Internal and Emergency Medicine
https://doi.org/10.1007/s11739-023-03297-6

IM - ORIGINAL

Cardiovascular disease (CVD) outcomes and associated risk factors in a medicare population without prior CVD history: an analysis using statistical and machine learning algorithms

Gregory Yoke Hong Lip[1,2] · Ash Genaidy[3,4] · Cara Estes[3]

Received: 7 February 2023 / Accepted: 26 April 2023

Abstract

There is limited information on predicting incident cardiovascular outcomes among high- to very high-risk populations such as the elderly (≥ 65 years) in the absence of prior cardiovascular disease and the presence of non-cardiovascular multi-morbidity. We hypothesized that statistical/machine learning modeling can improve risk prediction, thus helping inform care management strategies. We defined a population from the Medicare health plan, a US government-funded program mostly for the elderly and varied levels of non-cardiovascular multi-morbidity. Participants were screened for cardiovascular disease (CVD), coronary or peripheral artery disease (CAD or PAD), heart failure (HF), atrial fibrillation (AF), ischemic stroke (IS), transient ischemic attack (TIA), and myocardial infarction (MI) for a 3-yr period in the comorbid history. They were followed up for up to 45.2 months. Analyses included descriptive approaches in terms of incidence rates and density ratios, and inferential in terms of main effect statistical/complex machine learning modeling. The contemporary risk factors of interest spanned across the domains of comorbidity, lifestyle, and healthcare utilization history. The cohort consisted of 154,551 individuals (mean age 68.8 years; 62.2% female). The overall crude incidence rate of CVD events was 9.9 new cases per 100 person-years. The highest rates among its component outcomes were obtained for CAD or PAD (3.6 for each), followed by HF (2.2) and AF (1.8), then IS (1.3), and finally TIA (1.0) and MI (0.9).

Model performance was modest in terms of discriminatory power (C index: 0.67, 95%CI 0.667–0.674 for training; and 0.668, 95%CI 0.663–0.673 for validation data), equal agreement between predicted and observed events for calibration purposes, and good clinical utility in terms of a net benefit of 15 true positives per 100 patients relative to the All-patient treatment strategy. Complex models based on machine learning algorithms yielded incrementally better discriminatory power and much improved goodness-of-fitness tests from those based on main effect statistical modeling. This Medicare population represents a highly vulnerable group for incident CVD events. This population would benefit from an integrated approach to their care and management, including attention to their comorbidities and lifestyle factors, as well as medication adherence.

Keywords Cardiovascular disease events · Statistical and machine learning modeling · Integrated care management

✉ Gregory Yoke Hong Lip
gregory.lip@liverpool.ac.uk

✉ Ash Genaidy
ashgenaidy@gmail.com

[1] Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart and Chest Hospital, Liverpool L7 8TX, UK

[2] Danish Center for Clinical Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[3] Anthem Inc, Indianapolis, IN, USA

[4] Anthem Clinical Health Economics Team, Cincinnati, OH 45249, USA

Introduction

Cardiovascular diseases (CVD) remain the leading cause of disease burden in the world and in non-high income countries rates of morbidity and mortality continue to increase [1]. Consequently, there remains a global need to focus on implementing cost-effective strategies via integrated care policies and interventions to improve CVD diagnosis, risk assessment and aid management decision-making.

Current algorithms used to predict CVD risk are usually designed for the general populations who are free of prior CVD risk and with median age mostly < 50 years [2–8]. These algorithms also typically target future CVD risk of 5 to 10 yr.

Published online: 09 June 2023

A major drawback of these algorithms is that they tend to overestimate the CVD risk, and age limits were mostly around 75 years and were represented in relatively smaller proportions compared to the general population [9]. Importantly, limited information is available when applied to the 75–84 age limit and non-existent for the 85–94 age limit. Another major drawback is the lack of data on non-cardiovascular multimorbidity (e.g., chronic kidney disease, chronic obstructive pulmonary disease/bronchiectasis), and healthcare utilization such as emergency room visits. The latter is usually regarded as a proxy for healthcare use and a key determinant of disease burden. Finally, CVD risk equations do not include CVD outcomes such as incident atrial fibrillation which is associated with a high risk of mortality and morbidity, particularly from stroke and heart failure [10].

With the above in mind, there are limited insights into the evolution of CVD and cardiovascular risk prediction in special populations such as the elderly without prior CVD history in the presence of non-cardiovascular multi-morbidity. Recently, Neuman et al. [9] studied an elderly population without prior CVD events and dementia/physical disability but with limited information on any prior non-cardiovascular co-/multi-morbidity. They reported 594 major adverse cardiovascular events in a median follow-up of 4.7 years involving 18,548 participants aged 70 years and above. The risk factors for CVD events included age, gender, smoking, systolic blood pressure, high-density lipoprotein cholesterol (HDL-c), non-HDL-c, serum creatinine, diabetes mellitus, and intake of anti-hypertensive agents [9]. These results are limited by the smaller sample for the elderly populations and model performance validation metrics, and cannot be extrapolated to Medicare cohorts in the US which are mostly elderly and typically have higher prevalence rates of non-cardiovascular multi-morbidity relative to the general population [11–15]. Therefore, the development of contemporary risk tools would benefit an integrated approach customized to their care management, including attention to their current comorbidities and lifestyle factors.

The specific aims of this study are: (i) to examine the incidence of CVD and its components i.e., coronary or peripheral artery disease (CAD or PAD), heart failure (HF), atrial fibrillation (AF), ischemic stroke (IS), transient ischemic attack (TIA), and myocardial infarction (MI)) in a multi-morbid Medicare population without prior CVD history; and (ii) to develop and validate an algorithm to predict incident CVD outcomes/components in Medicare populations using statistical and machine learning algorithms.

Methods

Population description and criteria for participation

The participant patients are derived from the Medicare health plan which is funded by the US government for individuals aged 65 years and above, as well as including those with some level of disability for those age > 18 years. In essence, Medicare provides medical insurance to about 60 million of the US population who are 65 years and above as well as younger populations under age 65 receiving Social Security Disability Insurance and those diagnosed with end-stage renal disease and amyotrophic lateral sclerosis. Details of the Medicare health plan characteristics are provided elsewhere [16].

