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

Population health research is becoming increasingly based on data-driven methods (as opposed to those designed solely by clinical experts) for patient-centred care through the advent of accessible software and a relative abundance of electronic data. Population health research concerns itself with better understanding the healthcare needs and behaviours of a population, so to further that end it can be helpful to find an appropriate segmentation of that population; such a segmentation allows for finer-grained analysis of groups in the population that share some form of homogeneity. One commonly used method for such patient-centred analysis is that of patient flow and their interaction with the healthcare system.

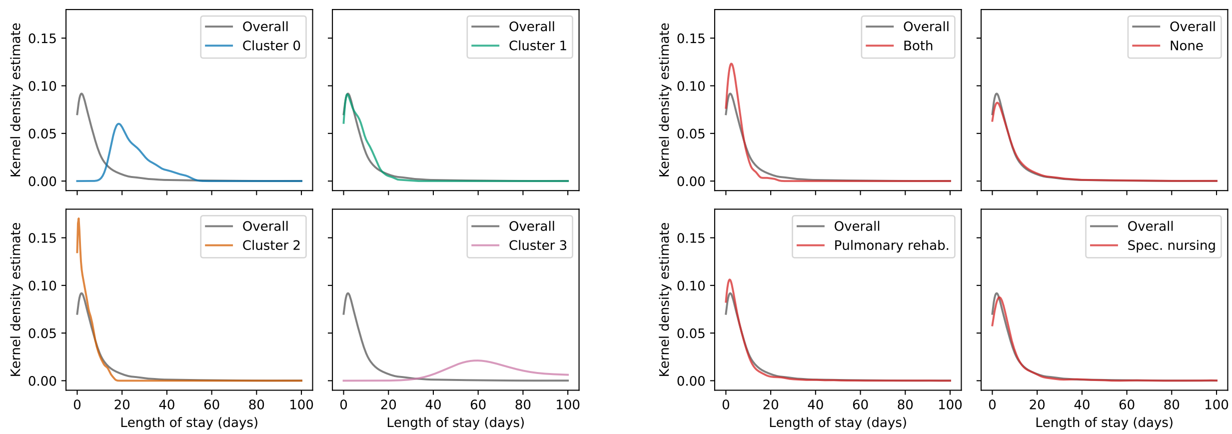
However, this process relies heavily on detailed data — about both the system and the population within that system — which may limit research where sophisticated data pipelines are not yet in place. This work demonstrates how this issue may be overcome using administrative, spell-level hospital data to build a patient clustering that feeds into a multi-class queuing model. Specifically, this work examines patient records from the NHS Wales Cwm Taf Morgannwg University Health Board (UHB) that present chronic obstructive pulmonary disease (COPD). COPD is of particular interest to Cwm Taf Morgannwg as it was found that they had the highest prevalence of the condition across all the Welsh health boards in an internal report by NHS Wales.

1.1 Literature review

Given the subject matter of this work, the relevant literature spans much of operational research in healthcare and the focus of this review is on the principal topics of segmentation analysis, the handling of missing data in healthcare settings and queuing theory applied to hospital systems.

Segmentation analysis allows for the targeted analysis of otherwise heterogeneous datasets and encompasses several techniques from operational research, statistics and machine learning. One of the most attractive qualities simplified communication of patient needs to stakeholders within a healthcare system [4, 6]. For instance, profiling often forms part of the wider analysis where each segment can be summarised in a phrase or infographic [3, 5]. The review for this work identified three groups of characteristics used to cluster a patient population: a patient's system utilisation metrics, their clinical attributes and their pathway. The latter does not strictly segment the patients directly but rather groups their movements through a healthcare system. This is typically done via process mining. [1] and [2] demonstrate how this technique can help to improve the efficiency of a hospital system rather than tackling the issue of patient-centred care as is the focus of this work.

Clustering is the process of separating the instances of a dataset into distinct parts to maximise homogeneity within each cluster and heterogeneity between each cluster. Within healthcare, and particularly population health, this concept has many applications. When applying this to a population, some form of patient corpus is considered the *de facto* dataset



(a) (b)

Figure 1: Kernel density estimate plots for length of stay by (a) cluster and (b) intervention

but the choice of characteristics with which to describe a patient or record remains open.

