# Understanding COPD patients in the hospital system via administrative data

Henry Wilde, Vincent Knight, Jonathan Gillard

**Abstract**

This work presents an analysis of how patients with chronic obstructive pulmonary disorder (COPD) interact with the hospital system in South Wales.

## 1 Introduction

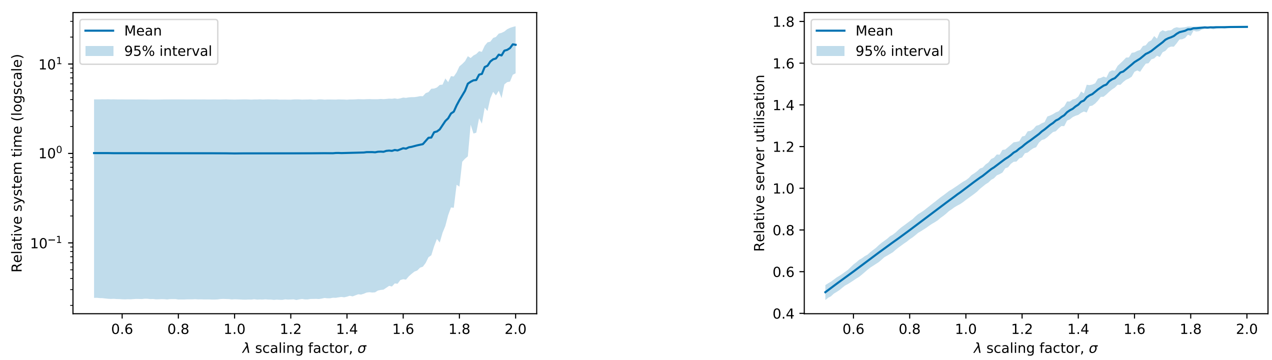
This introduction will briefly summarise the literature review for studying a patient corpus via clustering. Following this, a condensed data analysis is presented highlighting the main conclusions of the clustering and the overall benefits compared with traditional condition-treatment segmentation.

## 2 Estimating queuing parameters

Reiterate the objective of the paper — to model a COPD ward within a hospital — and draw attention to lack of fine-grain data. Lead into how this can be overcome with the Wasserstein distance (a lot of this has been written up in `nbs/wasserstein.ipynb`). A brief summary of how the parameter set is chosen and a nice image of the queue we are building. Close out the section with best and worst case parameter set plots.

## 3 Adjusting the queuing model

With the queuing model established and validated in Section 2, an investigation into the parameters of the model can be conducted. This section is comprised of several 'what-if' scenarios — a classic component of healthcare operational research — under this novel parameterisation. The outcomes of interest in this work are server (resource) utilisation and system times as these capture the driving forces of cost and flow as well as the overall state of the system, its staff and its patients. Specifically, the objective of these experiments is to address the following questions:




(a) (b)

Figure 1: Plots of $\sigma$ against relative (a) system time and (b) server utilisation.

- How would the system be affected by a change in overall patient arrivals?
- How is the system affected by a change in resource availability (i.e. a change in $c$)?
- How is the system affected by patients moving between clusters?