Plan participants have both medical and pharmacy benefits, and their information was extracted from administrative databases between January 1, 2016 and September 30, 2022, with the criterion of having at least 48 months of continuous enrollment or more for each individual participant.

The plan participants did not have any cardiovascular conditions in the comorbid history for the first three years of the study. For the fourth year, at least one year of follow-up or more was dedicated to detect any incident cardiovascular outcomes. No cardiovascular conditions were allowed in the comorbid history period of three years including coronary or peripheral artery disease (CAD or PAD), heart failure (HF), atrial fibrillation (AF), ischemic stroke (IS), transient ischemic attack (TIA), and myocardial infarction (MI). The outcomes tracked during the follow-up time consisted of incident cardiovascular disease (CVD), i.e., composite outcome including any of the individual seven cardiovascular outcomes and its component outcomes.

Risk factors and cardiovascular outcomes

The risk factors included medical conditions in the comorbid history, lifestyle/personal factors, healthcare utilization variables, and demographic attributes. Description of these factors and their codes in the administrative databases are provided in Suppl Tables S1 and S2. This comprehensive array of risk factors was based on prior published literature [11–15].

The above risk factors were treated as binary variables with "1" for the condition presence and "0" for its absence. The healthcare utilization factors were nominal categorical variables and are described in Suppl Table S2. Gender was treated as a binary variable with "1" for females and "0" for males. Age was analyzed as a continuous

variable in years and categorical in groups (i.e., nominal variable). The age brackets included 18–44 years or "0", 45–54 years or "1", 55–64 years or "2"; 65–74 years or "3"; 75–84 years or "4"; and 85–94 years or "5".

Analysis of variables

The analysis of variables included descriptive and inferential statistics as well as model prediction using statistical and machine learning (ML) algorithms. The statistical analyses were performed using the Statistical Analysis Software (SAS) Enterprise, with the ML computations conducted using the SAS Enterprise Miner. The descriptive analyses included calculation of individual counts (%) for demographic parameters and co-morbid history, with the exception of mean (SD) for age as a continuous variable. The clinical outcomes were analyzed in terms of incidence rates in new cases/100 person-years by age groups, gender, and overall population.

Prediction modeling was performed using statistical (i.e., main effect modeling via logistic regression analysis) to examine the independent effects of co-morbid, lifestyle/personal history, healthcare utilization, and demographic variables. The incident cardiovascular outcomes were binary and 8 in total (one for composite outcome or CVD and 7 for its component outcomes, that is, CAD, PAD, HF, AF, IS, TIA, and MI). The analyses were performed for all 8 outcomes.

The ML techniques were pursued for complex relationships using two parametric (i.e., neural network and logistic regression) methods. The ML-based logistic regression algorithm included main effects, interaction terms and polynomial effects, with the model selection based on the stepwise method. Only quadratic terms were included in the polynomial formulation to ensure proper conversion in a timely fashion of the optimization algorithm from numerical analysis perspective. Neural network used a multilayer perceptron architecture with direct connection for a feed-forward multilayer network architecture composed of several layers of neurons (i.e., input, output, and hidden layers). Five hidden layers were deemed appropriate to handle the model complexity in this study.

Model validation was based on calibration, discrimination and clinical utility. Each model was trained on 67% of the data, with the remaining 33% data used for external validation. The training and validation samples were extracted at random. Discriminant validity was assessed using C-indices for both training and validation samples, separately. Clinical utility of each model was evaluated using decision curve analysis, with the net benefit calculated for the prediction model at hand in comparison to default strategies of treating all or no patients [17, 18].

In decision curve analysis, the net benefits of the model predicting patients at risk are compared against treating all patients or not treating any patient. Furthermore, the comparison between the prediction model and all patients is made for a given probability threshold. In general, net benefit is calculated across a range of threshold probabilities, defined as the minimum probability of disease at which further intervention would be warranted, as net benefit = true positive rate – (false positive rate x weighting factor) where the weighting factor = threshold probability / (1—threshold probability). As such, it is a measure of true positive events after accounting for false positives. Indeed, the risk model use would then provide a more clinically effective care strategy via reduction of potential harm or false positive if the induced net benefit is higher than one produced by treating all patients. Model calibration was assessed as described elsewhere for assessing the degree of agreement between the predicted probabilities and actual values [19].

Results

The Medicare cohort consisted of 154,551 individuals (mean age 68.8 years; 62.2% female) (Table 1). About 81% of the population was aged ≥ 65 years (with 70.6% of those in the 65–74 year age bracket), with 9.3% and 9.8% for the 18–54 and 45–54 age groups, respectively. The highest prevalence rates in the comorbid history (i.e., ≥ 25%) were observed for hypertension, diabetes mellitus, hyperlipidemia, spondylosis, and osteoarthritis (Table 1). In terms of lifestyle factors, obesity and tobacco use/dependency had the highest prevalence rates (21% and 10%, respectively). In terms of healthcare utilization, the rates of ER visits and hospitalizations in the last 6 months of the 3-yr period of the comorbid history were 9.8% and 2.8%, respectively. The rate of spending 1 day or longer in the last 30 days in the 3-yr period of the comorbid history prior to the incidence of CVD events was 0.51%.

Incidence of CVD events and its component outcomes, and their survival times

The overall crude incidence rate of CVD events was 9.9 new cases per 100 person-years (Suppl. Table S3). The highest rates among its component outcomes were CAD or PAD (3.6), followed by HF (2.2) and atrial fibrillation (1.8), then ischemic stroke (1.3), and finally transient ischemic attack (1.0) and myocardial infarction (0.9). The incidence density ratio followed a similar path, that is, CVD (29%) and its component outcomes (CAD, 11.9%; PAD, 12.2%; HF, 7.6%; AF, 6.2%; IS, 4.4%; TIA, 3.4%; MI, 3.0%).

