Accepted Manuscript

British Journal of General Practice

Predicting unplanned hospitalisations in older adults using routinely recorded general practice data

Klunder, Jet; Heymans, Martijn; van der Heide, Iris; Verheij, Robert; Maarsingh, Otto; van Hout, Hein; Joling, Karlijn

DOI: https://doi.org/10.3399/BJGP.2023.0350

To access the most recent version of this article, please click the DOI URL in the line above.

Received 13 July 2023
Revised 22 February 2024
Accepted 26 February 2024

Predicting unplanned hospitalisations in older adults using routinely recorded general practice data

Development and validation of a prediction model

Jet H Klunder, Martijn W Heymans, Iris van der Heide, Robert A Verheij, Otto R Maarsingh, Hein PJ van Hout, Karlijn J Joling

Jet H Klunder, MD, GP registrar, PhD candidate (corresponding author)
Department of General Practice, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam; Aging and Later Life, Amsterdam Public Health research institute
Email: j.h.klunder@amsterdamumc.nl
ORCID: 0000-0002-0463-6909

Martijn W Heymans, PhD, assistant professor
Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit, Amsterdam; Methodology, Amsterdam Public Health research institute
ORCID: 0000-0002-3889-0921

Iris van der Heide, PhD, senior researcher
Netherlands Institute for Health Services Research (NIVEL), Utrecht, the Netherlands; Department of Languages, Literature and Communication, Faculty of Humanities, Utrecht University, Utrecht
ORCID: 0000-0002-9709-7261

Robert A Verheij, PhD, professor
Netherlands Institute for Health Services Research (NIVEL), Utrecht; Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg
ORCID: 0000-0003-2234-4819

Otto R Maarsingh, MD, PhD, GP, associate professor
Department of General Practice, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam; Aging and Later Life, Amsterdam Public Health research institute
ORCID: 0000-0002-3747-9217

Hein PJ van Hout, PhD, professor
Department of General Practice, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam; Aging and Later Life, Amsterdam Public Health research institute
ORCID: 0000-0002-2495-4808

Karlijn J Joling, PhD, assistant professor
Department of Medicine for Older People, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam; Aging and Later Life, Amsterdam Public Health research institute
ORCID: 0000-0001-5301-6370

Abstract

Background: Unplanned hospitalisations represent a hazardous event for older persons. Timely identification of high-risk individuals using a prediction tool may facilitate preventive interventions.

Aim: To develop and validate an easy-to-use prediction model for unplanned hospitalisations in community-dwelling older adults using readily available data to allow rapid bedside assessment by general practitioners.

Design and Setting: Retrospective study using general practice electronic health records of 243,129 community-dwelling adults aged ≥65 years linked with national administrative data.

Methods: The dataset was geographically split into a development (58.7%) and validation (41.3%) sample to predict unplanned hospitalisations within 6 months. We evaluated the performance of three different models with increasingly smaller selections of candidate predictors (i.e. optimal, readily-available and easy-to-use model, respectively). We used logistic regression with backward selection for model development. The models were validated internally and externally. We assessed predictive performance by area under the curve (AUC) and calibration plots.

Results: In both samples, 7.6% had at least one unplanned hospitalisation within 6 months. The discriminative ability of the three models was comparable and remained stable after geographic validation. The easy-to-use model included age, sex, prior hospitalisations, pulmonary emphysema, heart failure and polypharmacy. Its discriminative ability after validation was AUC 0.72 [95% confidence interval: 0.72-0.71]. Calibration plots showed good calibration.

Conclusion: Our models showed satisfactory predictive ability. Reducing the number of predictors and geographic validation did not impact predictive performance, demonstrating the robustness of the model. We developed an easy-to-use tool that may assist general practitioners in decision-making and targeted preventive interventions.

Keywords: unplanned hospitalisations, older adults, prediction model, general practice, dementia

How this fits in: Unplanned hospital admissions in older adults are a critical concern for both patients, family caregivers, healthcare professionals, and service planners. We developed and validated a robust and easy-to-use prediction model using routinely recorded data from general practices to predict the risk of unplanned hospital admissions in community-dwelling older adults. Identifying older adults at high risk can facilitate targeted preventive interventions, such as case management, telemedicine or anticipatory care planning. Moreover, the model could also be utilised by policymakers for capacity planning of hospital beds.

Introduction

Increasing rates of unplanned hospitalisations in older adults are a major burden on healthcare systems worldwide. For patients, unplanned hospital admissions are associated with functional decline and reduced quality of life(1). People with dementia are at particularly high risk of unplanned admissions, associated with worsening of pre-existing cognitive problems and an increased risk of readmission and death(2-4).

Preventing unplanned admissions is critical to ensure patient safety and well-being and aligns with the World Health Organisation's philosophy of providing tailored care in appropriate settings for older adults(5). A proactive approach optimises the allocation of scarce healthcare resources and addresses a pervasive concern in healthcare systems worldwide, where increasing demand outpaces the capacity of healthcare professionals. In the Netherlands, the Integral Care Agreement (ICA) of 2022 prioritises preventive measures for acute care, particularly for older adults. Through education, prevention and early signalling initiatives, the ICA aims to reduce unplanned hospitalisations(6). Interventions such as providing an anticipatory care plan, telemedicine and integrating a multidisciplinary geriatric team have been shown to reduce the number of unplanned hospitalisations(7-10). However, timely identification of high-risk groups is essential for implementing proactive and targeted interventions.

