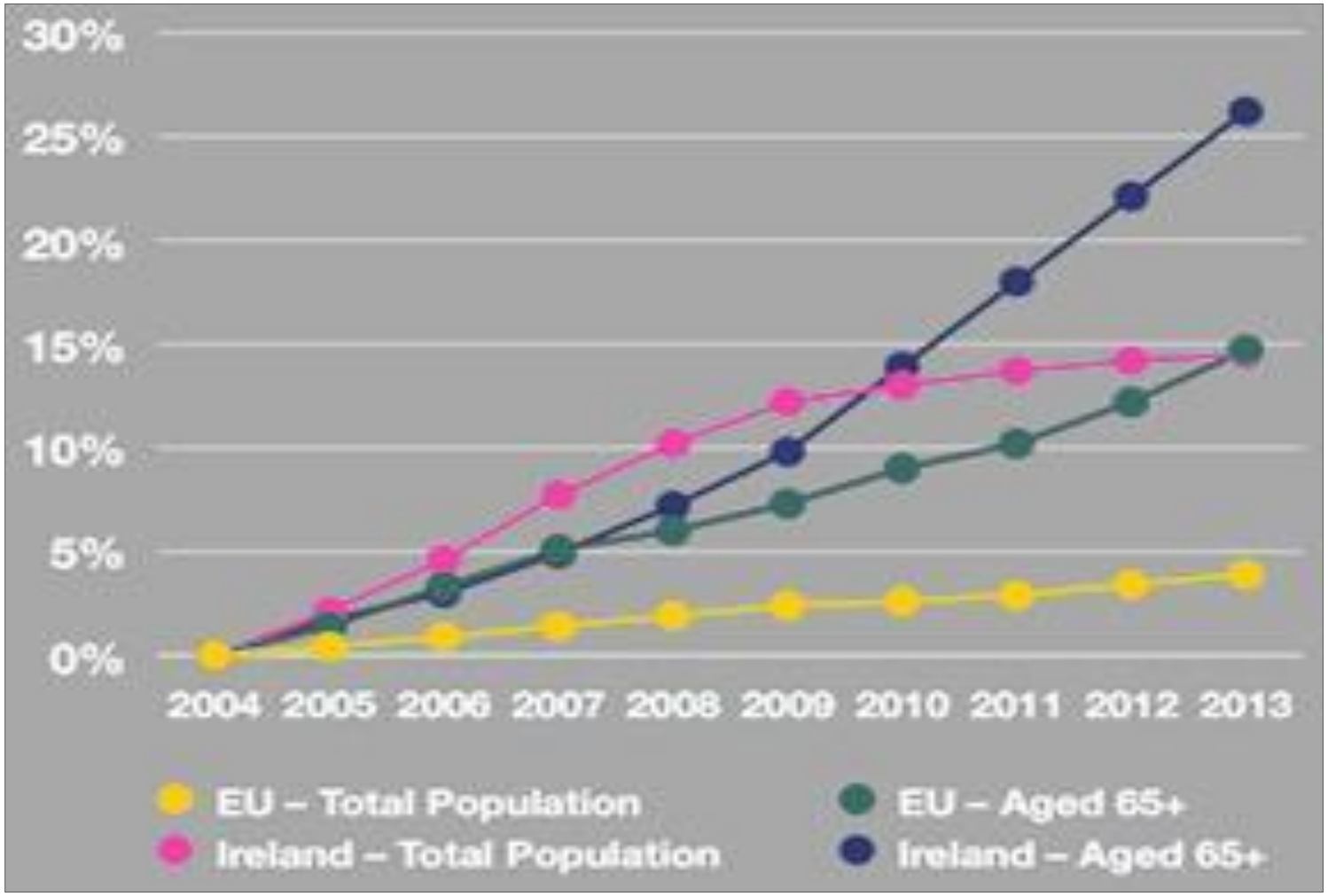
Data-Driven Patient Segmentation Using K-Means Clustering: The Case Of Hip Fracture Care In Ireland

https://dl.acm.org/citation.cfm?id=3014874

Mahmoud Elbattah, Owen Molloy
mahmoud.elbattah@nuigalway.ie

National University of Ireland, Galway

Challenge to Healthcare: Population Ageing



Source : Health Service Executive. Annual Report and Financial Statements, 2014.