1.2 Overview of the dataset

The dataset used in this work was provided by the Cwm Taf Morgannwg UHB as part of an ongoing research project with the authors. The dataset contains a spell-level summary of 5,243 patients presenting COPD from February 2011 through March 2019 totalling 10,881 spells.

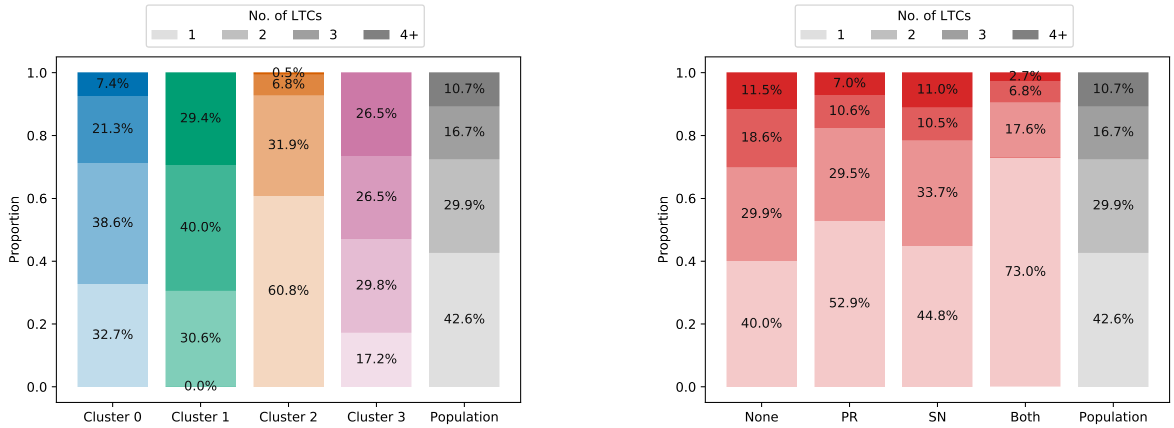
1.3 Cluster analysis

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.

