Domain led time series analysis of cardiovascular disease using open data – does reduction in coronary disease increase heart failure prevalence?

Alicja Jasinska-Piadlo[1], Raymond Bond[2], Pardis Biglarbeigi[2], Patricia Campbell[3], D. McEneaney[1]

[1] Cardiovascular Research Department, Craigavon Area Hospital, Craigavon, UK
[2] Faculty of Computing, Engineering & Built Environment, Ulster University, Northern Ireland, UK
[3] Department of Cardiology, Craigavon Area Hospital, Craigavon, Uk

Abstract

Accessing healthcare data for research purposes is lengthy and requires a data controller and ethical approval. There are, however, open data that exist in the public domain that can be freely accessed and used for research purposes. We performed a time series analysis of cardiovascular disease trends in Northern Ireland during years 2010-2020. We accessed the Raw Disease Prevalence Data for NI from the Quality Outcomes Framework (QOF) published on the government website. Statistical analysis was performed to explore trends in prevalence for patients with coronary artery disease (CAD), heart failure (HF) and atrial fibrillation (AF). Data analysis was performed using MATLAB version R2021b. We observed a significant increase in the prevalence of patients with HF and AF against a steady decrease of coronary artery disease prevalence in NI population. Our analysis shows that publicly available data may have useful research value when research questions are asked by clinical experts. We advocate for the use of publicly available open data from government or other sources for research purposes and care quality improvement.

1. Introduction

Coronary artery disease (CAD) remains the leading cause of mortality in western countries [1]. In the UK, CAD affects approximately 2.6 million people [2]. Acute or chronic myocardial ischemia, a consequence of CAD, is one of major causes for developing heart failure (HF) with reduced left ventricular ejection fraction (LVEF). HF is associated with high mortality and lengthy recurrent hospital admissions. In the UK, data on prevalence of CAD and HF are collected by various organizations. Data on coronary events leading to myocardial infarction (MI) are routinely collected by UK-wide Myocardial Ischaemia National Audit Project (MINAP)[3]. British Cardiovascular Intervention Society (BICIS) collects data on percutaneous coronary interventions (PCI) [4]. Data on admissions of patients with HF in England and Wales are collected by the National Cardiac Audit Program (NICOR)[5]. In Scotland HF data are collected by the Information Services Division and published as an annual report - Scottish Heart Disease Statistics [6]. In Northern Ireland (NI) however there is no auditable registry collecting data on HF admissions or provision of HF secondary care. Nevertheless, there is an open access prevalence data of specific conditions in primary care published by Quality Outcome Framework (QOF) [7]. The QOF data are used to calculate payments for procedures within each clinical domain (6). Based on available statistics, the prevalence of CAD has been decreasing in Europe over past decades (citation) [2]. Most of this reduction is due to primary prevention measures including smoking cessation and targeted risk factor modification in primary care including management of blood pressure, cholesterol level and screening for diabetes. In addition to the decreasing number of acute coronary events such as acute myocardial infarction with ST segment elevation (STEMI) or without ST segment elevation (NSTEMI) the survival from these events has improved due to coronary interventions and secondary prevention measures including cardiac rehabilitation. Because of improved survival from myocardial events the likelihood of surviving albeit with heart failure increases. We hypothesize that HF prevalence increases as patients are surviving to a more advanced age with access to the modern evidence-based pharmacotherapy and device therapy like implantable cardiac defibrillators and cardiac resynchronization therapy. To test our hypothesis, we performed time series analysis of trends of CAD and HF in NI based on the QOF data. We further explored the differences of HF prevalence at the regional and granular level, looking at the health and social care trusts prevalence and the prevalence at the GP practice level.

2. Materials

The QOF Register with row prevalence data was available from the Department of Health website [7]. The register runs from 31st March to 31st March of the following year between years 2009 and 2020. The register gives the postal address of 359 GP practices in NI and includes prevalence data for the practice population. Register provides the prevalence of: coronary heart disease, heart failure, stroke, hypertension, diabetes, chronic obstructive pulmonary disease (COPD), epilepsy, hypothyroidism, cancer, depression, asthma, dementia, chronic kidney disease, atrial fibrillation (AF), obesity, palliative care, peripheral arterial disease and rheumatoid arthritis. Since the register began in 2004, 39 out of 359 GP practices were either closed or merged between themselves, creating bigger practices. Each GP practice in NI operates within one of 5 health and social care (HSC) trusts: Belfast, South-Eastern, Northern, Southern and Western Trust. Over the years, there have been changes to the criteria for the inclusion of certain clinical conditions and categories on the register and those changes are described in the QOF register's definitions section [7].