Incidence rate or density ratio increased steadily with an increase in age; however, there was a plateau between the 55–64 and 65–74 age groups (Suppl Table S3). Males had slightly higher incidence rates and density ratios for

Table 1 Frequency and prevalence of baseline characteristics for participant population

Baseline characteristic	Level	Frequency/Prevalence
Age group, *n* (%)	18–44	5792 (3.8)
	45–54	8491 (5.5)
	55–64	15,097 (9.8)
	65–74	88,338 (57.2)
	75–84	31,480 (20.4)
	85–94	5242 (3.4)
Age in years (mean, SD)	68.8 (10.3)	68.8 (10.3)
Gender, *n* (%)	Males	58,450 (37.8)
	Females	95,990 (62.2)
Overall, *n* (%)		154,551 (100.0)
Comorbid condition history, *n* (%)	Hypertension	97,087 (62.8)
	Diabetes mellitus	38,697 (25.0)
	Hyperlipidemia	96,631 (62.5)
	Low high-density lipoprotein (HDL)	512 (0.3)
	Hyperthyroidism	3292 (2.1)
	Chronic obstructive pulmonary disease/bronchictasis	16,051 (10.4)
	Asthma	12,586 (8.1)
	Sleep apnea	9923 (6.4)
	Chronic kidney disease	12,795 (8.3)
	Liver disease	13,492 (8.7)
	Anemia	24,178 (15.6)
	Spondylosis/intervertebral disks	56,093 (36.3)
	Osteoarthritis	48,042 (31.1)
	Depression	20,208 (13.1)
	Major bleeding	10,753 (7.0)
	Cognitive impairment	3566 (2.3)
Lifestyle/personal history, *n* (%)	Obesity	31,808 (20.6)
	Tobacco use and dependency	15,233 (9.9)
	Alcohol use and dependency	2690 (1.7)
	Inappropriate diet	3743 (2.4)
	Inadequate physical exercise	400 (0.3)
	Life stresses	84 (0.1)
	Family history of diseases	9705 (6.3)
Healthcare utilization history, *n* (%)	*ER usage in last 6 months, count/nominal scale*	
	0	139,354 (90.2)
	1	11,851 (7.7)
	2	2209 (1.4)
	3	637 (0.4)
	4 or more (4)	500 (0.3)
	Hospital inpatient admissions in last 6 months, count/nominal scale	
	0	150,218 (97.20)
	1	3790 (2.45)
	2	434 (0.28)
	3	71 (0.05)
	4 or more (4)	38 (0.02)
	Length of hospital stay in last 30 days, total days/nominal scale	
	0	153,766 (99.49)
	1	94 (0.06)
	2	157 (0.10)
	3	131 (0.08)
	4–6 (4)	223 (0.14)
	7–13 (5)	133 (0.09)
	14–30 (7)	47 (0.03)

Table 1 (continued)

Values are in count (*n*) and percent except where noted

Note: *ER* emergency room

CVD events and its component outcomes. The survival times for the incident CVD events and component outcomes averaged from 21 to 24 months, with the median values similarly spanning from 21 to 26 months past the 3-yr period of comorbid history (see Suppl Table S4).

Derivation and validation of an algorithm to predict CVD risk using statistical/ML algorithms

Several individual risk variables were statistically significant in predicting incidence cardiovascular disease events and its component outcomes and produced good performance measures (see Table 2 and Figs. 1, 2, 3; Tables S5, S6, S7, S8, S9, S10, S11, S12 and Figs. S1, S2, S3) at the 0.001 level.

Table 2 Odds ratios (*PE* point estimate; *CI* confidence interval; *LL* lower limit; *UL* upper limit) for incident cardiovascular disease using individual risk/protective variables

		Incident cardiovascular event							
		Training data				Validation Data			
		PE	LL	UL	Pr > ChiSq	PE	LL	UL	Pr > ChiSq
Comorbid condition history									
	Hypertension	1.44	1.40	1.49	< .0001	1.41	1.34	1.48	< .0001
	Diabetes mellitus	1.24	1.20	1.28	< .0001	1.21	1.16	1.27	< .0001
	Hyperlipidemia	1.10	1.06	1.13	< .0001	1.13	1.08	1.18	< .0001
	Low high-density lipoprotein (HDL)								
	Hyperthyroidism	1.13	1.03	1.24	0.0137	1.17	1.03	1.33	0.0167
	Chronic obstructive pulmonary disease/bronchiectasis	1.63	1.55	1.70	< .0001	1.58	1.48	1.68	< .0001
	Asthma	1.11	1.05	1.17	< .0001	1.17	1.09	1.26	< .0001
	Sleep apnea	1.27	1.21	1.35	< .0001	1.23	1.14	1.34	< .0001
	Chronic kidney disease	1.21	1.16	1.27	< .0001	1.18	1.10	1.26	< .0001
	Liver disease	1.07	1.02	1.13	0.0053	1.07	1.00	1.15	0.0435
	Anemia	1.21	1.17	1.26	< .0001	1.20	1.14	1.26	< .0001
	Spondylosis/intervertebral discs	1.25	1.21	1.29	< .0001	1.28	1.23	1.34	< .0001
	Osteoarthritis	1.22	1.18	1.26	< .0001	1.25	1.20	1.31	< .0001
	Depression	1.12	1.07	1.17	< .0001	1.13	1.06	1.20	0.0001
	Major bleeding	1.19	1.13	1.26	< .0001	1.20	1.11	1.29	< .0001
	Cognitive impairment	1.43	1.31	1.56	< .0001	1.29	1.14	1.46	< .0001
Lifestyle history									
	Obesity	1.17	1.13	1.21	< .0001	1.15	1.09	1.21	< .0001
	Tobacco use and dependency	1.35	1.28	1.42	< .0001	1.38	1.28	1.48	< .0001
	Alcohol use and dependency								
	Inappropriate diet								
	Inadequate physical exercise								
	Life stresses								
	Family history of diseases	1.30	1.23	1.37	< .0001	1.29	1.19	1.39	< .0001
Healthcare utilization history									
	ER usage in last 6 months	1.12	1.09	1.16	< .0001	1.20	1.15	1.25	< .0001
	Hospital inpatient admissions in last 6 months								
	Length of hospital stay in last 30 days	1.09	1.03	1.14	0.0015	1.05	0.98	1.12	0.184
Demographics	Gender	0.81	0.78	0.83	< .0001	0.79	0.75	0.82	< .0001
	Age (years)	1.04	1.04	1.05	< .0001	1.04	1.04	1.05	< .0001
C index		0.669				0.669			