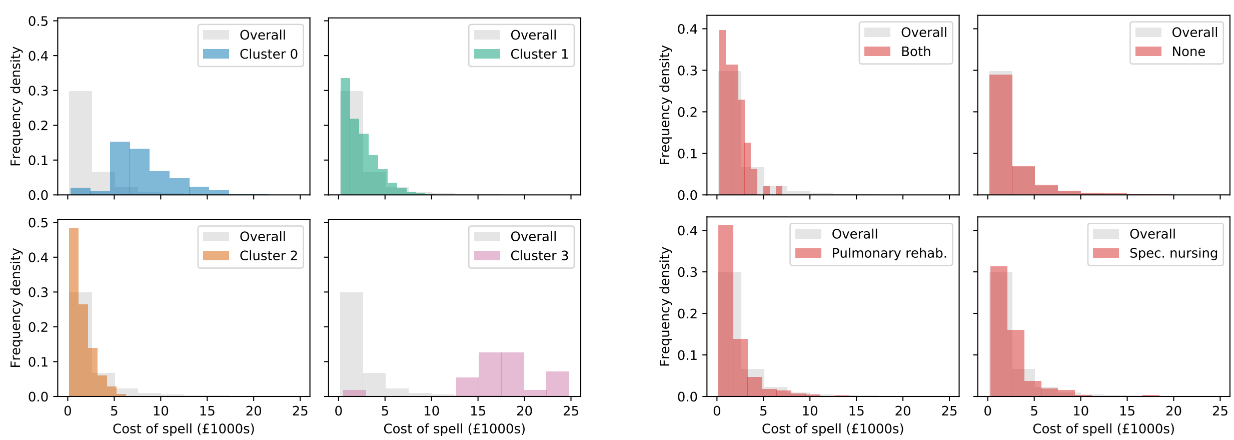
		Cluster 0	1	2	3	Population (mean)
Characteristics	**Mean spell cost**	8083.69	2312.39	1509.22	17847.80	2280.54
	Mean age	76.15	77.16	71.01	81.36	72.22
	Mean COPD adm. in year	1.91	1.51	1.31	1.98	1.29
	Min. LOS	12.82	-0.00	-0.02	48.82	5.41
	Mean LOS	25.30	6.45	3.79	74.65	7.47
	Max. LOS	51.36	30.86	16.94	224.93	10.40
	Median no. of LTCs	2.00	3.00	1.00	3.00	2.00
	Median no. ICDs	9.00	8.00	5.00	11.00	6.58
	Median CCI	9.00	20.00	4.00	18.00	9.72
Intervention prevalence	**None, %**	80.26	83.40	65.76	89.81	70.95
	Pulmonary rehab., %	15.77	13.41	27.96	8.92	23.66
	Spec. nursing, %	3.78	2.91	4.63	1.27	4.16
	Both, %	0.18	0.29	1.66	0.00	1.22
LTC prevalence	**Pulmonary disease, %**	100.00	100.00	100.00	100.00	100.00
	Diabetes, %	20.12	28.70	16.71	25.83	20.07
	CHF, %	14.47	54.62	0.00	27.15	14.83
	AMI, %	14.35	23.42	10.86	16.56	14.28
	Renal disease, %	8.59	22.52	3.02	18.54	8.50
	Cancer, %	8.12	13.69	4.11	10.60	6.94
	Dementia, %	7.53	20.08	0.00	25.17	6.10
	CVA, %	9.65	15.01	1.21	19.21	5.85
	PVD, %	5.06	8.90	3.18	5.96	4.78
	CTD, %	5.18	5.07	4.01	4.64	4.42
	Obesity, %	2.82	3.75	2.22	7.95	2.78
	Metastatic cancer, %	1.88	5.70	0.00	0.66	1.54
	Paraplegia, %	1.29	3.96	0.25	0.66	1.23
	Sepsis, %	2.12	1.25	0.25	1.99	0.76
	Peptic ulcer, %	1.76	1.04	0.35	1.32	0.72
	Diabetic compl., %	0.24	0.69	0.33	1.99	0.44
	Liver disease, %	0.35	0.56	0.33	0.00	0.37
	Severe liver disease, %	0.24	0.63	0.00	0.00	0.17
	C. diff, %	0.82	0.14	0.03	0.66	0.17
	MRSA, %	0.35	0.07	0.05	1.32	0.12
	HIV, %	0.00	0.00	0.03	0.00	0.02

Table 1: A summary of patient-level clinical attributes and disease prevalence by cluster and by population



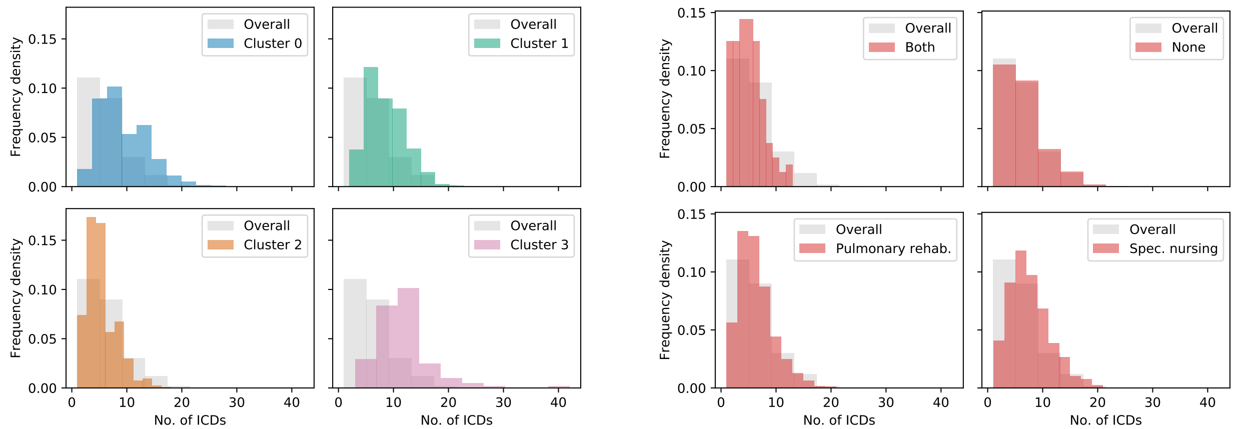
(a) (b)

Figure 2: Proportions of concurrent LTC counts presented by patients by (a) cluster and (b) intervention



