**JSPS Studentship Reflections**

I undertook a remote Studentship funded by the Jean Shanks Foundation and the Pathological Society in the summer of 2021. As part of my project, I was tasked with analysing a series of non-small cell lung cancer (NSCLC) pathology reports from several hundred patients enrolled in a large-scale research study. My supervisors were keen to examine the associations between variables such as tumour type, genomic characteristics of the tumour, with background lung disease, particularly interstitial lung disease.

This was a beneficial task as it allowed me to familiarise myself with the pathology of NSCLC, have direction from a specialist pathologist, and to learn about the structure of a pathology report. Since this was part of a multicentre study, there was considerable heterogeneity in the diagnostic reports, and this was challenging to navigate. There were several discussions around data quality and how to manage this during analysis. This experience has highlighted the limitations of large-scale analyses; they are only as good as the data provided.

Part of the discussion on missing data concerned the possibility of reassessing slides to identify the presence or absence of certain pathologies, but this was deemed to be too time-consuming a task to complete the dataset. To my mind, this highlights the future relevance of machine learning in studying histology slides, since the information provided might offer a more comprehensive picture of results.

Following the studentship, I presented my work to my peers in poster form, and this was useful as I gained my first experience of writing a structured abstract and creating a poster. This was a useful opportunity to condense the work of the summer into a form that I could communicate to students without research experience.

More recently I have been able, with support, to perform statistical analyses to investigate associations. I continue to work on the project, utilising SPSS to plot the data and to perform appropriate statistical tests. At present, I have performed a one-way ANOVA test to quality control the data. The incidence of background lung diseases appears to be associated with poor lung function, and thus my classification of disease based on the pathology reports seems accurate. This supports continued use of this data for analysis. I am excited to learn more about biostatistics and software through plotting these data and performing further statistical tests.

Overall, this Studentship has been an invaluable opportunity through which I have learned about the pathology of a hugely prevalent disease and come to appreciate the importance of clinical and pathological data in analysing pathogenesis. I have come to appreciate the scale and scope of prospective studies and understand the relationships between the fields of genomics and immunology and histopathology. My supervisors have been very supportive, and it has been fascinating and beneficial to learn about their work and to receive instruction from them. Through the Studentship, I have been exposed to a clinical and research field I would not otherwise have considered, and it has been a privilege to receive this opportunity.

**Sumayyah Imran, UCL Medical School Year 3**