

# Understanding cost variation with clustering

---

Henry Wilde

*Supervised by:* Dr Jonathan Gillard and Dr Vincent Knight

February 8, 2019



GIG  
CYMRU  
NHS  
WALES

Bwrdd Iechyd Prifysgol  
Cwm Taf  
University Health Board

# The data

- The Cwm Taf University Health Board
- April 2012 through April 2017
- 2.4 million patient-episode records with 260 attributes

# The data

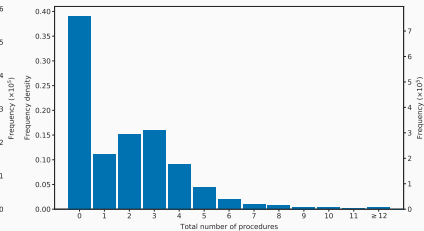
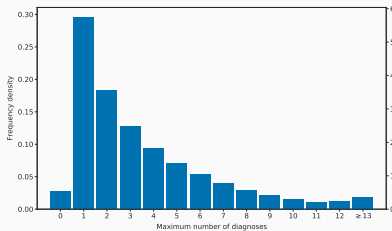
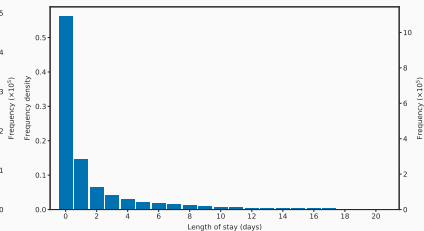
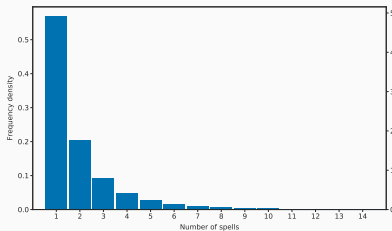
- The Cwm Taf University Health Board
- April 2012 through April 2017
- 2.4 million patient-episode records with 260 attributes

PATIENT ID	SPELL ID	EPISODE ID	Net Cost	Age (years)	HRG	Admission Date	Discharge Date	Length of Stay (days)
ID_123456	M1001	M1001-1	858.14	74	EA05Z	2015-05-06	2015-05-06	0.0
	M1211	M1211-1	333.95	74	FZ38F	2015-07-15	2015-08-01	17.0
		M1211-2	706.09	74	FZ38F	2015-07-15	2015-08-01	17.0
		M1211-3	8671.31	74	RC16Z	2015-07-15	2015-08-01	17.0

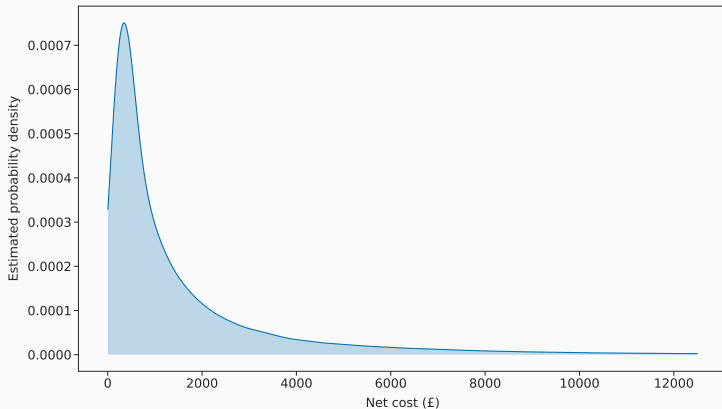
# **An overview of the data**

---

# An overview

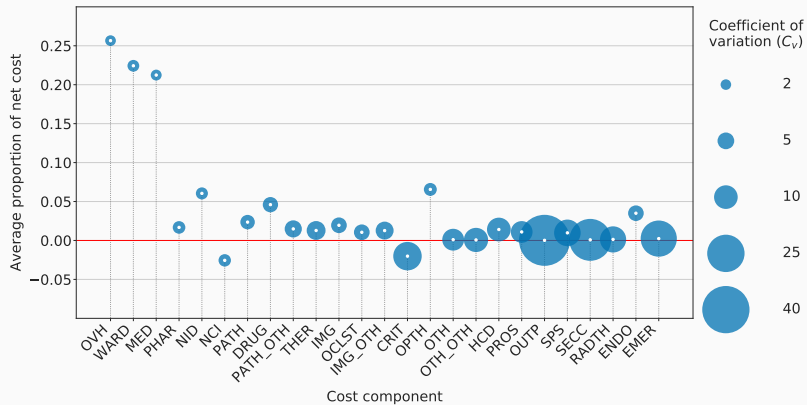


# An overview



	NetCost
mean	1,737.65
std	3,160.54
min	4.50
1%	62.55
25%	347.07
50%	745.51
75%	1,859.00
95%	6,554.91
99%	14,183.23
max	369,168.93

# An overview

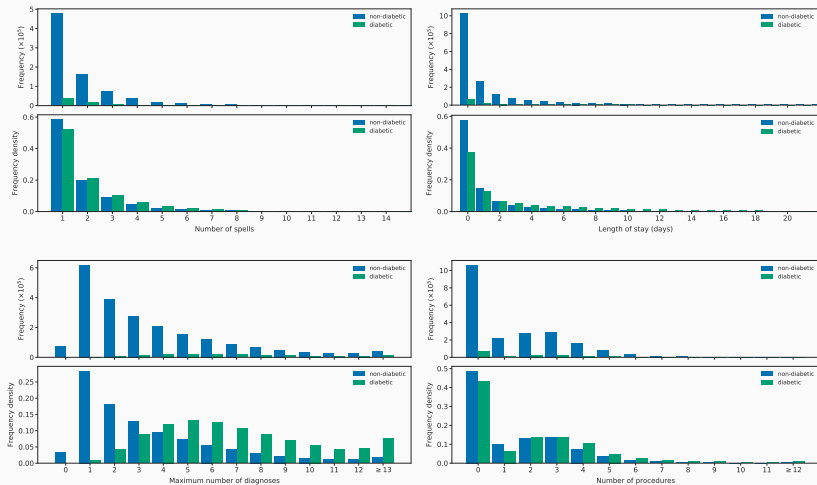


# Diabetic patient analysis

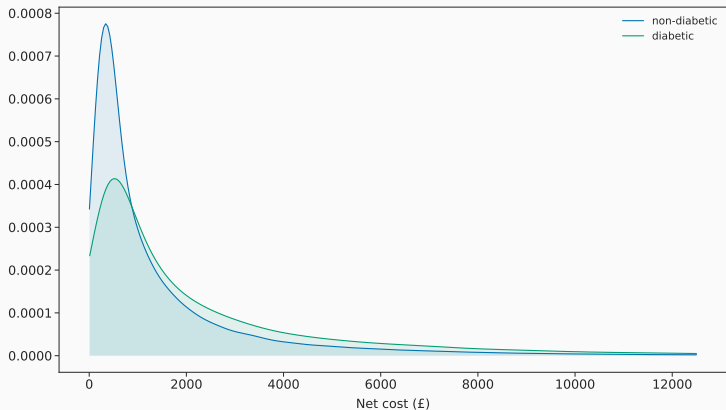
---



# Diabetic overview



# An overview



	Non-diabetic	Diabetic
mean	1,647.01	2,648.98
std	3,019.54	4,152.20
min	4.50	10.91
1%	62.55	139.65
25%	338.67	490.64
50%	709.33	1,227.95
75%	1,756.90	3,106.44
95%	6,179.92	9,591.06
99%	13,414.47	19,128.45
max	369,168.93	273,450.30

# An overview

