

Segmentation analysis and the recovery of queuing parameters via the Wasserstein distance: a study of administrative data for patients with chronic obstructive pulmonary disease

To whom it may concern,

We present a novel approach for dealing with lack of data as inputs to queuing theoretic models. All research software designed for the work and all data is made available for all to use according to the best open scientific principles.

The paper also presents a novel application of queue modelling within healthcare by considering COPD patients. Clustering of patients using an unsupervised algorithm ensures accurate groupings informed by the data. This is used in combination with the queuing model to conclude that no quick wins are possible for the particular healthcare system and in fact a public health intervention is required to improve the general health within the population.

A precise aspect of the work here is that the clustering algorithm identifies a specific group of patients for which an intervention would be most beneficial to the efficiency of the healthcare system.

All of the above is made possible by the major contribution of the paper: dealing with lack of data. Indeed, as is often the case in healthcare models actual service length data is not available due to delay of admission or delay of discharge. The approach here is to use the Wasserstein distance as a metric on length of stay distributions to recover the service length from the data.

This approach could be readily used in similar modelling exercises where accurate service length data is not available.

The word count for this manuscript (calculated with `texcount`) is 6,960.

Sincerely,

The authors