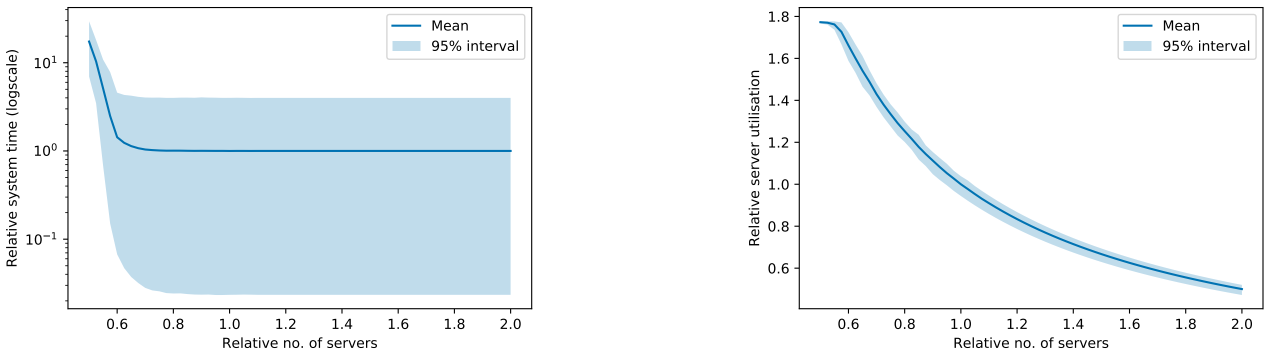
Owing to the nature of the observed data, the queuing model parameterisation and its assumptions, the effects of each scenario are given in relative terms with respect to the base case. The base case being those results generated from the best parameter set recorded in Table **??**. In particular, each piece of data in each scenario is simply scaled by the corresponding median value in the base case.

As mentioned in Section 1, the source code used throughout this work is available online and has been archived. In addition to this, the datasets generated from the simulations in this section have been archived.

## 3.1 Changes to overall patient arrivals

Changes in overall patient arrivals to a queue reflect real-world scenarios where some stimulus is improving (or worsening) the condition of the patient population. Examples of positive stimuli include increased community care and campaigns against harmful behaviours such as smoking. Within this model, overall patient arrivals are altered using a scaling factor denoted by $\sigma \in \mathbb{R}$. This scaling factor is applied to the model by multiplying each cluster's arrival rate by $\sigma$. That is, for cluster $i$, its new arrival rate, $\hat{\lambda}_i$, is given by:

$$\hat{\lambda}_i = \sigma \lambda_i \tag{1}$$




(a) (b)

Figure 2: Plots of the relative number of servers against relative (a) system time and (b) server utilisation.

### 3.2 Changes to resource availability

As is discussed in Section **??**, the resource availability of the system is captured by the number of parallel servers in the system, $c$. Therefore, to modify the overall resource availability, only the number of servers need be changed. This kind of sensitivity analysis is usually done to determine the opportunity cost of adding service capacity to a system, e.g. would adding $n$ servers sufficiently increase efficiency without exceeding a budget?

To reiterate the beginning of this section, all suitable parameters are given in relative terms. This includes the number of servers here. By doing this, the changes in resource availability are more easily seen, and overrule any concerns as to what a particular number of servers exactly reflects in the real world.

### 3.3 Moving patients between clusters

In order to model the effects of patients moving between two clusters, the assumption is that services remain the same (and so does each cluster's $p_i$) but their arrival rates are altered according to some transfer proportion. Consider two clusters indexed at $i, j$, and their respective arrival rates, $\lambda_i, \lambda_j$, and let $\delta \in [0, 1]$ denote the proportion of arrivals to be moved from cluster $i$ to cluster $j$. Then the new arrival rates for each cluster, denoted by $\hat{\lambda}_i, \hat{\lambda}_j$ respectively, are:

$$\hat{\lambda}_i = (1 - \delta)\, \lambda_i \quad \text{and} \quad \hat{\lambda}_j = \delta \lambda_i + \lambda_j \tag{2}$$

By moving patient arrivals between clusters in this way, the overall arrivals are left the same since the sum of the arrival rates is the same. Hence, the (relative) effect on server utilisation and system time can be measured independently.