Our Focus: Hip Fracture Care in Ireland

- A good exemplar of elderly healthcare.
- Exponentially increasing with age.[1]
- Identified as one of the most serious injuries resulting in lengthy hospital admissions and high costs.[2]
- High quality data available through the Irish Hip Fracture Database (IHFD).

Sources :[1] Gullberg, B., Johnell, O. and Kanis, J.A., 1997. World-wide projections for hip fracture. Osteoporosis international, 7(5), pp.407-413.
[2]http://www.hse.ie/eng/services/publications/olderpeople/Executive_Summary_Strategy_to_Prevent_Falls_and_Fractures_in_Ireland%E2%80%99s_Ageing_Population.pdf

Our Focus: Hip Fracture Care in Ireland

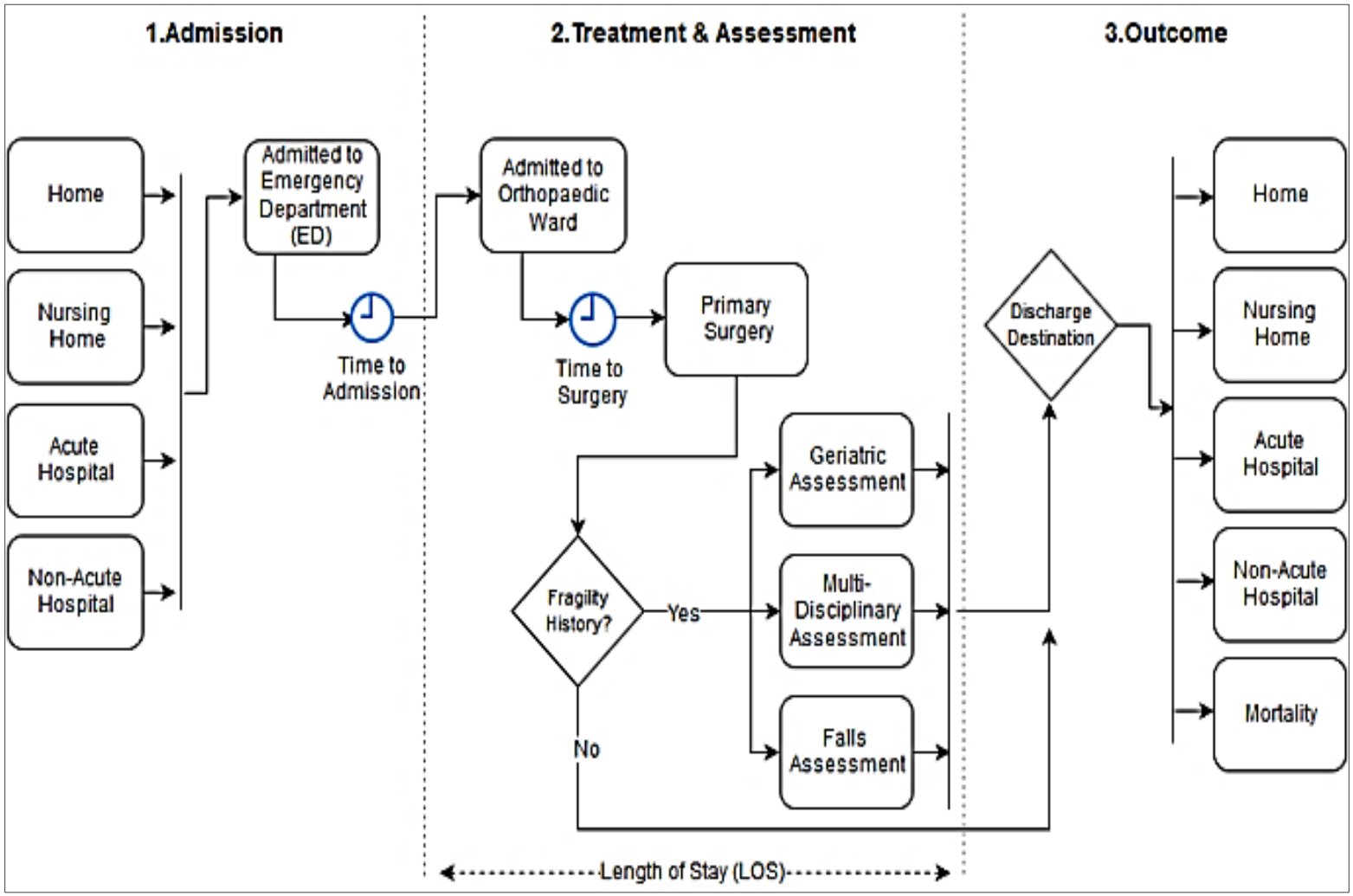


Figure1. Elderly Patient Journey

Questions of Interest

Principal Question

- What does the underlying structure of data infer in terms of composing coherent clusters of elderly patients based on specific similarity measures?