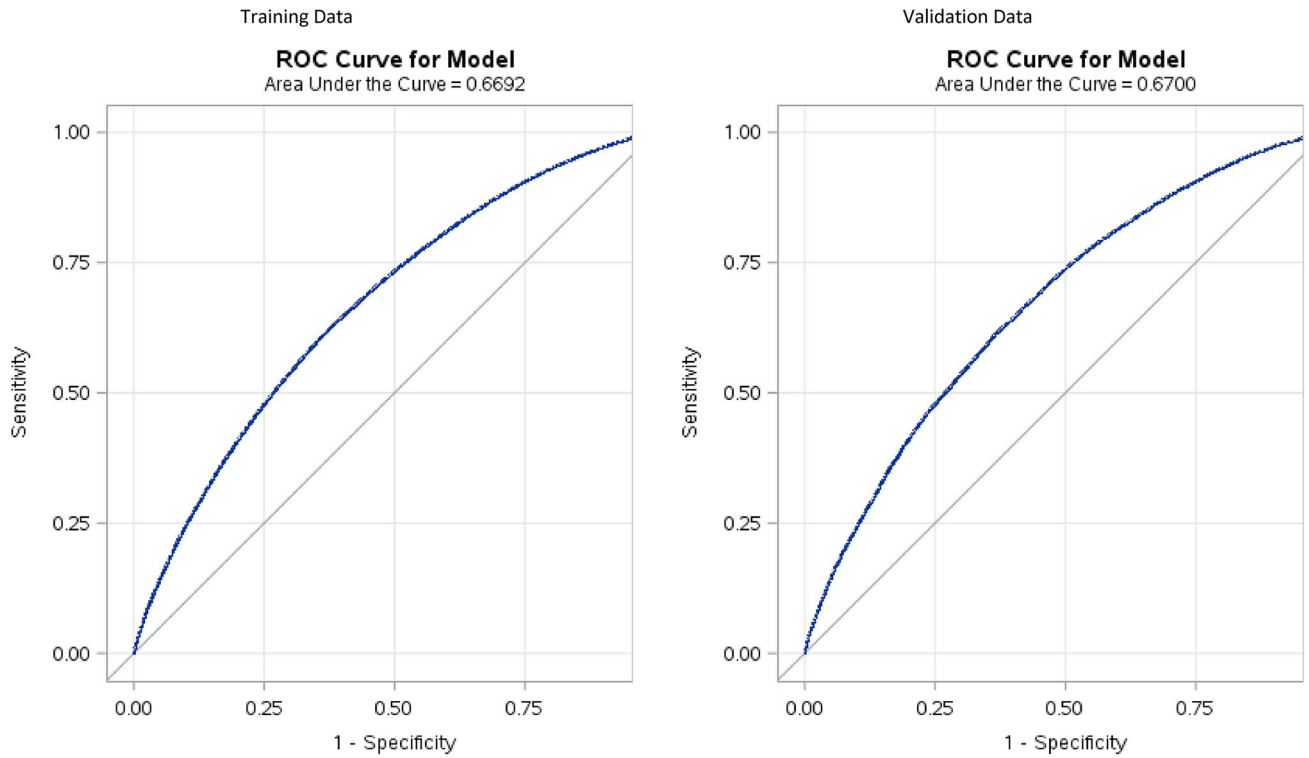

Fig. 1 Discriminant validity performance of incident cardiovascular disease outcome model with individual risk/protective factors for training and validation data

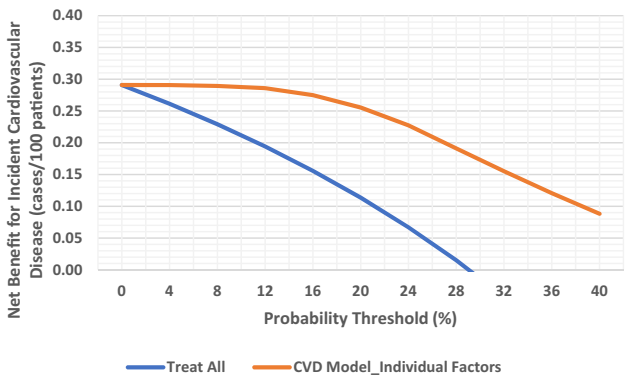

Fig. 2 Decision curve analysis in terms of net benefit (new cases/100 patients) for incident composite cardiovascular disease outcome model with individual risk/protective factors

The validation data showed comparable performance measures to the training data for the various incident cardiovascular outcomes. The models for the incident composite CVD outcomes showed good performance values for both training and validation data, with similar c indexes for the training and validation data, as follows: 0.67 (95%CI 0.667–0.674) and 0.668 (95%CI 0.663–0.673), respectively with the ROC curves shown in Fig. 1.

The performance of main effect modeling in terms of clinical utility is shown in Fig. 2 for the validation data of incident CVD events using decision curve analysis. The predicted outcomes produced much higher true positives (after accounting for false positives) than those obtained on the basis of the strategy of treating all patients for the entire spectrum of probability thresholds. For example, if one is to adopt a 20% probability threshold to predict an incident CVD event, the model can detect 25.4 true positives per 100 patients in comparison to treating all patients which can target only about 11.3% of the patients. The prediction model provides a clear benefit over and above treating all patients by 14.1 cardiovascular cases per 100 patients after adjusting for any false positives.

