.30)	0.626		0.067
Gender (%)	Male	97 (59.1)	38 (46.9)	0.094		0.247
	Female	67 (40.9)	43 (53.1)			
ECOG Performance Status (%)	0	42 (25.6)	31 (38.3)	0.115		0.280
	1	58 (35.4)	22 (27.2)			
	2	64 (39.0)	28 (34.6)			
CNS/Extramedullary disease (%)	No	83 (50.6)	39 (48.1)	0.821		0.049
	Yes	81 (49.4)	42 (51.9)			
BCR::ABL1 Isoform (%)	p210	81 (49.4)	44 (54.3)	0.555		0.099
	p190	83 (50.6)	37 (45.7)			
>=1 Cardiovascular Comorbidity (%)	No	90 (54.9)	37 (45.7)	0.223		0.185
	Yes	74 (45.1)	44 (54.3)			

	level	Ponatinib	Imatinib	p	test	SMD
>=2 Cardiovascular Comorbidities (%)	No	84 (51.2)	42 (51.9)	1.000		0.013
	Yes	80 (48.8)	39 (48.1)			
Hypertension (%)	No	86 (52.4)	42 (51.9)	1.000		0.012
	Yes	78 (47.6)	39 (48.1)			
Diabetes (%)	No	89 (54.3)	34 (42.0)	0.094		0.248
	Yes	75 (45.7)	47 (58.0)			
Obesity (%)	No	84 (51.2)	40 (49.4)	0.893		0.037
	Yes	80 (48.8)	41 (50.6)			
Dyslipidemia (%)	No	86 (52.4)	43 (53.1)	1.000		0.013
	Yes	78 (47.6)	38 (46.9)			
History of Smoking (%)	No	88 (53.7)	44 (54.3)	1.000		0.013
	Yes	76 (46.3)	37 (45.7)			

## Question 5 (4 points)

Answer the following questions about the Table 1 that you generated.

1. Would you reject the null hypothesis of no difference between the treatment arms for any of the baseline characteristics in Table 1 (using a 5% alpha level)? **No, there are no p-values less than .05.**
2. If you were to interpret the p-value as continuous, rather than using a threshold value like 5%, then are there any factors for which the p-value is suggesting the evidence leans in favor of the alternative hypothesis (i.e., that the distribution of a factor is actually different in patients assigned to Ponatinib vs. Imatinib)? Discuss why this doesn't make any sense in this context. Refer to Altman's paper as a guide as you think about this. **There are several factors with small p-values: gender, ECOG Performance Status, and Diabetes. We can't interpret these small values as evidence in favor of the alternative because we know H0 is true here. Under H0 all p-values are equally likely so there's actually no information contained in the magnitude of the p-value.**
3. Are there any factors that are imbalanced between the groups when you look at the standardized mean differences (use 0.2 as a threshold)? List the factors and describe the imbalance that you see. **Gender, ECOG score, and Diabetes. The Ponatanib group has a larger proportion of males, poor performance status, and diabetes.**
4. Why do you think the standardize mean differences are a more satisfactory approach to assessing imbalance than using p-values? **The standardized mean differences don't require us to condition on the truth of H0 for interpretation. The SMD is simply a ratio of signal (size of difference) to noise (amount of variability) in the data. This implies the SMD relates directly to the importance of the imbalance in this study specifically, which is what we care about. Making inference about differences in the population, which is what we do with p-values, isn't sensible here.**
5. Suppose that the factors you identified as being imbalanced were also strongly associated with the primary outcome of the study. Using Altman's argument, what might you consider doing if you were the statistician for the study? **Adjust for these factors since they would by definition create confounding. This is a subtle distinction between confounding in the data vs. confounding in the population.**
6. Suppose that the factors you identified as being imbalanced had no association with the primary outcome of the study. Again, referring to Altman's paper, is the imbalance a concern and what, if anything, would you consider doing about it? **No need to do anything. Imbalance on a factor that isn't related to the outcome cannot create confounding.**