Further Questions:

- How do clusters vary with respect to patient characteristics such as age, gender or fragility history for example?
- How do clusters vary with respect to patient outcomes in terms of LOS and discharge destination?
- How do clusters vary with respect to the demographic profile of patients?
- Is there a possible correlation between patient outcomes and other care-related factors, such as “Time to Surgery” for example?

Data Description

- Irish Hip Fracture Database (IHFD).
- Patient records in the years 2013-2014.
- Patients aged 60 and over.
- 38 data fields such as gender, age, type of fracture, date of admission, and LOS.

Data Anomalies: Outliers

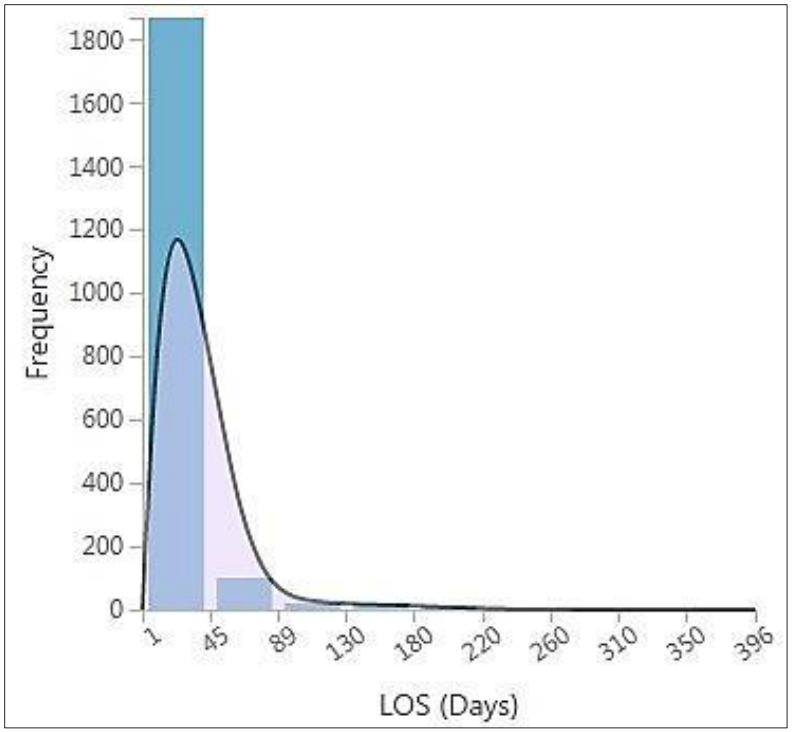


Figure 2. Histogram and probability density of the LOS variable. The outliers can be observed when the LOS becomes longer than 40 days.