The calibration curves in Fig. 2 showed good accuracy agreements between the observed and predicted events for the widest range of probability levels (0–60%), with slight overestimation in the higher probability thresholds. Calibration of the training and validation models for various outcomes showed comparable predicted and observed events in the allowable range (Suppl Fig. S2).

Among the individual factors, a number of influential conditions were associated with the onset of incident composite CVD events (see Table 2). The most common included hypertension, diabetes mellitus, chronic kidney disease, sleep

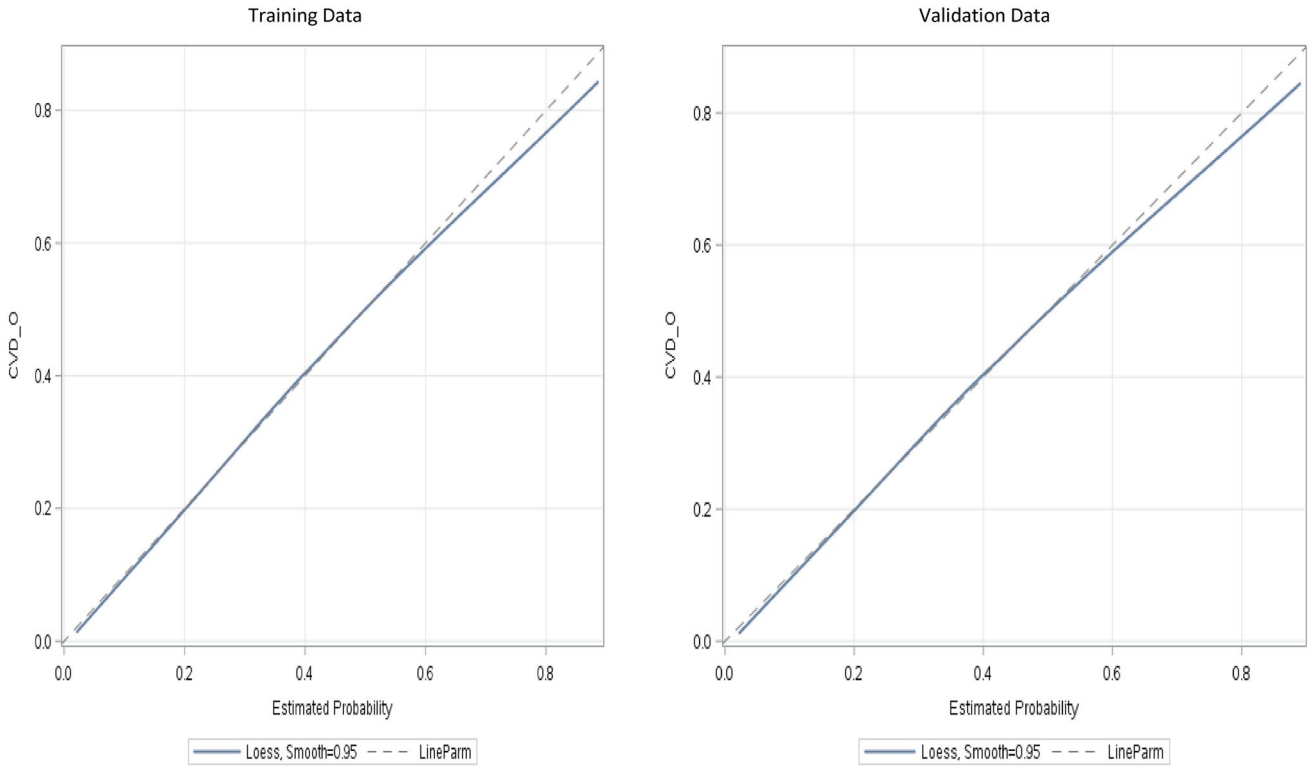

Fig. 3 Calibration performance of incident cardiovascular disease outcome model with individual risk/protective factors for training and validation data

apnea, spondylosis, osteoarthritis, COPD, anemia, major bleeding, and cognitive impairment. Obesity, tobacco use and dependency, and family history were significant predictors in the lifestyle/personal history index. ER utilization in the last 6 months of the 3-year comorbid history period as well as days of hospitalizations in the last 30 days of the 3-yr comorbid history were also associated with CVD events. Females were at a lesser risk than males. Inching up one year of age at a time increased the risk of cardiovascular onset by 4%.

Building complex models using ML-based formulations yielded slightly higher discriminatory power (c index values for neural network: 0.68 for training data and 0.68 for validation data; c index values for logistic regression: 0.674 for training data and 0.68 for validation data) than the main effect modeling based on logistic regression (see Suppl Figs. S4, S5, S6). The goodness-of-fit tests for the ML formulations were much better than those for the main effect models therefore reducing the misspecification errors. The two parametric-based ML algorithms produced very similar results.

Discussion

The main findings from this study are as follows: (1) the overall crude incidence rate for CVD events was high amounting to 9.9 new cases/100 person-years and an incidence density equaling 29%; (2) the highest incidence rates and density ratios were obtained for CAD and PAD; (3) the model developed to predict incident CVD events demonstrated good performance in terms of discrimination, calibration, and clinical utility; (4) a number of important contemporary individual risk factors emerged from the comorbid (e.g., COPD), lifestyle/personal (e.g., tobacco use/dependency) and healthcare utilization (e.g., ER visit counts) history; and (5) the more complex ML-based algorithms yielded slightly better discriminatory performance than those based on traditional main effect statistical modeling, yet it produced a better model with a goodness-of-fit test thus reducing misspecification errors.

The crude overall incidence density ratio of 29% was much higher than those recently reported in 2019 for the general population in the US [20]. For the adult population aged 18 years and above, the prevalence rate of self-reported heart disease (i.e., coronary heart disease, angina or angina pectoris, and myocardial infarction) was 6.4% [18]. For an age group equal to 65 years and above, the prevalence rate was 18.3%. Based on an American Heart Association report, [21] the prevalence of CVD (defined as comprised of CAD, HF and stroke) was equal to 9.3% overall in adults ≥ 20 years of age. The 29% incidence ratio was much higher than the prevalence rate of 7.9% for CVD for the general population we previously reported [11] and was slightly lower than that of 35.1% for the age ≥ 65-year group.

