ESA DOSING OF END-STAGE RENAL DISEASE PATIENTS

Howard Hu

August 13, 2022

1 Exclusive Summary

This report aims to find a statistical framework that can help doctors better understand kidney disease treatment performance. Since we need to compare the UVA's patients versus national patients, we would like to use the standardization framework. There are two main approaches. The first one is direct standardization which means we use the distribution of UVA's population (stratum-specific population) to standardize the national patients. The second one is indirect standardization which indicates we can apply the standard rate of national patients to the UVA's population. We want to use the median Hgb levels to separate patients into high strata and low strata.

2 Introduction

2.1 General Background

The client of this report is Dr. Benjamin Lobo, a research scientist from the school of Data Science at UVA. This report's main topic discusses kidney disease, more specifically End-Stage Renal Disease (ESRD). Over 30 million Americans suffer from some form of kidney disease. ESRD is the most severe form of renal disease, affects over 650,000 people, and is more prevalent in minorities than among whites. Doctors can diagnose people whose kidney's ability to filter blood toxins falls below 15 percent. Then there are two possible treatment options. They are dialysis and kidney transplant. Fortunately, UVA owns 12 dialysis units across Virginia with about one thousand patients. For this report, we focus on dialysis treatment. Although there is not an extremely large patient population, dialysis patients occupy seven percent of the medicating budget with one percent of the medicate population.

According to the Medicare guidelines, patients accept Erythropoiesis stimulating agents (ESAs) treatment to reduce the target red blood cell levels to 10 - 12 g/dl. For the current protocol, every patient has the same treatment based on the current Hgb level and the magnitude and direction of change in Hgb levels between the current and one-month prior levels. Then doctors monthly record patients' Hgb levels and their ESAs doses value. Since ESAs treatment is very expensive, we want to develop a statistical framework to help doctors understand the performance of ESAs treatment. However, we face some problems, such as unreliable DOPPS data in recent updates. The pandemic makes it hard to trace patients, there is some cohort variability in the treatments, and there are diminishing returns in real-world practice.

2.2 Objectives

The main objective of our client is that are there any statistical frameworks for comparing performance over time while accounting for the differences among the cohorts treated?

3 Approach

According to the given sides from Dr. Kafadar, which references Data Analysis and Regression[1], One of the possible approaches is using standardization to compare different rates across regions. We want

to compare UVA's kidney disease patients with the national kidney disease population for our report. We know we will have a different sample size between UVA's and national samples; we would like to use direct standardization to make a fair comparison. For example, suppose we try to compare two methods of teaching (Online synchronous (OS) and asynchronous (OA)) with two groups of students (easy to teach and hard to teach), and we measure the success by the passing grade on the test. In that case, we can have two following figures.

	(1) ASynch	(2) OSynch	Total
Easy	$800 = n_{E1}$	$100 = n_{E2}$	$N_{easy} = 900$
Hard	$200 = n_{H1}$	$900 = n_{H2}$	$N_{hard} = 1100$
Total	$N_1 = 1000$	$N_2 = 1000$	2000

Figure 1: 1000 students to each method

	(1) ASynch	(2) OSynch	Total
Easy	$560 = s_{E1}$	$80 = s_{E2}$	$S_{easy} = 640$
Hard	$40 = s_{H1}$	$360 = s_{H2}$	$S_{hard} = 400$
Total	$S_1 = 600$	$S_2 = 440$	1040

Figure 2: Number of students passed test in each group

From the previous two figures, we know the success rate of method 1 is better than method 2 ($600/1000 = 60\% > 440/1000 = 44\%$). However, if we calculate the success rate within each category, we will have the following figure.

	(1) ASynch	(2) OSynch	Total
Easy	$560/800 = 70\%$	$80/100 = 80\%$	$p_{easy,1} < p_{easy,2}$
Hard	$40/200 = 20\%$	$360/900 = 40\%$	$p_{hard,1} < p_{hard,2}$

Figure 3: The success rate within each category

We will know the previous conclusion is misleading because method 1 has a lower passing rate in each category. The reason for having this problem is that the sample size of each case is different. To solve this problem, we can use the idea of direct standardization by applying the distribution of asynchronous cases ($80\% easy and 20\% hard$) to both two groups. Similarly, we can apply the distribution of online synchronous cases ($10\% easy and 90\% hard$) on both two groups. Then we will have two new figures.

