



EDITORIAL COMMENT

Epidemiology of atrial fibrillation in Galicia: Insights from 'big data'



Epidemiologia da fibrilhação auricular na Galiza: perspectivas de uma vasta quantidade de dados

Diogo Magalhães Cavaco

Serviço de Cardiologia, Hospital de Santa Cruz, CHLO, Carnaxide, Portugal

Available online 23 January 2019

Epidemiological data on most medical conditions have traditionally been derived from registries and clinical trials. Although such data are obtained under highly controlled conditions, it is not certain whether the results correlate with key epidemiological indicators such as incidence, prevalence, mortality, morbidity and treatment status in a given population. The larger the number of subjects included, the greater the probability of obtaining an accurate result.

Over the last decade, there has been a rapid digital transformation in healthcare, with the increased use of electronic medical records, healthcare information systems, and handheld, wearable and smart devices. A massive quantity and variety of health-related data today is in digital form, including not only clinical information, but also data on genomics and proteomics, sociodemographics and insurance claims.¹

This kind of information is what is known as 'big data', many of the current applications of which are not related to medicine. It is by analyzing big data that companies such as Google, Amazon and Netflix are able to predict the behavior of their customers (what products they can be persuaded to

buy, what kind of advertisements or movies they want to see, and so on).

New data analysis methods have also been developed to handle big data. Machine learning automates analytical model building. Data mining uses machine learning processes for knowledge discovery and predictive modeling in large datasets, searching for hidden insights. Unlike traditional statistical methods, these tools do not start with a predefined model; the model is built at the same time as underlying patterns are discovered. Nevertheless, traditional statistical methods can also be applied to big data.²

Electronic medical records for entire regions and populations constitute a vast quantity of data and many epidemiological and clinical insights can be obtained from this information.

In the current issue of the *Journal*, Rodríguez-Mañero et al.³ present a study using electronic health records from the Spanish region of Galicia. The population is composed of around 380 000 individuals according to the population information technology system of the Galician regional health service. Information on diagnosis, treatment status and complications in atrial fibrillation patients was obtained from various sources, including primary health care, hospital discharge and pharmacy records. Traditional statistical methods were used to handle this large amount of data.

The authors were able to estimate the overall prevalence of atrial fibrillation in Galicia. The figure they obtained (2.08%) is similar to that obtained from Spanish registries.

DOI of original article: <https://doi.org/10.1016/j.repc.2018.06.008>

E-mail address: cavacodm@gmail.com

<https://doi.org/10.1016/j.repc.2018.12.002>

0870-2551/© 2019 Sociedade Portuguesa de Cardiologia. Published by Elsevier España, S.L.U.

This is an open access article under CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

This is not the first such study to be published. In an Israel-based population study, the prevalence of atrial fibrillation was 3.0% in individuals older than 21 years.⁴ As in other studies, the prevalence increases with age and is the same overall in men and women, although women tend to develop atrial fibrillation later in life.

The treatment status of patients diagnosed with atrial fibrillation was also assessed. Overall, two-thirds of the patients were anticoagulated, a slightly higher figure according to the authors than that reported in previous Spanish studies. Most of these patients (71.6%) were treated with vitamin K antagonists. As expected, the rate of oral anticoagulation prescription increased with higher CHA₂DS₂VASc scores.

There are some advantages of this kind of study derived from the longitudinal nature of the data (although they are not explored in the paper). It will be interesting to see the changes in patterns of anticoagulation. Will a higher percentage of patients be anticoagulated in the future? How will direct oral anticoagulants (DOACs) affect this? It is likely that in 2019, the number of patients anticoagulated with vitamin K antagonists will be lower compared to DOACs than in 2014.

With regard to mortality, the study also confirms data obtained from registries. Mortality of atrial fibrillation patients was high, and was related to the disease – patients with higher CHA₂DS₂VASc scores had higher mortality (confirming the power of this score), as did females and patients with dementia.

A limitation of this kind of study (as in any large-scale epidemiological database-based study) is variability in coding by multiple physicians at different steps of the process, which means the system is prone to error.

Although the information in this paper is not new, it confirms that epidemiological indicators obtained from clinical

registries and small epidemiological studies are of good quality and reliable.

The main strength of the study is, of course, the size of the database and the reliable coverage of the entire population of a particular region.

It will be interesting to see if genuine ‘big data’ applications will be developed for this subset of the Spanish population – more could be learned about the prognosis, unknown risk factors and outcomes of these patients using these new tools.

Conflicts of interest

The author has no conflicts of interest to declare.

References

1. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol.* 2016;13:350–9.
2. Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: an Acute Coronary Syndrome Israeli Survey data mining study. *Int J Cardiol.* 2017;246:7–13.
3. Rodríguez-Mañero M, López-Pardo E, Cordero-Fort A, et al. Prevalence and outcomes of atrial fibrillation in a European healthcare area gained through the processing of a health information technology system. *Rev Port Cardiol.* 2019;38:21–9.
4. Haim M, Hoshen M, Reges O, et al. Prospective national study of the prevalence, incidence, management and outcome of a large contemporary cohort of patients with incident non-valvular atrial fibrillation. *J Am Heart Assoc.* 2015;4:1–12.