In this study, the crude overall incidence rate was 9.9 cases/100 person-years, a value that is higher than the overall rates of 0.66 and 0.95 cases/100 person-years previously reported for the UK women and men, respectively [2]. Furthermore, the 85–94 year age group studied in the present study had an incidence rate that is almost double the overall rate (18.3); indeed, almost 1 in 2 end up having an incident CVD event based on the incidence density ratio of 48%. This age bracket has limited prior data and demonstrates that it is extremely vulnerable for incident CVD events.

The overall incidence rate obtained for MI was 0.9 events/100 person-years, which was higher than our prior report for the general population across three health plans (i.e., Medicare, Medicaid, Commercial) and a comparable age spectrum [11]. Hence, the MI incidence rate for this Medicare cohort without CVD in the comorbid history is almost 38% to 104% of the incidence rate for the general population. The overall incidence rate of AF was 1.8 new events/100 person-years, much higher than that reported by Lip et al. [12] for a Medicaid cohort typified by a lower socio-economic status (relative to those participating in this study), high disability status, presence of cardiovascular conditions in comorbid history, and a much younger average age. In our prior study [12], the incidence rate was 0.49 cases/100 person-years. Furthermore, the overall rate in the present study was much higher than that reported by Lip et al. [13] for the general population (incidence rate = 0.33 cases/100 person-years) with the presence of cardiovascular conditions in the comorbid history.

Lip et al. [14] reported that the incidence rate for stroke (i.e., ischemic stroke, transient ischemic attack, and thrombo-embolic events) is 0.95 cases/100 person-years in a general population across three health plans (Commercial, Medicare, and Medicaid). The population had 3.4 million participants and had cardiovascular conditions in the comorbid history including stroke. Yet, the value obtained in the present study for ischemic stroke was higher and equal to 1.3 new events/100 person-years, although the incidence rate for ischemic stroke alone is lower than that for stroke at large. Indeed, adding TIA events to those for ischemic stroke would yield even higher incidence rates [15].

The incidence statistics in the present study suggest that the cohort may represent a high to very high CVD risk population. Therefore, because of the lack of information in the published literature, a simple model predictive of incident CVD events and consisting of individual risk factors was devised, demonstrating good performance metrics and good model calibration. The clinical utility of the model also showed good results by truly detecting 15 true positive events (after accounting for the false positives) per 100 patients over and above those produced by the all-patient treatment strategy at a probability threshold of 20%.

Collectively, this predominantly elderly population with a varied mix of disability and non-cardiovascular multi-morbidity and the absence of cardiovascular disease in comorbid history therefore represents a high- to very high-risk group for CVD events. This suggests that significant non-cardiovascular illnesses put excessive overload on the cardiovascular system and weaken the general immune system. Indeed, these effects would make these cohorts vulnerable for any kind of cardiovascular illness as well as the easier path for the effects of external factors on the cardiovascular system due to reduced immunity (e.g., COVID-19).

In light of the above, it appears that these subpopulations would benefit from an integrated approach to their care management, including attention to their comorbidities and lifestyle factors. From a practicing clinician standpoint, such an integrated care management approach can be devised to manage comorbid history symptoms or disorders [22–24]. This would be in a way similar to (a) the ABC (Atrial fibrillation Better Care) strategy to manage AF [25], which has been associated with improved outcomes in AF patients [26] and has been recommended in guidelines [27] and (b) other integrated approaches reported by researchers and clinicians in the medical literature for chronic long term conditions [24, 28]. The lifestyle/personal variables can be modified to reduce the future CVD risk in a comprehensive manner consistent with the guidelines [29]. For example, the American Heart Association introduced "Life's Simple 7 initiative" including three cardiovascular risk factors (glucose, blood pressure, and cholesterol) and four lifestyle behaviors (body mass index, smoking, physical activity, and diet), and the majority of these factors have been associated with longevity in prospective observational studies [30–33]. In addition to an integrated care management approach, it will be essential to find ways to improve medication adherence given the likely polypharmacy in such patients [34].

With respect to the statistical methods and ML algorithms utilized in this study, it appears that ML algorithms provide better solutions than the traditional statistical techniques, given that the ML formulations provide complex non-linear equations, thereby, exploiting the detailed interactions and

non-linear effects within and across the classes of clinical and non-clinical parameters utilized in the built-in models. Parametric ML techniques were utilized with the aim to provide detailed equations for use by clinicians, and neural network algorithms provided comparable results to the complex logistic regressions equations. Therefore, the latter solution was utilized due to their explicit mathematical formulations.

Indeed, there is a potential economic value of integrating ML in usual care, for detecting patients at higher risk for an integrated, personalized approach. This is particularly vital for the patients at risk of serious CVD (e.g., ischemic heart disease, stroke, and heart failure) with important co-morbidities. For example, Szymanski et al. [35] examined the use of an AF risk prediction algorithm in improving AF detection compared with regular screening in primary care and assessed the associated budget impact, potentially saving millions in the UK healthcare system. Other examples are also reported for other cardiovascular conditions [36–39].

Limitations

The findings of this study were based on observational research derived from administrative databases with potential subject and methodological biases compared to the well-controlled clinical trials. The use of observational studies using administrative data may be subject to confounding bias by unadjusted factors (e.g., disease severity, blood pressure control, exact estimated glomerular filtration rate, adverse drug effect, and reasons for ceasing medication) or by a residual channeling bias. Finally, residual bias is still possible, especially with regard to unmeasured variables related to disease severity and clinical data.

Despite the above-mentioned limitations, the methodological procedures deployed in this investigation are based on best available practices. Additionally, the potential biases may have been lessened by the truly diversified population utilized in this study with large numbers. Despite the biases to which observational studies are subject, these studies complement clinical trial via generalization of results through the use of real world data.