Data Anomalies: Imbalances

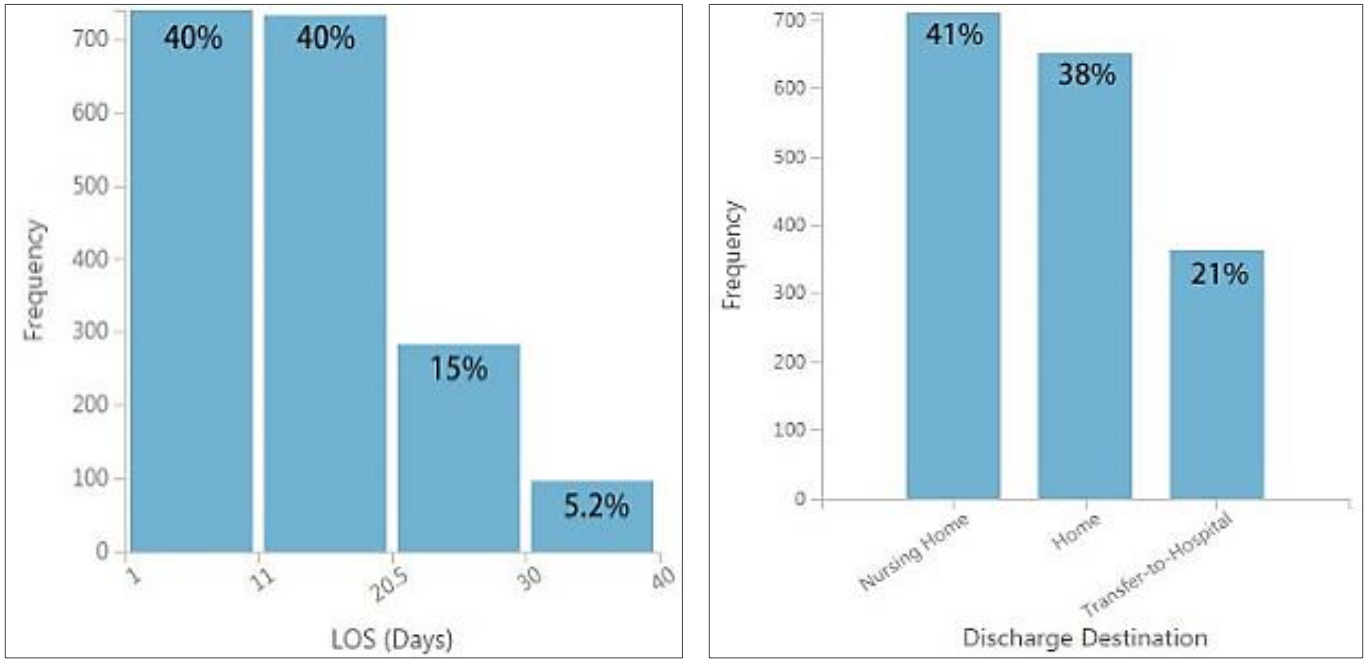
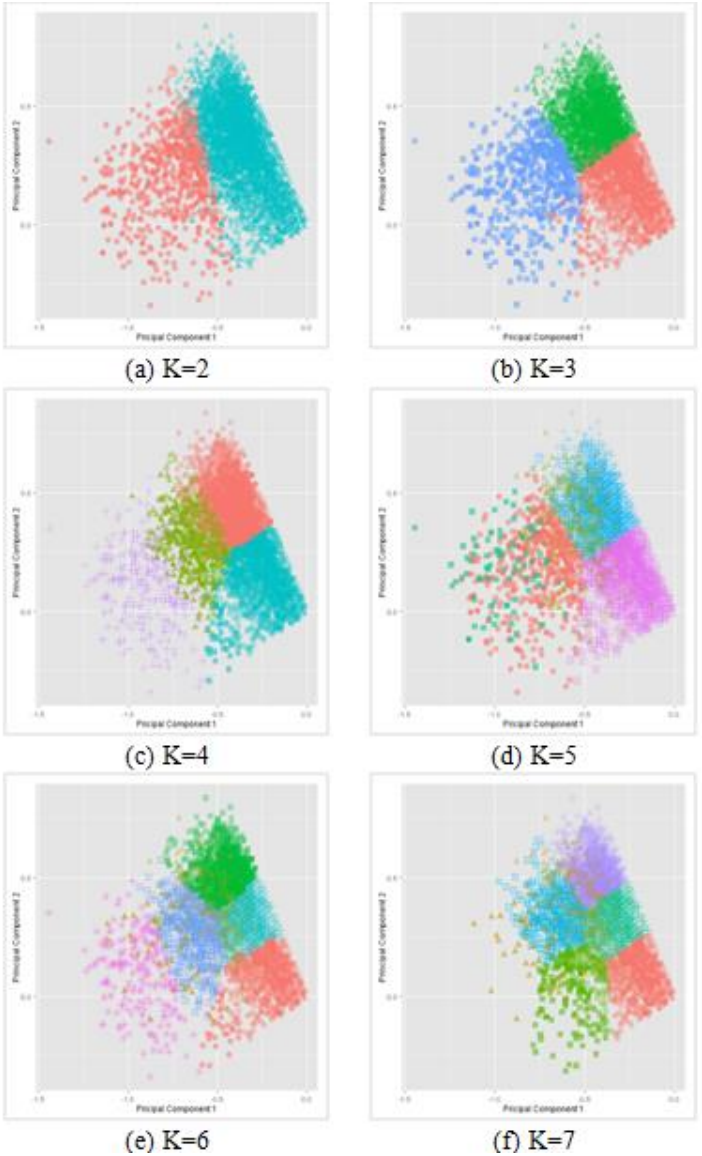


Figure 3. The imbalanced training samples, where figures (a) and (b) plot histograms of inpatient LOS and discharge destination respectively.