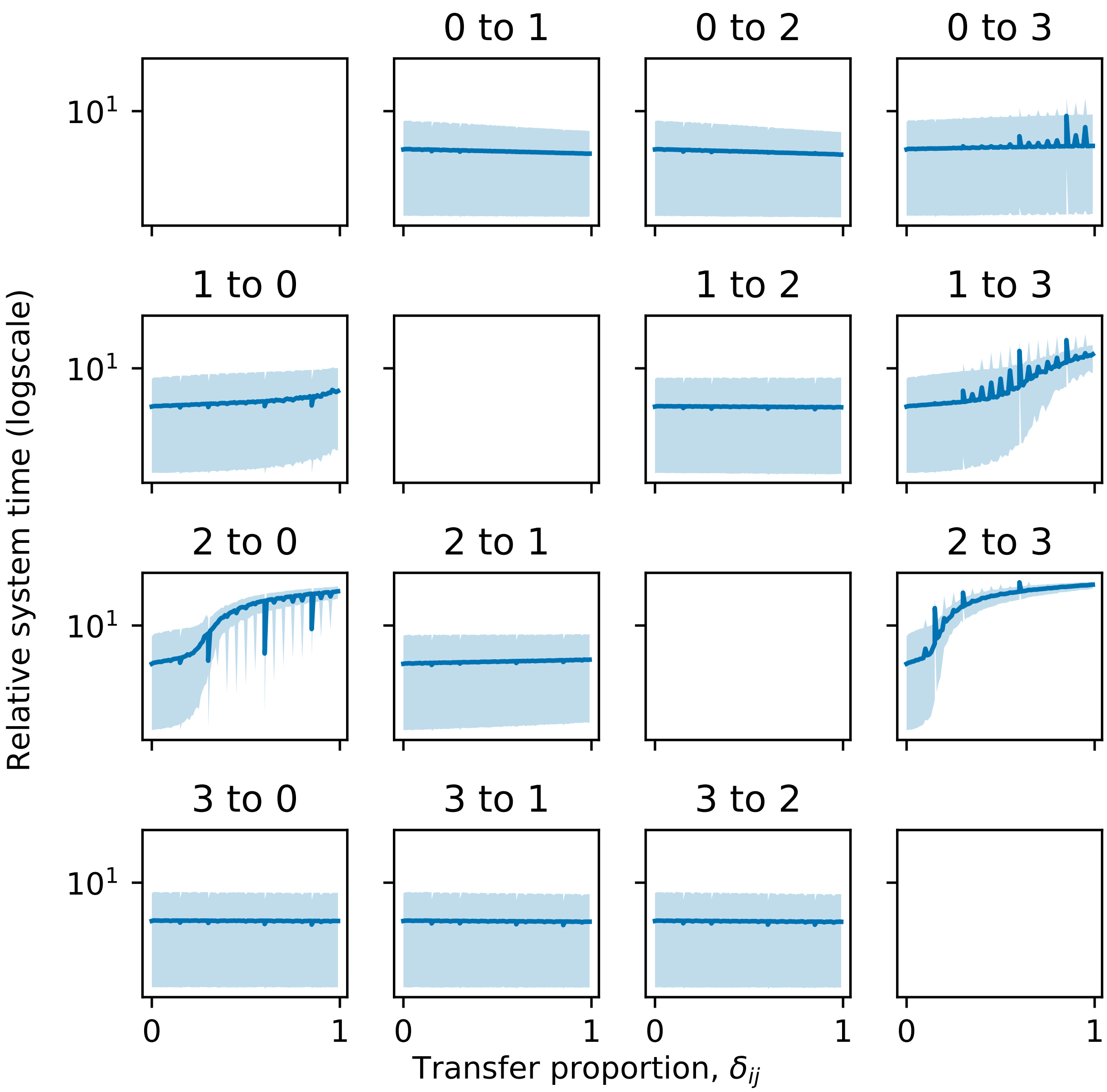



Figure 3: Plots of proportions of each cluster moving to another against relative system time.