Conclusions

This Medicare population represents a high- to very high-risk group for CVD events and would benefit from an integrated care approach to their management, including attention to their comorbidities, lifestyle factors, and healthcare utilization patterns.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11739-023-03297-6.

Author contributions G.Y.H.L. was involved in conception and design of the study and critical revision of the manuscript. A.G. was involved in the conception and design of the study, data acquisition, statistical analysis and interpretation of data, drafting and critical revision of the manuscript. C.E. was involved in the interpretation of data and critical revision of the manuscript.

Data availability Data are available as presented in the paper. According to US laws and corporate agreements, our own approvals to use the Anthem and Ingenio-Rx data sources for the current study do not allow us to distribute or make patient data directly available to other parties.

Declarations

Conflict of interest The authors report no conflicts of interest in this work.

Human and animal rights statement and Informed consent As this work used fully anonymised administrative datasets, no written consent was obtained.

References

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, Bonny A, Brauer M, Brodmann M, Cahill TJ, Carapetis J, Catapano AL, Chugh SS, Cooper LT, Coresh J, Criqui M, DeCleene N, Eagle KA, Emmons-Bell S, Feigin VL, Fernández-Solà J, Fowkes G, Gakidou E, Grundy SM, He FJ, Howard G, Hu F, Inker L, Karthikeyan G, Kassebaum N, Koroshetz W, Lavie C, Lloyd-Jones D, Lu HS, Mirijello A, Temesgen AM, Mokdad A, Moran AE, Muntner P, Narula J, Neal B, Ntsekhe M, Moraes de Oliveira G, Otto C, Owolabi M, Pratt M, Rajagopalan S, Reitsma M, Ribeiro ALP, Rigotti N, Rodgers A, Sable C, Shakil S, Sliwa-Hahnle K, Stark B, Sundström J, Timpel P, Tleyjeh IM, Valgimigli M, Vos T, Whelton PK, Yacoub M, Zuhlke L, Murray C, Fuster V, GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group (2020) Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. J Am Coll Cardiol 76(25):2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010
2. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P (2008) Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 94(1):34–39. https://doi.org/10.1136/hrt.2007.134890
3. Hippisley-Cox J, Coupland C, Brindle P (2017) Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. https://doi.org/10.1136/bmj.j2099
4. Anderson KM, Odell PM, Wilson PWF, Kannel WB (1991) Cardiovascular disease risk profiles. Am Heart J 121:293–298
5. D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB (2008) General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation 117(6):743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579
6. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, SCORE project group (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 24:987–1003