3. Methods and Results

In the first instance we used data from the register to plot the prevalence of conditions of interest (Fig. 1).

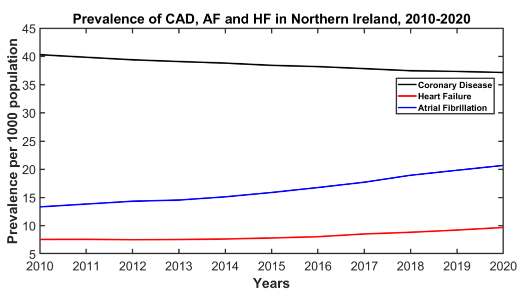

Figure 1. Prevalence per 1000 population of Heart Failure (HF), coronary artery disease (CAD) and Atrial Fibrillation (AF) in NI, 2010-2020.

We observed a small increase in the prevalence of HF around the year 2014 to 2015. Although the overall annual prevalence remained below 1% of the population. There was a decreasing trend of CAD prevalence, however AF prevalence has been on a steady increase since the beginning of the register. We then looked at each of 5 HSC trusts and plotted HF prevalence for each of them (Fig. 2), observing a quite rapid increase of HF prevalence since 2015.

To be able to see more granular data on the HF prevalence within each GP practice, we used histogram to look

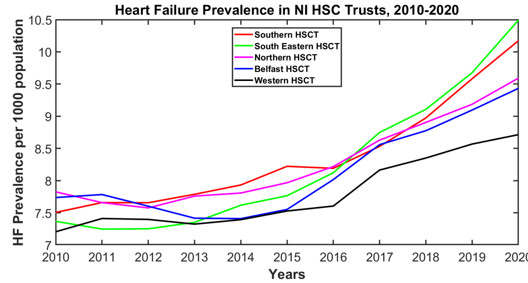

Figure 2. Prevalence of heart failure per 1000 population in each Health and Social Care Trust in NI, years 2010-2020.

for the HF prevalence distribution within practices (Fig. 3). The prevalence of HF in most GP practices was between

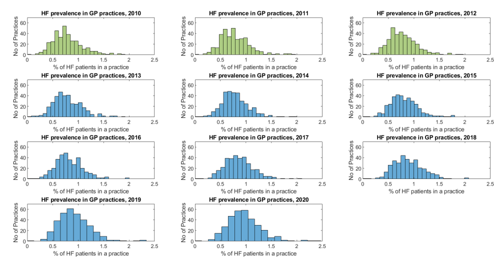

Figure 3. Histogram of annual prevalence of heart failure in a GP practice population, expressed by percentage, Northern Ireland, years 2010-2020.

0.5% to 1% of the practice population. This may suggest possible under-reporting and possible under-recognition of HF given that the WHO estimate the HF prevalence at 2% [2]. We used a Chi-squared test to test if there was a statistically significant difference between the frequencies of HF in consecutive years. We used proportions of the total number of HF patients in each year to the total NI population for that year to calculate p-value as in the following equation:

p = (n1+n2) / (N1+N2);

where:

p = p - value for the Chi squared test

n1 = number of patients with a condition in year No 1

n2 = number of patients with a condition in year (No 1) +1

N1 = total NI population for year No 1

N2 = total NI population for year (No 1) +1 (subsequent year)

Because the register inclusion criteria for HF have changed in 2013 [7], we have applied Chi square test from that year.

We achieved a drop in p-value (Fig. 4) for proportion of number of patients between year 2014 and 2015, which lead us to reject the null hypothesis and we were able to show that there was a significant increase in the proportion

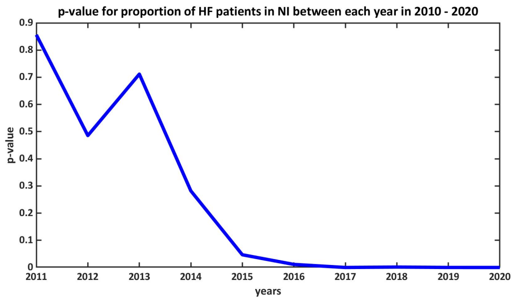

Figure 4. P-values for comparing heart failure prevalence rates for patients in consecutive years in, NI, 2010 -2019.

of patients with HF in consecutive years resulting in an increasing trend of HF prevalence in NI, see Table 1. Similarly, we were able to show a decreasing trend of p-value for AF prevalence, suggesting that there was a significant increase in AF patients within the total population. Contrary to HF and AF, there was a decreased trend of CAD in NI which was supported statistically by applying a Chi-square test to proportions of patients with CAD (Table 1).