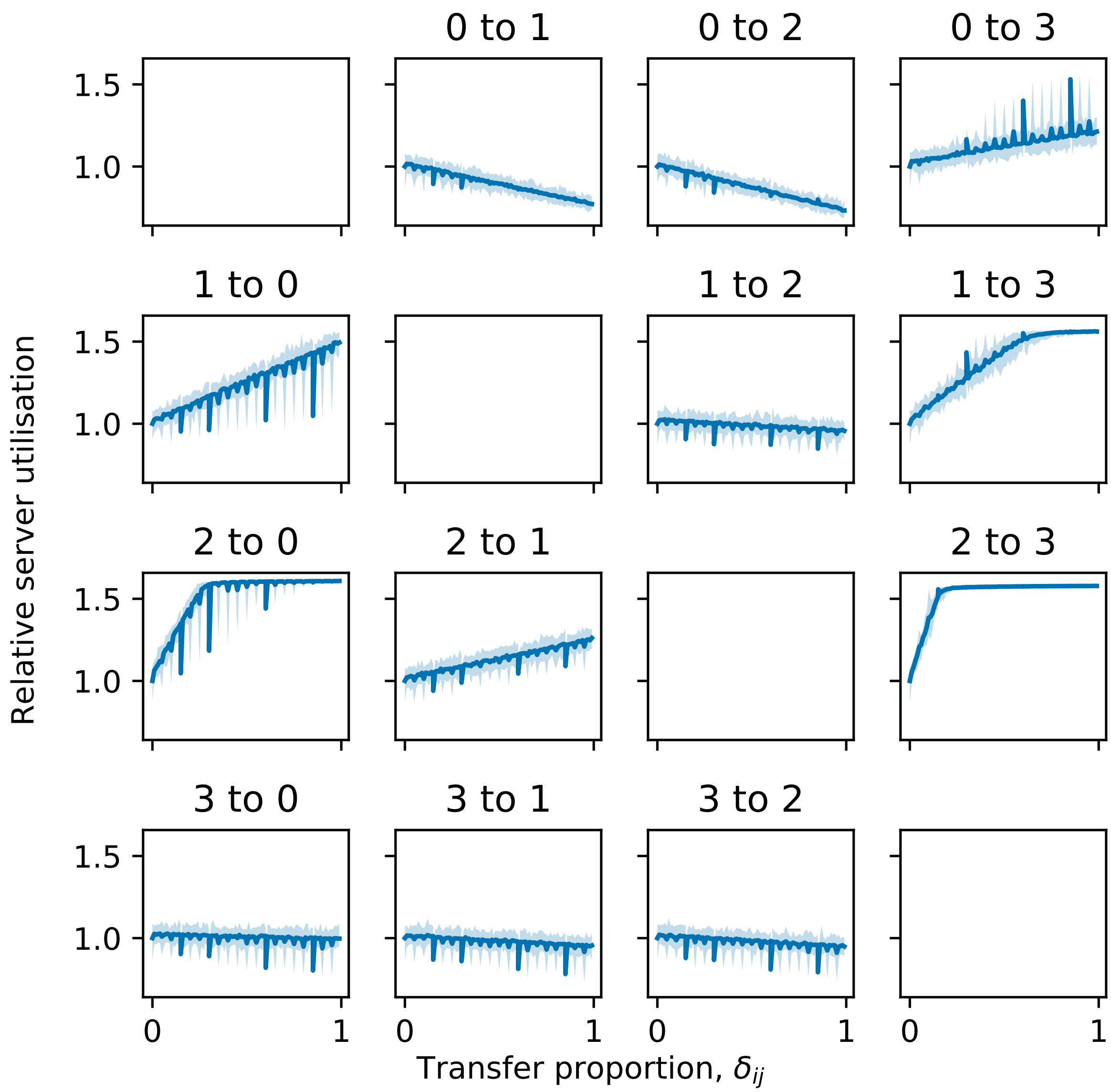



Figure 4: Plots of proportions of each cluster moving to another on relative server utilisation.

# 4 Conclusion

Summarise the findings and novelty of the paper: sensitivity analysis and queuing models are within reach despite a lack of data. The chosen modelling discipline for service times is very simplistic but can return good results (refer back to best-case parameter plot).