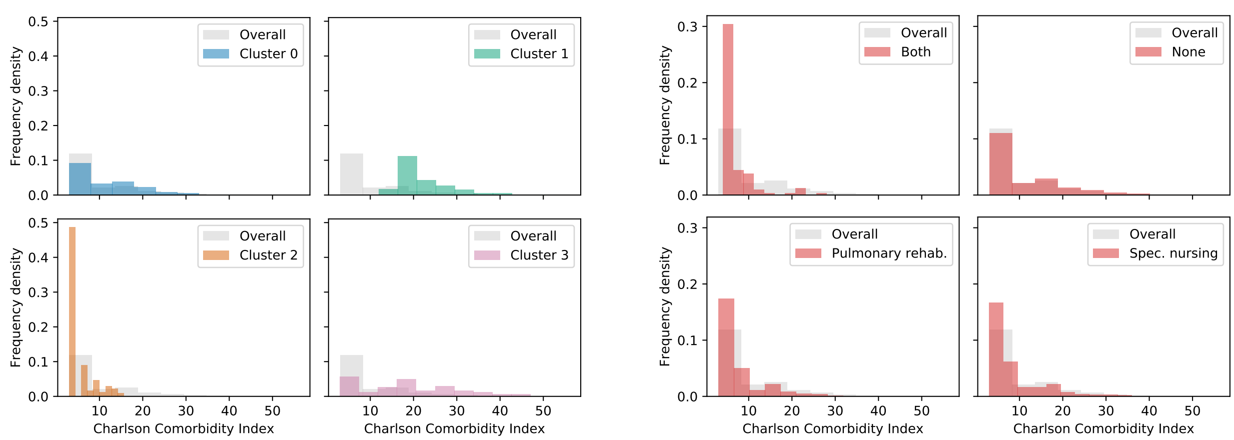
(a) (b)

Figure 3: Histograms for spell costs by (a) cluster and (b) intervention



(a) (b)

Figure 4: Histograms for number of ICDs by (a) cluster and (b) intervention



(a) (b)

Figure 5: Histograms for CCI by (a) cluster and (b) intervention

3 Adjusting the queuing model

Body of the writing and plots come here. What can we see in the what-if scenarios? The main scenarios are:

- How would server utilisation (i.e. resource consumption) be affected by an increase in overall patient arrivals?
- How is the system affected by certain types of patients (e.g. short-stay, low-impact) arriving less frequently?
- What are the sensitivities of mean system times and server utilisation based on a change in c?

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).

References

[1] I. V. Arnolds and D. Gartner. Improving hospital layout planning through clinical pathway mining. *Annals of Operations Research*, 263:453 – 477, 2018. doi: https://doi.org/10.1007/s10479-017-2485-4.

[2] P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas, and N. Matsatsinis. Supporting healthcare management decisions via robust clustering of event logs. *Knowledge-Based Systems*, 84:203 – 213, 2015. doi: https://doi.org/10.1016/j.knosys.2015.04.012.

[3] S. Vuik, E. Mayer, and A. Darzi. A quantitative evidence base for population health: Applying utilization-based cluster analysis to segment a patient population. *Population Health Metrics*, 14, 2016. doi: 10.1186/s12963-016-0115-z.

[4] S. I. Vuik, E. K. Mayer, and A. Darzi. Patient segmentation analysis offers significant benefits for integrated care and support. *Health Affairs*, 35(5):769–775, 2016. doi: 10.1377/hlthaff.2015.1311.

[5] S. Yan, B. J. J. Seng, Y. H. Kwan, C. S. Tan, J. H. M. Quah, J. Thumboo, and L. L. Low. Identifying heterogeneous health profiles of primary care utilizers and their differential healthcare utilization and mortality – a retrospective cohort study. *BMC Family Practice*, 20(54), 2019. doi: https://doi.org/10.1186/s12875-019-0939-2.

[6] S. Yoon, H. Goh, Y. H. Kwan, J. Thumboo, and L. L. Low. Identifying optimal indicators and purposes of population segmentation through engagement of key stakeholders: A qualitative study. *Health Res Policy Syst.*, 18(1):26, 2020. doi: 10.1186/s12961-019-0519-x.