General practitioners (GPs) are patients' primary point of contact and act as gatekeepers in many healthcare systems, such as the Netherlands(11). Therefore, they play a pivotal role in identifying those at risk for unplanned hospitalisation and targeting preventive interventions. A prediction model that can accurately predict high-risk individuals by reusing patient registration data could help GPs identify these individuals. The use of electronic health record (EHR) data offers opportunities for the development, integration and automated calculation of an individual's risk, because it contains comprehensive patient information and is derived from routine healthcare. The utilisation of these readily available data for the development of a prediction model facilitates ease of use and reduces time burden on GPs. Previous research has shown that administrative data can be useful in accurately predicting unplanned hospitalisations(12, 13). However, the methodological quality of these studies was limited and many models required additional data collection, making clinical use difficult. Models based on routine care data have a lower threshold and might therefore be used more frequently. As a result, their potential impact would be greater, even if the predictive power is similar.

The aim of this study was to develop and validate a practical and easy-to-use prediction model for unplanned hospitalisations using a Dutch representative sample of older persons in general practice. The model was developed using current state-of-the-art methods and incorporating readily available

EHR data complemented with national administrative data. Also, we specifically assessed the predictive performance of the model in a subsample of individuals with cognitive decline or dementia.

Methods

We reported this study according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines(14).

Sources of data

We used pseudonymised EHR data from GPs linked with data from national administrative databases. The baseline data covered the year 2013, outcomes were assessed in 2014.
We used routine EHRs from a nationally representative sample of 417 Dutch general practices participating in NIVEL-Primary Care Database (NIVEL-PCD)(15). This database covers about 10% of the Dutch population and is representative in terms of practice type, urbanisation level and age and gender distribution(15, 16). Data includes information on chronic conditions, medication and GP consultations. GPs receive support in coding and feedback in the quality of recording(16). In the Netherlands, all Dutch inhabitants are registered with a GP and have mandatory health insurance. GP care is fully insured, therefore the threshold for consultation is low. Nine out of ten people aged ≥65 visit their GP at least once a year, with an average of eight consultations per year(17).
Administrative data was provided by Statistics Netherlands, the governmental institution responsible for processing statistical data in the Netherlands. These included demographic information and data on institutionalisation and death. Data on hospitalisations were derived from the Dutch Hospital Data (DHD) database, made available by Statistics Netherlands. In 2013 and 2014, DHD contained data from 87 out of 88 general and academic hospitals in the Netherlands.

Study population

The study population consisted of individuals aged ≥65 years, living at home, and registered uninterrupted in one practice between 1 January 2013 and 31 December 2013 (baseline period). To avoid potential noise from admissions to a long-term care facility and deaths in predicting the outcome, individuals without experiencing an unplanned admission among those who died or were admitted to an LTCF within the prediction period were excluded from the analysis (Supplementary Figure 1). The number of excluded individuals varied depending on the follow-up period (3, 6, and 12 months) (Figure 1).

Outcome

The primary outcome was unplanned hospitalisation with at least one overnight stay within 6 months and derived from national administrative data. Admissions were defined unplanned, when immediate treatment or assistance within 24 hours was necessary according to the medical specialist(18). Admissions without overnight stay and admissions for psychiatric conditions were excluded, since these often require different care trajectories. Secondary outcomes were unplanned admissions within 3 and 12 months.

Predictors

Updating existing prediction models was not feasible, due to the incomparability of the predictors in our dataset with the predictors in existing models as well as the low methodological quality of these studies(12). To (partially) incorporate information from existing models, we selected variables commonly included in existing models as candidate predictors for our model, e.g. prior hospitalisations and several chronic conditions(12). In addition, we selected variables based on the insights from a focus group study we organised among primary healthcare professionals (to be published) and based on the clinical expertise of the authors. Ultimately, we selected 29 candidate predictors including age, sex, migration background, income, living situation, chronic conditions, prescription medications, and healthcare utilisation (see Supplementary Table 1 for a detailed description). Chronic conditions were derived from ICPC-1 coded EHR data (WHO International Classification of Primary Care(19)) recorded up to the end of the baseline period. In NIVEL-PCD, GPs received feedback on the quality of recording and support in coding(16). Chronic conditions were selected because of their high prevalence in older adults(20). Dementia was added because of its strong association with hospitalisation(2). Medication variables were derived from prescription data coded with the Anatomical Therapeutic Chemical (ATC) Classification system and included when prescribed in a chronic fashion (i.e. >2 prescriptions(21)) in the year before baseline. Consult declarations (CTG-codes in Dutch) were derived from coded claims data recorded in general practices in the year before baseline.

