Abstract for ISB Conference (Deadline: 15 February 2017)
Current Character Count = 1750/1750, which includes characters in abstract title and abstract body.

Title
Data-driven cluster analysis for identifying groups within users of anti-osteoporosis medication, using real-world primary care data.

Authors: Sara Khalid, M Sanni Ali, Alan Silman, Daniel Prieto-Alhambra

Background and Objectives: Data-driven methods can be used for pattern recognition within a clinical population, enriching the existing analytical tools for clinical data analysis. We clustered anti-osteoporosis drug users with similar risk factors, to better determine the influence of therapy on their fracture risk. **Methods:** Using the SIDIAP Database (anonymized primary care records for >80% of Catalonian population), complete data from 37,996 incident users (2007-2014) of anti-osteoporosis drugs (AODs) were analysed. Hierarchical clustering [1] was used to derive sub-groups based on risk factors including age, gender, body mass index, smoking, drinking, Charlson index, steroid /sedative use, and fracture history. For each sub-group, on-treatment incident fracture rate (/100 person-years) was estimated. **Results:** Patients could be stratified into one of five clusters: 1) elderly multi-morbid men with high prevalence of smoking and drinking; 2) elderly women with high co-morbidity; 3) systemic steroid users; 4) secondary prevention (previous fracture history); and 5) younger (early post-menopause) women with low-medium co-morbidity. Group 4 had the highest fracture incidence (1.05 (95%CI 0.88-1.22), and 4.63 (95%CI 4.29-4.97), for hip and non-hip fractures, respectively); whilst Group 5 had lowest fracture incidence (0.15 (95%CI 0.11-0.20), 1.72 (95%CI 1.58-1.87), for hip and non-hip fractures, respectively). **Conclusion:** Cluster analysis identified sub-groups within AOD users, including expected patient groups but also a surprising cluster of younger women with low fracture risk, where therapy is probably not recommended. Further work should explore the usefulness of such data-driven algorithms for clinical data analysis.

Acknowledgements

References

[1] F. Murtagh. A Survey of Recent Advances in Hierarchical Clustering Algorithms. *The Computer Journal* (1983) 26 (4): 354-359.

[2] Fine, J.P. and R.J. Gray, A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association, 1999. 94(446): p. 496-509.

Disclosures/ Conflicts of interest

DPA's research group has received unrestricted research grants from Amgen and Laboratoires Servier.

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	4589	6901
Age > 60 years	0.99	1.00	0.88	0.85	0.24
Female	0.01	1.00	0.94	0.99	0.85
BMI > 30	0.29	0.36	0.40	0.34	0.27
Smoking	0.51	0.00	0.10	0.12	0.51
Drinking	0.46	0.18	0.14	0.06	0.27
Charlson index > 2	0.71	0.47	0.72	0.55	0.34
Steroid use	0.18	0.00	1.00	0.00	0.02
Sedative use	0.37	0.49	0.54	0.50	0.41

Fracture history	0.16	0.03	0.17	1.00	0.02
Hip BMD, median (IQR)	-1.88 (1.26)	-1.90 (1.30)	-2.00 (1.33)	-2.20 (1.20)	-1.60 (1.28)
Spine BMD, median (IQR)	-2.50 (1.42)	-2.60 (1.44)	-2.60 (1.34)	-2.69 (0.98)	-2.10 (1.34)

Table 1: baseline characteristics of patients within the derived clusters. Units: %, unless specified.

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	4589	6901
All fractures, SHR (95%CI), p-val	ref	1.71 (1.57, 1.87), 0	1.66 (1.47, 1.86), 0	2.4 (2.16, 2.66), 0	0.938 (0.838 1.05), 0.26
Hip fractures, SHR (95%CI),p-val	ref	1.41 (1.15, 1.72), 0.0011	1.38 (1.04, 1.83), 0.026	2.32 (1.82, 2.97), 1.2e-11	0.393 (0.282 0.548), 3.5e-08
Non-hip fractures, SHR (95%CI), p-val	ref	1.81 (1.64, 1.99), 0	1.74 (1.53, 1.98), 0	2.47 (2.21, 2.77), 0	1.07 (0.945 1.2), 0.3

Table 2: estimated relative fracture risk within the derived clusters, with cluster one as the reference group.