7. SCORE2 working group and ESC Cardiovascular risk collaboration (2021) SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 42(25):2439–2454. https://doi.org/10.1093/eurheartj/ehab309
8. Pylypchuk R, Wells S, Kerr A, Poppe K, Riddell T, Harwood M, Exeter D, Mehta S, Grey C, Wu BP, Metcalf P, Warren J, Harrison J, Marshall R, Jackson R (2018) Cardiovascular disease risk prediction equations in 400000 primary care patients in New Zealand: a derivation and validation study. Lancet 391:1897–1907
9. Neumann JT, Thao LTP, Callander E, Chowdhury E, Williamson JD, Nelson MR, Donnan G, Woods RL, Reid CM, Poppe KK, Jackson R, Tonkin AM, McNeil JJ (2022) Cardiovascular risk prediction in healthy older people. Geroscience 44(1):403–413. https://doi.org/10.1007/s11357-021-00486-z
10. Burdett P, Lip GYH (2022) Atrial fibrillation in the UK: predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. Eur Heart J Qual Care Clin Outcomes 8(2):187–194. https://doi.org/10.1093/ehjqcco/qcaa093
11. Lip G, Genaidy A, Tran G, Marroquin P, Estes C, Shnaiden T, Bayewitz A (2022) Incident and recurrent myocardial infarction (MI) in relation to comorbidities: prediction of outcomes using machine-learning algorithms. Eur J Clin Invest 52(8):e13777. https://doi.org/10.1111/eci.13777
12. Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C (2022) Incidence and complications of atrial fibrillation in a low Socioeconomic and high disability United States (US) population: a combined statistical and machine learning approach. Int J Clin Pract 2022:8649050. https://doi.org/10.1155/2022/8649050
13. Lip GYH, Tran G, Genaidy A, Marroquin M, Estes C, Harrell T (2021) Prevalence/incidence of atrial fibrillation based on integrated medical/pharmacy claims, and association with co-morbidity profiles/multi-morbidity in a large US adult cohort. Int J Clin Pract 75(5):e14042. https://doi.org/10.1111/ijcp.14042
14. Lip G, Genaidy A, Tran G, Marroquin P, Estes C, Sloop S (2022) Improving stroke risk prediction in the general population: a comparative assessment of common clinical rules, a new multimorbid index, and machine-learning-based algorithms. Thromb Haemost 122(1):142–150. https://doi.org/10.1055/a-1467-2993
15. Lip GYH, Genaidy A, Estes C, McKay D, Falks T (2022) Transient ischemic attack events and incident cardiovascular and non-cardiovascular complications: observations from a large diversified multimorbid cohort. European J Stroke; in press
16. Henry J. Kaiser foundation. An overview of Medicare. https://files.kff.org/attachment/issue-brief-an-overview-of-medicare. Consulted on 04/18/2023
17. Fitzgerald M, Saville BR, Lewis RJ (2015) Decision curve analysis. JAMA 313(4):409–410. https://doi.org/10.1001/jama.2015.37
18. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW (2018) Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol 74(6):796–804. https://doi.org/10.1016/j.eururo.2018.08.038
19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, On behalf of topic group 'evaluating diagnostic tests and prediction models' of the STRATOS initiative (2019) Calibration: the achilles heel of predictive analytics. BMC Med 17:230. https://doi.org/10.1186/s12916-019-1466-7
20. National Centre for Health Statistics. Health, United States, 2020–2021. Respondent-reported prevalence of heart disease in adults aged 18 and over, by selected characteristics: United States, selected years 1997–2019. https:\\www.cdc.gov/nchs/hus/data-finder.htm
21. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng S, Delling FN, Elkind MSV, Evenson KR, Ferguson JF, Gupta DK, Khan SS, Kissela BM, Knutson KL, Lee CD, Lewis TT, Liu J, Loop MS, Lutsey PL, Ma J, Mackey J, Martin SS, Matchar DB, Mussolino ME, Navaneethan SD, Perak AM, Roth GA, Samad Z, Satou GM, Schroeder EB, Shah SH, Shay CM, Stokes A, Van Wagner LB, Wang N-W, Tsao CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee (2021) Heart disease and stroke statistics-2021 update: a report from the American Heart Association. Circulation 143(8):e254–e743. https://doi.org/10.1161/CIR.0000000000000950
22. Buckley BJR, Lip GYH (2022) Current concepts: comprehensive "cardiovascular health" rehabilitation-an integrated approach to improve secondary prevention and rehabilitation of cardiovascular diseases. Thromb Haemost 122(12):1966–1968. https://doi.org/10.1055/s-0042-1757403
23. Lip GYH, Ntaios G (2022) "Novel clinical concepts in thrombosis": integrated care for stroke management-easy as ABC. Thromb Haemost 122(3):316–319. https://doi.org/10.1055/a-1632-1777
24. Field M, Kuduvalli M, Torella F, McKay V, Khalatbari A, Lip GYH (2022) Integrated care systems and the aortovascular hub. Thromb Haemost 122(2):177–180. https://doi.org/10.1055/a-1591-8033
25. Lip GYH (2017) The ABC pathway: an integrated approach to improve AF management. Nat Rev Cardiol 14(11):627–628. https://doi.org/10.1038/nrcardio.2017.153
26. Romiti GF, Pastori D, Rivera-Caravaca JM, Ding WY, Gue YX, Menichelli D, Gumprecht J, Kozieł M, Yang P-S, Guo Y, Lip GYH, Proietti M (2022) Adherence to the "Atrial Fibrillation Better Care" pathway in patients with atrial fibrillation: impact on clinical outcomes-a systematic review and meta-analysis of 285,000 patients. Thromb Haemost 122(3):406–414. https://doi.org/10.1055/a-1515-9630
27. Chao TF, Joung B, Takahashi Y, Lim TW, Choi EK, Chan YH, Guo Y, Sriratanasathavorn C, Oh S, Okumura K, Lip GYH (2022) 2021 focused update consensus guidelines of the Asia Pacific Heart Rhythm Society on Stroke Prevention in Atrial Fibrillation: executive summary. Thromb Haemost 122(1):20–47
28. Lip GYH, Ntaios G (2022) Novel clinical concepts in thrombosis: integrated care for stroke management - easy as ABC. Thromb Haemost 122(3):316–319. https://doi.org/10.1055/a-1632-1777
29. Buckley BJR, Lip GYH (2022) Current concepts: comprehensive "cardiovascular health" rehabilitation - an integrated approach to improve secondary prevention and rehabilitation of cardiovascular diseases. Thromb Haemost 122(12):1966–1968
30. Brenn T (2028) Survival to age 90 in men: the Tromso Study 1974–2018. Int J Environ Res Public Health 2019(16):6
31. Heir T, Erikssen J, Sandvik L (2013) Life style and longevity among initially healthy middle-aged men: prospective cohort study. BMC Public Health 13:8317
32. Wilhelmsen L, Svardsudd K, Eriksson H et al (2011) Factors associated with reaching 90 years of age: a study of men born in 1913 in Gothenburg, Sweden. J Int Med 269:441–51.8
33. Yates LB, Djousse L, Kurth T, Buring JE, Gaziano JM (2008) Exceptional longevity in men: modifiable factors associated with survival and function to age 90 years. Arch Int Med 168:284–290
34. Lip GYH, Genaidy A, Jones B, Tran G, Marroquin P, Estes C, Shnaiden T (2023) Adherence levels and patterns for multiple cardiac medications prescribed to patients with incident atrial fibrillation events. Br J Clin Pharmacol. https://doi.org/10.1111/bcp.15627
35. Szymanski T, Ashton R, Sekelj S, Petrungaro B, Pollock KG, Sandler B, Lister S, Hill NR, Farooqui U (2022) Budget impact analysis of a machine learning algorithm to predict high risk of atrial fibrillation among primary care patients. Europace 24(8):1240–1247. https://doi.org/10.1093/europace/euac016

36. Casebeer A, Horter L, Hayden J, Simmons J, Evers T (2021) Phenotypic clustering of heart failure with preserved ejection fraction reveals different rates of hospitalization. J Cardiovasc Med (Hagerstown) 22(1):45–52. https://doi.org/10.2459/JCM.0000000000001116
37. Yasmin F, Shah SMI, Naeem A, Shujauddin SM, Jabeen A, Kazmi S, Siddiqui SA, Kumar P, Salman S, Hassan SA, Dasari C, Choudhry AS, Mustafa A, Chawla S, Lak HM (2021) Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. Rev Cardiovasc Med 22(4):1095–1113. https://doi.org/10.31083/j.rcm2204121
38. Polo Friz H, Esposito V, Marano G, Primitz L, Bovio A, Delgrossi G, Bombelli M, Grignaffini G, Monza G, Boracchi P (2022) Machine learning and LACE index for predicting 30-day readmissions after heart failure hospitalization in elderly patients. Intern Emerg Med 17(6):1727–1737. https://doi.org/10.1007/s11739-022-02996-w
39. Bernabeu-Wittel M, Para O, Voicehovska J, Gómez-Huelgas R, Václavík J, Battegay E, Holecki M, EFIM Multimorbidity Working Group (2023) van Munster BC (2023) Competences of internal medicine specialists for the management of patients with multimorbidity. EFIM multimorbidity working group position paper. Eur J Intern Med 109:97–106. https://doi.org/10.1016/j.ejim.2023.01.011