Patterns of drug use in actual clinical practice are not well-understood. We propose a data-driven approach to detect features from a set of clinical risk factors related to anti-osteoporosis drug (AOD) use. Our aim is to identify under-studied population sub-groups, so as to ultimately inform regulatory decision making.

**METHODS**

**Source** – The SIDIAP Database (anonymized, electronic primary care medical records for >80% of the population of Catalonia, Spain).

**Data** – Data from 37,996 incident users of AOD (2007-2014) with complete data on fracture risk factors was analysed.

**Variables** – 12 risk factors (based on NICE guidelines): age, gender, body mass index, smoking, alcohol drinking, Charlson index, fracture history, and current use of steroids, sedatives, HRT/contraceptives, anticoagulants, and aromatase inhibitors.

**Methods** – An unsupervised machine learning model called autoencoder was used to detect important features of the population. An autoencoder learns which of a given set of risk factors (features) “characterise” a dataset by attempting to reconstruct (i.e. estimating the identity function \( f(x) = x \)).

The autoencoder model detected 9 risk factors as important features: age, gender, body mass index, smoking, alcohol drinking, fracture history, co-morbidity, sedative use, and steroid use. Anticoagulants, aromatase inhibitors, and HRT/contraceptives were not considered important features. Results agreed with that of independent clinical expert opinion, on factors considered by GPs when prescribing AODs.

**CONCLUSION**

Our study proposed a novel approach for determining the characteristics of a population of AOD users. Further research is needed on the use of this method for drug utilisation research, especially for high-volume, high-dimensional data.

**Contact:** sara.khalid@ndorms.ox.ac.uk