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	4589	6901
All fractures, absolute risk (%)	9.35	15.67	14.08	19.19	8.79
Hip fractures, absolute risk (%)	1.96	2.76	2.53	4.04	0.77
Non-hip fractures, absolute risk (%)	7.68	13.59	12.09	16.31	8.14

Table 3: estimated absolute fracture risk within the derived clusters.

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	4589	6901
All fractures incidence rate (95%CI) [per 100 py]	3.13 (2.88, 3.38)	1.22 (1.17, 1.27)	6.33 (5.79, 6.88)	5.53 (5.16, 5.89)	2.71 (2.49, 2.92)
Hip fractures incidence rate (95%CI) [per 100 py]	0.44 (0.36, 0.52)	0.57 (0.51, 0.63)	0.63 (0.49, 0.77)	1.05 (0.88, 1.22)	0.15 (0.11, 0.20)
Non-hip fractures, incidence rate (95%CI) [per 100 py]	1.79 (1.63, 1.95)	2.99 (2.86, 3.11)	3.20 (2.90, 3.50)	4.63 (4.29, 4.97)	1.72 (1.58, 1.87)

Table 4: estimated incidence rate of fracture within the derived clusters.

Code List

Sidiap_clusters_fx_risk_2017_01_03.R

Sidiap_cluster_bmd_2016_12_05.m

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	6901	4589
Age > 60 years	0.99	1.00	0.88	0.24	0.85

Female	0.01	1.00	0.94	0.85	0.99
BMI > 30	0.29	0.36	0.40	0.27	0.34
Smoking	0.51	0.00	0.10	0.51	0.12
Drinking	0.46	0.18	0.14	0.27	0.06
Charlson index > 2	0.71	0.47	0.72	0.34	0.55
Steroid use	0.18	0.00	1.00	0.02	0.00
Sedative use	0.37	0.49	0.54	0.41	0.50
Fracture history	0.16	0.03	0.17	0.02	1.00
Hip BMD, median (IQR)	-1.88 (1.26)	-1.90 (1.30)	-2.00 (1.33)	-1.60 (1.28)	-2.20 (1.20)
Spine BMD, median (IQR)	-2.50 (1.42)	-2.60 (1.44)	-2.60 (1.34)	-2.10 (1.34)	-2.69 (0.98)

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	6901	4589
All fractures, SHR (95%CI), p-val	ref	1.71 (1.57, 1.87), 0	1.66 (1.47, 1.86), 0	0.938 (0.838 1.05), 0.26	2.4 (2.16, 2.66), 0
Hip fractures, SHR (95%CI),p-val	ref	1.41 (1.15, 1.72), 0.0011	1.38 (1.04, 1.83), 0.026	0.393 (0.282 0.548), 3.5e-08	2.32 (1.82, 2.97), 1.2e-11
Non-hip fractures, SHR (95%CI), p-val	ref	1.81 (1.64, 1.99), 0	1.74 (1.53, 1.98), 0	1.07 (0.945 1.2), 0.3	2.47 (2.21, 2.77), 0

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
***n*. patients**	6661	16124	3721	6901	4589
All fractures incidence rate (95%CI) [per 100 py]	3.13 (2.88, 3.38)	1.22 (1.17, 1.27)	6.33 (5.79, 6.88)	2.71 (2.49, 2.92)	5.53 (5.16, 5.89)
Hip fractures incidence rate (95%CI) [per 100 py]	0.44 (0.36, 0.52)	0.57 (0.51, 0.63)	0.63 (0.49, 0.77)	0.15 (0.11, 0.20)	1.05 (0.88, 1.22)
Non-hip fractures, incidence rate (95%CI) [per 100 py]	1.79 (1.63, 1.95)	2.99 (2.86, 3.11)	3.20 (2.90, 3.50)	1.72 (1.58, 1.87)	4.63 (4.29, 4.97)