Table 1. p-value for proportions of HF patients in consecutive years. * Can't use p-values for years 2010-2013 due to change of the HF register criteria in years 2012/2013.

Tax Year	p-value HF	p-value CAD	p-value AF
2010/2011	*	0.0240	0.0000294
2011/2012	*	0.0250	0.0000327
2012/2013	0.7116	0.1251	0.098357
2013/2014	0.2820	0.1518	0.000003
2014/2015	0.0468	0.0454	0.000000000995
2015/2016	0.0113	0.2511	09.4602e-12
2016/2017	0.0012	0.0626	6.8612e-13
2017/2018	0.0016	0.0626	0
2018/2019	0.0030	0.5127	0.000000000194
2019/2020	0.0027	0.3153	0.00000000108

Bearing in mind that since September 2012 there was a regional NI roll out of the 24 hour Primary Percutaneous Coronary Intervention (PPCI) service, we examined HF prevalence in rural and urban areas with different travel time to PPCI. Timely access to coronary revascularization is correlated with better outcomes [8], whereas delays of revascularization in setting of the acute coronary syndrome increases risk of myocardial damage, which in consequence can lead to HF. We cross checked postcodes for each GP practice with the most recent version of the National Statistics Postcode Lookup (NSPL) for the UK from February 2021 and matched it with the data from lookup table of the 890 Super Output Areas (SOA2011) to obtain defaulted urban/rural classification for each practice and travel times to Belfast and Derry/Londonderry, where 2 PPCI centers are located. Geo data were accessed from NI Statistics and Research Agency and Office for National Statistics websites. Urban area has been defined as a settlement with a population greater than or equal to 5,000 people, whereas rural area has a population of less than 5,000 people. There were 97 practices in rural areas and 261 in urban areas. 1 practice was allocated to the mixed rural/urban area. This practice has been excluded from the analysis. We plotted the prevalence of HF in rural and urban GP practices (Fig. 5).

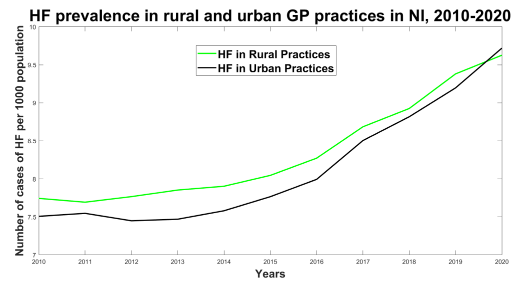

Figure 5. Heart Failure prevalence in GP practices located in rural and urban areas, Northern Ireland, years 2010-2020.

In comparison to urban practices, HF prevalence in rural GP practices was higher than in urban GP practices. We then divided practices from rural and urban areas into 2 subgroups: "close" and "distant" practices. In first subgroup: "close" we included all practices that were within 0–20-minute travel to Belfast or Derry/Londonderry. In the second subgroup: "distant" we included all practices with travel time of ¿60 minutes to Belfast and Derry/Londonderry. We then applied the unpaired Mann-Whitney U-test to 2 subgroups within rural and urban practices. Mann-Whitney U-test examines the null hypothesis that data in x and y are samples from continuous distributions with equal medians, against the alternative that they are not. We hypothesised whether there were statistically significant differences in HF prevalence between "distant" and "close" practices regardless of the default status of the practice as rural or urban. We calculated U statistics and p-value for rural and urban practices. The statistic did not suggest to be significant even though we observed a consistent year on year difference in the HF prevalence between rural and urban areas. This may be the shortcoming of statistical analysis.