K-Means Clustering Experiments



(a) K=2 (b) K=3

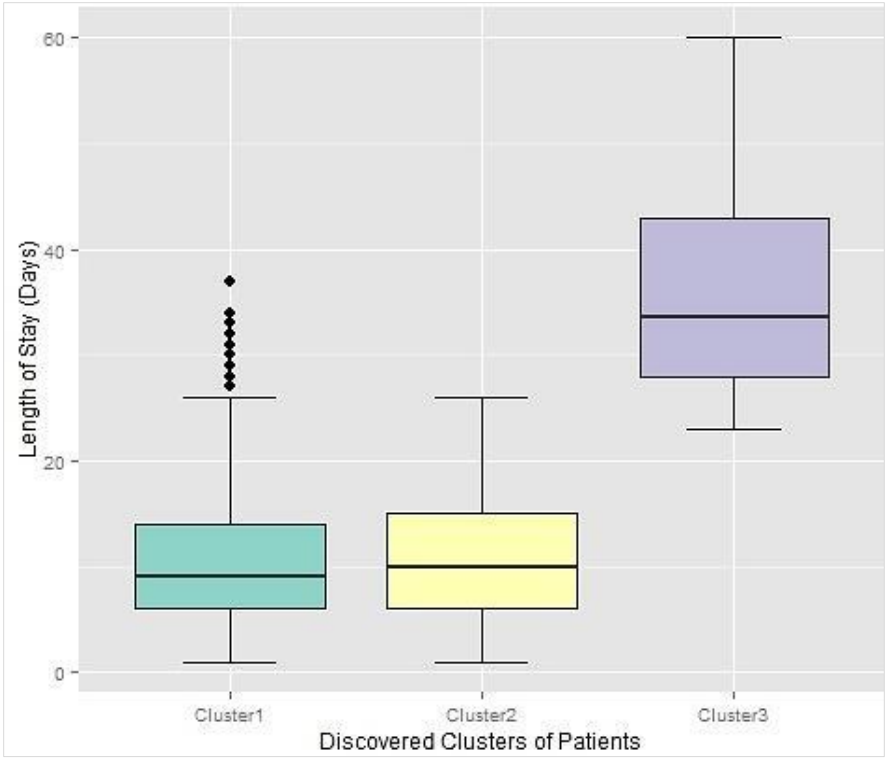
(c) K=4 (d) K=5

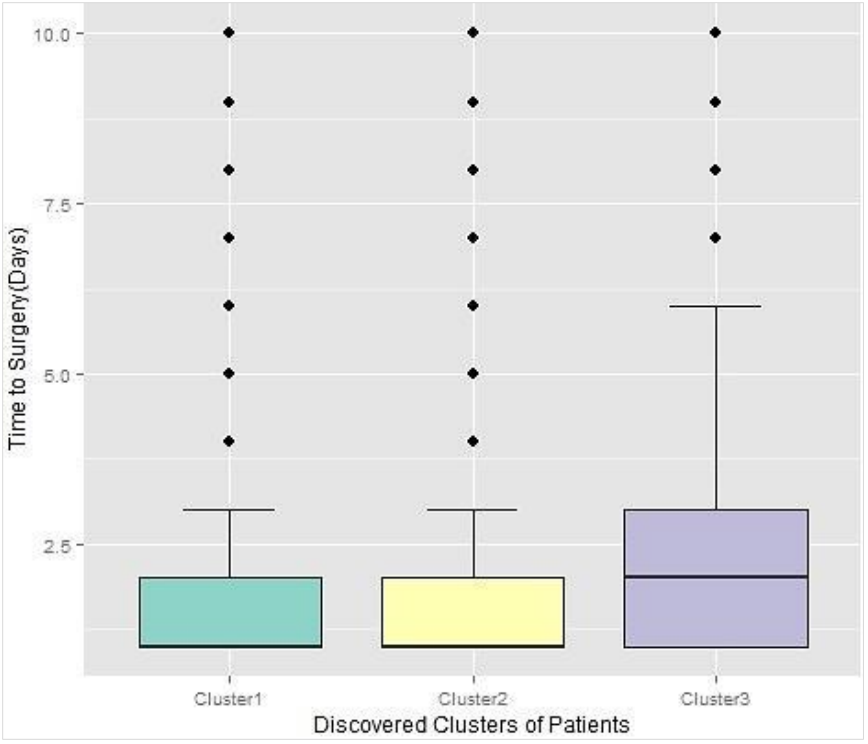
(e) K=6 (f) K=7

Patients clustered based on:

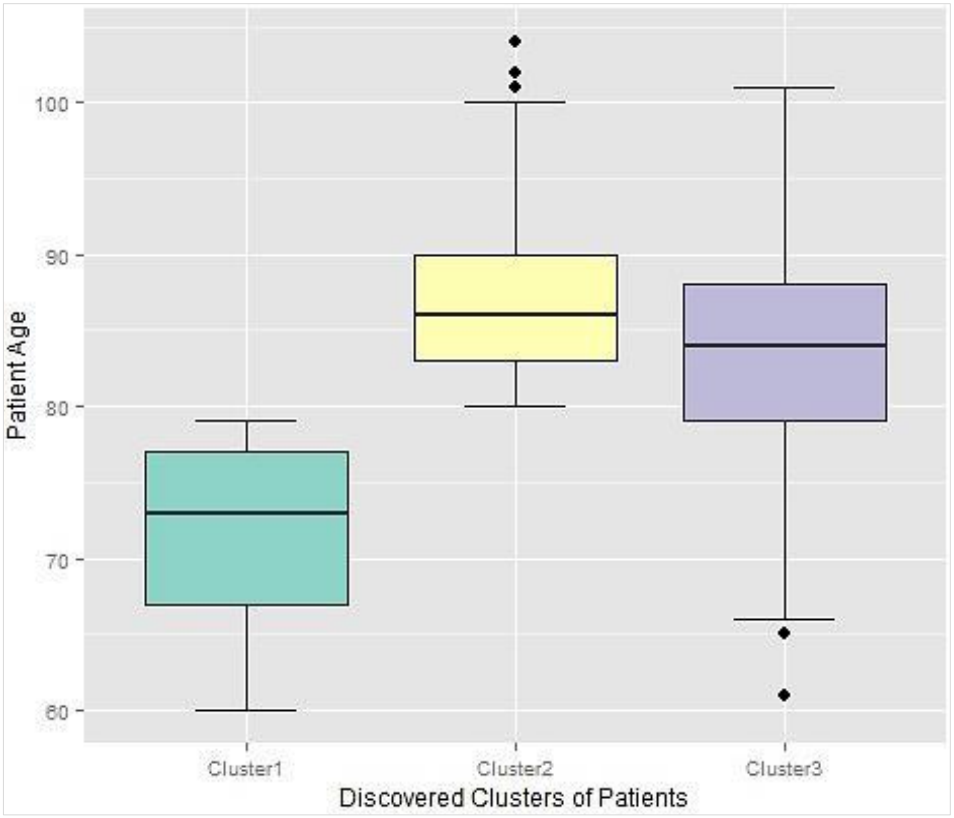
- LOS
- Age
- Time to Surgery

Cluster Analysis: LOS & Time to Surgery





Cluster Analysis: Age & Discharge Dest.



Conclusions

- The study presented an example of data-guided segmentation of patients using unsupervised machine learning.
- The group of patients who experienced longer periods of time to surgery, tended to have a considerably longer length of stay. This conforms well with the standards of hip fracture care.

Conclusions (cont'd)

- The two clusters that comprised relatively older patients consistently had the highest proportion of patients discharged to long-stay care facilities such as nursing homes.
- This can translate into the significance of early intervention programs for very elderly patients, which might help reduce highly costs of prolonged stays in nursing homes.

Acknowledgements

- National Office of Clinical Audit (NOCA), Ireland.

THANK YOU!

mahmoud.elbattah@nuigalway.ie