The December Science café was held at the Bilik Serbaguna , Faculty of University Malaya on the 14th of December 2017. The talk was mainly focused on statistic in medicine and missing data problems which was conducted by Dr Manimalar Selvi Naicker from the department of Pathology and Prof Dr Sanjay Rampal a/l Lekhraj Rampal and Associate Prof Dr Karuthan Chinna from department of social and preventive medicine.

In a nutshell the talk emphasised on the misuse of p-values and missing data in clinical studies. Dr Manimalar Selvi Naicker started introducing her title “regression analysis with missing data: The problem with p-values” with a brief history on how p-values were initially introduced by Dr Ronald Fisher (agricultural-statistician) into agriculture in 1925 which was later used in medicine around the 1950’s. She informed that Sir Austin Bradford Hill (pilot turned statistician!) in 1965 had alerted the medical community about the misuse of the p-value in medicine for treating patients. Dr Manimalar further showed articles by the American statistical Association which have notified the usage of p-values in clinical study to be wrong. She also stated the reason p-values are not accepted in clinical studies are actually due to the misinterpretation of the concepts of the p-values and its inappropriateness with missing data. Moreover, to understand the significance of p-value one must first have strong foundation in mathematics.

Dr Manimalar also gave a little insight on p-value firstly explaining that p-values are not one but a composite value. She also stated sample size and effect size affect the p-value, however a missing date leads to changes to both sample size and effect size changes resulting in change to p-value. Interestingly Dr Manimalar mentioned the usage of a common mathematical formula used to calculate sample size to be wrongly applied in clinical studies as it was designed for factory statistics, whereby experiments are repeated on a daily bases. Before ending her talk Dr Manimalar stated the usage of t-test to also be misused as they correspond with sample size and are inversely proportional to p-value. In fact, missing data creates bias, lack of precision and results is us not being able to isolate the real magnitude of confounding. Also, when sample size increases the t-value also increase, therefore t-value is only reliable as statistical significance but not clinical significance. She also stated that data imputation was essentially data fabrication but if the amount of missing data is small it is better to impute data to save the dataset.

Associate prof Dr Karuthan took over next to discuss ‘Types of missing data’. He supported Dr Manimalar’s statement by justifying the insignificance of p-values and how missing data could affect it. He mentioned that sample size are needed to minimise type 1 and type 2 errors and how p-values could be easily manipulated by sample size. Subsequently he stated in presence of missing data if the sample size increased the p-value would decrease resulting in increase in power. On the contrary, in clinical study if the sample size is too big the effectiveness of the study will decrease according to Dr Karuthan. Furthermore, he also stated in a clinical study the goal is always to obtain evidence and not perfect data or statistics hence the p-value is insignificance in missing data.

Consequently, missing data and higher p-value decreases the power hence to obtain a significance study some imputation should take place said Dr Karuthan. Imputing data could be done at random to maintain the sample size. For instance, if the variable containing missing data is not vital one is suggested to omit the variable in the study. Similarly to perform regression with known value and predict the unknown value by taking the average. Finally, Dr karuthan stated that there is no right way to impute data hence one should always avoid missing data in a study.

Finally the talk ended with Dr Sanjay explaining further about missing data stating that it could be prevented with high data quality. He stressed that data collection should be always monitored to prevent loss of data and immediate action should be taken if any had occurred. On the other hand, Dr Sanjay mentioned briefly about ignorable and non-ignorable data. Ignorable data can be changed whereby non-ignorable data acts as study’s limitation. He also explained how missing complete at random is associated with unobserved variable while data missing at random is associated with observed variable. Hence, this effects the validity, precision and power of the study. Agreeing with Dr Karuthan, Dr Sanjay mentioned if the missing data is of variable not of interest than it should be just avoided and no imputation required.

Contacts:

Dr Manimalar Selvi Naicker, mala@ummc.edu.my

Associate Prof Karuthan Chinna, karuthan@um.edu.my

Professor Dr Sanjay Rampal Lekhraj Rampal, srampal@ummc.edu.my

Research Management Centre, Faculty of Medicine, UM, http://resfom.um.edu.my/