4. Discussion

Data obtained from various open sources allowed us to present trends of cardiovascular conditions for the NI population over the last decade. We were able to show that there is a significant increase in HF prevalence in NI from 2015 onward. The QOF team published their own analyses however they did not describe the trends that we have discovered in this dataset. This supports a theory that close collaboration between clinicians and data scientists can lead to knowledge discovery from information included in health care data [9]. We were not able to show, however, factors directly influencing the increase in HF prevalence on this register. We may only speculate that this could be due to the increased awareness of HF care, as new life-saving evidence-based therapies such as sacubitril/valsartan were introduced and as a result of significant increase in HF education dissemination. Advancements of HF device therapy could pleay a role too [10]. There may be another reason for the increase in HF prevalence between 2013 and 2014. This could be a result of the introduction of NI regional 24/7 revascularization service in September 2012. This however could be correlation not causation. AF showed the most rapid increase in prevalence in this time series. This may be a result of intensification of active screening for AF, to which primary care physicians are encouraged to. The detection of AF carries a benefit on multiple levels. Detection of AF reduces the risk of stroke, which is the major cause of disability and increased costs to healthcare, society and economy due to the loss of employment years. We had an opportunity to evaluate the dynamics of HF prevalence in different NI HSC trusts and at the GP practice level. We were able to see the difference in HF prevalence between distant and close GP practices within both rural and urban areas. This may be caused by differences in timely access to specialized cardiac teams for patients of distant GP practices. Interestingly, the QOF administrators claim that the apparent increases in prevalence of certain conditions may be due to improvement in recording and case finding by GPs, rather than a true increase in the prevalence in the population. There are limitations to our analysis due to the quality of the available data. The QOF register represents aggregated population information hence it lacks detailed information on gender and age ranges of included patients. From the clinicians' perspective however, this dataset is not optimal. There is a need for the high quality data gathered at the central level. Prevalence and health outcome data are needed to monitor the efficiency of the healthcare service. Well-run secondary analyses by trained health data scientists can better inform policy makers how to adopt a healthcare system to maximize patient health outcomes. The value of population-based studies has been recognized previously and led to the implementation of standard data collection sets that can measure, analyze and improve outcomes achieved in the delivery of care of challenging groups of health conditions. We advocate for the development of automated data collection tools for the care quality improvement and research purposes.

References

[1] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart disease and stroke statistics—2020 update: a report from the american heart association. Circulation 2020;141(9):e139–e596.

[2] Timmis A, Townsend N, Gale CP, Torbica A, Lettino M, Petersen SE, Mossialos EA, Maggioni AP, Kazakiewicz D, May HT, et al. European society of cardiology: cardiovascular disease statistics 2019. European heart journal 2020; 41(1):12–85.

[3] Wilkinson C, Weston C, Timmis A, Quinn T, Keys A, Gale CP. The myocardial ischaemia national audit project (minap). European Heart Journal Quality of Care and Clinical Outcomes 2020;6(1):19–22.

[4] Ludman PF. British cardiovascular intervention society registry for audit and quality assessment of percutaneous coronary interventions in the united kingdom. Heart 2011; 97(16):1293–1297.

[5] Gale CP, Weston C, Denaxas S, Cunningham, et al. Engaging with the clinical data transparency initiative: a view from the national institute for cardiovascular outcomes research (nicor), 2012.

[6] Scotland PH. Scottish heart disease statistics-year ending 31^{st} march 2020, 2021. URL `https://publichealthscotland.scot/publications/scottish-heart-disease-statistics/scottish-heart-disease-statistics-year-ending-31-march-2020/`.

[7] of Health D. Raw disease prevalence data, 2020. URL `https://www.health-ni.gov.uk/publications/201920-raw-disease-prevalence-trend-data-northern-ireland`.

[8] Collet JP, et al. 2020 esc guidelines for the management of acute coronary syndromes in patients presenting without persistent st-segment elevation: the task force for the management of acute coronary syndromes in patients presenting without persistent st-segment elevation of the european society of cardiology (esc). European heart journal 2021; 42(14):1289–1367.

[9] Ben-Assuli O, Heart T, Shlomo N, Klempfner R. Bringing big data analytics closer to practice: A methodological explanation and demonstration of classification algorithms. Health Policy and Technology 2019;8(1):7–13.

[10] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, et al. 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure. Kardiologia Polska Polish Heart Journal 2016;74(10):1037–1147.

Address for correspondence:

Alicja Jasinska-Piadlo
Newtownabbey, BT37 0QB, Northern Ireland, UK.
jasinska_piadlo-a@ulster.ac.uk