Abstract

Infectious diseases can potentially develop into an outbreak due to transmission of virus by direct contact or caused by large scale exposure to undesirable environments or conditions. The longer duration that the outbreak goes undetected will result in greater adverse effects such as higher mortality and greater potential impact on tourism and trade. Therefore, it is important to detect the onset of an outbreak of these infectious diseases as early as possible and this serves as our main motivation for this thesis.

We chose to investigate on Gastro-Intestinal Disease (GID) in the military setting. The main purpose is to find the minimum number of GID cases that represents a potential outbreak in each medical center. The term 'potential outbreak' is used because an outbreak in the SAF will only be confirmed after verification by the investigation team. Currently, the SAF criterion that is used is not that useful as it is a fixed value while the distribution of data in every medical center is different. Therefore, we implemented an algorithm that considers the distribution of data.

We first obtained the data which comprises of the number of GID cases on a daily basis from start of 2010 until end of 2013. With this data, analysis is conducted on each medical center separately. Firstly, various Penalized Logistic Regression Models were fitted to the data of each medical centers. Next, a diagnostic test is conducted on each of these models. We computed the Area Under the ROC Curve (AUC) and checked the false positive rate at optimal threshold values on the ROC Curve for selected diagnostic tests so as to find the potential outbreak levels for each medical center.

We also introduced an existing algorithm to use as a benchmark to check the accuracy of the potential outbreak levels generated by the main algorithm. This existing algorithm is comprised of 4 different peak functions and involves statistical concepts such as conducting Simulation Study, finding Entropy and application of Chebyshev Inequality.

A comparison between these two algorithms demonstrates that the potential outbreak levels generated by the main algorithm are pretty accurate. Therefore, we can conclude that the potential outbreak levels generated are an improvement to the SAF criterion and may detect a potential outbreak of GID as early as possible with a low false alarm rate.