	(1) ASynch	(2) OSynch	Total
Easy	$560/800 = 70\%$	$640/800 = 80\%$	$p_{easy,1} < p_{easy,2}$
Hard	$40/200 = 20\%$	$80/200 = 40\%$	$p_{hard,1} < p_{hard,2}$
Total	$600/1000 = 60\%$	$720/1000 = 72\%$	$p_1 < p_2$ (12%)

Figure 4: Apply the distribution of asynchronous cases

	(1) ASynch	(2) OSynch	Total
Easy	$70/100 = 70\%$	$80/100 = 80\%$	$p_{easy,1} < p_{easy,2}$
Hard	$180/900 = 20\%$	$360/900 = 40\%$	$p_{hard,1} < p_{hard,2}$
Total	$250/1000 = 25\%$	$440/1000 = 44\%$	$p_1 < p_2$

Figure 5: Apply the distribution of online synchronous cases

Both figures show method two is better than method 1, which is consistent with the results of figure 3.
According to the idea of direct standardization, if we want to develop a framework for comparing performance of kidney disease treatment over time while taking into account the differences in the cohort being treated. We use the distribution of UVA's population (stratum-specific population) to standardize the national patients. Then we can avoid to make some misleading conclusion.

Moreover, there is another standardization idea called indirect standardization. The main idea is to take the standard population rates and apply them to stratum-specific populations. Suppose

we analyze the death rate in counties in California[1]. We separate the populations into their strata (Excellent, Fair, and Poor) by the number of exposed people. We will have the following figure.

	Stratum	I	II	PopRate
	Excellent	1	0.8	14/15,000
#Deaths	Fair	3	5.0	29/7,000
	Poor	1,000	100	3/21
	Total	20	26	

Expected Deaths = Rates × Stratum-specific Pop:

Stratum	I	II
Excellent	9.33	4.67
Fair	12.43	16.57
Poor	0.14	2.86
Total	21.90	24.10

Figure 6: The example of indirect standardization

Then we can use a t-test to compare the observed and expected populations.

$$\text{General: } \sqrt{4 \cdot obs + 2} - \sqrt{4 \cdot exp + 1} \sim N(0,1)$$
$$(\text{if } obs = 0, \text{ use } \sqrt{4 \cdot obs + 1})$$
$$\text{Method 1: } \sqrt{4 \cdot 20 + 2} - \sqrt{4 \cdot 21.90 + 1} = 9.06 - 9.41 = -0.35$$
$$\text{Method 2: } \sqrt{4 \cdot 26 + 2} - \sqrt{4 \cdot 24.10 + 1} = 10.03 - 9.87 = 0.43$$
$$\text{“t-test”: } (-0.35 - 0.43)/\sqrt{2} = -0.55$$

Figure 7: The T-test comparison in indirect standardization

The idea of indirect standardization applies the standard rates to stratum-specific populations. In our cases, we can apply the standard rate of national patients to the UVA population.

The critical question is how to separate samples into different stratum for direct standardization and indirect standardization. We would like to use the median Hgb levels as our separate point. Patients who have higher than median Hgb levels are marked as high. Otherwise, we can make it as low. Using the idea of a standardization framework can help us to avoid stating a misleading conclusion.

4 Conclusion

The goal of this report is to develop a statistical framework that can help doctors better understand kidney disease treatment outcomes. We need to compare UVA's patients with national patients so that we will use the standardization framework. There are two main approaches. The first is direct standardization, which means we use the distribution of UVA's population (strata-specific population) to standardize national patients. Secondly, indirect standardization involves applying the standard rate of national patients to UVA's patient population. By using median Hgb levels, we can classify patients into high versus low strata.

References

[1] Frederick Mosteller and John W. Tukey. *Data Analysis and regression: A second course in statistics*. Pearson, 2020.