**Revised Pre-course News Article**

**Predicting platelet usage using statistics and big data**

*Scientists at Stanford successfully predicted blood platelet usage in hospitals with remarkable accuracy using statistics, potentially saving $80 million in health care costs.*

Have you ever wondered how a small cut on your finger can heal in just a few days? It turns out that blood platelets present in our blood are responsible for healing injuries!

In the event of an injury, platelets will cluster together at the site of the wound. A clot will be formed, and this will prevent blood from leaking out. This clot also acts as a scaffolding where tissues and skin cells regrow, which is exactly the reason why wounds can heal very quickly.

Because of its ability, hospitals use platelets from donors to reduce blood loss in patients who are recovering from surgery and trauma, as well as prevent bleeding in patients with low platelet counts in their blood.

The problem is it is very hard to predict daily demand for platelets in hospitals. Because platelets donated from blood donation centers need to spend two days in testing before being certified as usable, hospitals have to predict their platelet usage three days in advance.

Of course, any rational hospital would just stockpile platelets as much as they can to prevent a shortage in case there is an unforeseen surge in demand, which is usually caused by natural disasters or accidents. However, unlike canned food, platelets only have a very short shelf life of three days before expiring. These expired platelets end up being wasted.

Approximately 10.5% of 2 million platelet units are wasted annually, and this represents $80 million in healthcare costs! This could definitely be better used to improve the quality and effectiveness of other areas in the healthcare industry. An accurate method for prediction of platelet usage is consequently necessary to minimize wastage of platelets, while at the same time maintaining a buffer stock of platelet units as a safeguard measure.

Numerous studies have aimed to predict platelet usage using patients’ data. However, these studies had very limited sample size and thus cannot be used as a general model for the larger population.

At Stanford University, researchers Leying Guan and colleagues from the Department of Statistics sought to solve the limitations in current forecasting methods. Their goal was to develop a mathematical model which would be able to forecast platelet usage three days in advance accurately.

Unlike previous studies, this time the Stanford researchers utilized large amounts of data collected over a very long period of time. They hypothesized that certain measurements in patients’ data, such as platelet count and blood composition, greatly influence the usage of platelet units. To test this claim, the researchers collected data from the university’s associated hospital, Stanford Health Care (SHC), as well as the number of platelets transfused and wasted each day for more than 800 days.

Imagine you have to read textbooks for more than 800 days non-stop. You have to take a quiz at the end of each day, and the suggested answers as well as your quiz score will be given. You will surely learn from your mistakes and perform well for your final exam at the end of the period.

In the same way, our model is being fed data from SHC and learns the trend continuously for 800 days. The model is also tested every day to see if the prediction matches the actual amounts of platelets used. Over time, the gap between prediction and actual value will narrow. The result at the end was a model that can accurately predict platelet usage with little error. The result of the forecast -- the optimal quantity of platelet units that minimizes wastage -- would then be used to order platelets from blood donation centers.

At the beginning of the study, the hospital wasted 1 in 10 platelet units on average daily. Using prediction from this new model, only 1 in 30 platelet units were wasted and there were no days where there was a shortage of platelets. If this model is used nationwide, it could translate to cost savings of over 80 million dollars.

Despite the potential cost savings, this model is still limited in use as it is based on the data from one hospital. Every hospital has different patient needs and populations, so in order for the model to work in a different hospital, it needs to be trained again. But training the model means that data must be collected over long periods of time so that the model is as accurate as possible. It is also costly to collect and manage big data.

In spite of the limitations, this model still has the potential to save millions of dollars of costs if applied widely. Hopefully, in the future, you could see your neighborhood hospital using this mathematical model.

(807 words)

**References:** (APA citation style)

Guan, L., Tian, X., Gombar, S., Zemek, A. J., Krishnan, G., Scott, R., Narasimhan, B., Tibshirani, R. J., & Pham, T. D. (2017). Big data modeling to predict platelet usage and minimize wastage in a tertiary care system. *Proceedings of the National Academy of Sciences*, *114*(43), 11368–11373. https://doi.org/10.1073/pnas.1714097114

*Hematology glossary*. Hematology Glossary - Hematology.org. (n.d.). Retrieved September 11, 2022, from https://www.hematology.org/education/patients/blood-basics