Missing data

As the data were derived from routine care processes, any undocumented information in the EHR was not indicated. For the data provided by Statistics Netherlands, income data had missing values

for 116 individuals (<0.01%). This justified conducting a complete case analysis considering the negligible proportion of missing data and the minimal potential impact on our results(12, 13, 22, 23).

Statistical analysis

We assessed linearity for continuous variables using restricted cubic splines(24). Non-linear variables were tested as splines and as categorical variables in the logistic model. If the spline did not improve performance, we chose the categorical variant because we wanted a practical model. Collinearity was evaluated using variance inflation factors (VIF). VIFs ranged between 1.01 and 2.43, therefore problematic collinearity was absent(25).

Model development

Our large sample provided sufficient statistical power to split the sample into a development and validation sample based on geographic region. We used the larger sample, i.e. the six southernmost provinces (58.7%), for development and the smaller sample for validation (Supplementary Figure 2). Geographic validation is considered a stronger approach compared to a random split sample procedure(14, 26).

For model building, we followed the recommended steps outlined in the TRIPOD guidelines (14) and by Steyerberg(27). We performed multivariable logistic regression with backward stepwise selection (P<0.01) using all 29 candidate predictors to design an optimal model (i). Given the sample size, we had sufficient power to fit a more parsimonious model by incrementally removing the variables with weakest association, until the area under the curve (AUC) deteriorated by ≥0.01. Internal validation was performed through bootstrapping (n=250).

This procedure was repeated twice with smaller subsets of candidate predictors to develop a model with only variables readily available from the EHR (readily-available model (ii)) and with only easy-to-use variables (easy-to-use model (iii)), using 24 and 22 candidate predictors, respectively (Supplementary Table 1). The easy-to-use model was designed to allow rapid completion by a GP, we therefore selected variables that are quick and easy to fill. All three models were validated in the northern sample.

Model performance

Discrimination was evaluated through AUC and calibration through calibration plots, intercept and slope. We determined the shrinkage factor to quantify overfitting. We also assessed classification measures, including sensitivity, specificity, positive and negative predictive values, for multiple probability thresholds. The optimal probability threshold was determined using the Youden index(28).

Sensitivity analysis

Sensitivity analyses were undertaken in the optimal model to assess performance for different follow-up periods (i.e. 3 and 12 months). Furthermore, we assessed performance in subsamples with cognitive decline or dementia (ICPC-1 P20 or P70). Additionally, we evaluated model performance in a sample including individuals who had died or been admitted to an LTCF within 6 months. Statistical analyses were performed using SPSS (version 26) and R Studio (version 4.1.2) using packages *rms, pROC,* and *psfmi.*

Results

Participants

Overall, 243,324 individuals were included in the six-month sample (58.7% development and 41.3% validation sample) (Figure 1). Prevalence of candidate predictors and incidence of outcome are shown in Table 1. In both samples, median age was 72 years, 54% was female, and approximately 40% had ≥2 comorbidities, of which osteoarthritis was most prevalent. Thirteen percent experienced at least one hospitalisation the year before baseline. In both samples, 7.6% experienced one or more unplanned hospitalisations within six months.

Model development and validation

The optimal model included eight predictors: sex, age, prior hospitalisations, chronic obstructive pulmonary disease (COPD), polypharmacy, use of blood thinners, number of GP or practice nurse consultations, and percentage of GP home visits (Table 2). When applied to the validation sample, the AUC was 0.73 (95%-CI 0.72-0.73). Youden's optimal probability threshold was 0.07, reflecting a sensitivity of 65.7% and a specificity of 68.5% in the validation sample (Table 3 and Figure 2). We reported performance measures for multiple probability thresholds to accommodate varying clinician preferences for risk estimation (Box 1).

The readily-available model contained all predictors of the optimal model except for prior hospitalisations (Table 2). Compared to the optimal model, the AUC in the validation sample was marginally lower (AUC 0.72 (95%-CI 0.71-0.72).

The easy-to-use model included age, sex, hospitalisations in the past year, heart failure, COPD, and polypharmacy (Table 2). When applied to the validation sample, this resulted in an AUC of 0.72 (95%-CI: 0.71-0.72). To allow for individualised predictions of this model, an Excel spreadsheet is provided as supplement (Supplementary File 2).

For all three models, bootstrapping resulted in an optimism of the AUC, intercept and slope <0.001, therefore no adjustments of the coefficients were required. Calibration of all models was good, the slope and intercept did not deviate to the extent that model updating was undertaken (Supplementary Figure 2).

Sensitivity analyses

Testing the optimal model in persons with cognitive decline resulted in an AUC of 0.67 (95% CI: 0.65 – 0.69) in both samples. The optimal model showed good predictive ability when fitted in the 3- and 12-month follow-up samples. However, the calibration plots showed systematic over- and underestimation in the 3- and 12-month samples, respectively. Finally, evaluating the optimal model in a sample including those who died or were admitted to an LTCF within 6 months resulted in an

AUC of 0.72 (95% CI: 0.72 – 0.73). See Supplementary Tables 2-4 and Supplementary Figure 3 for details of these analyses.

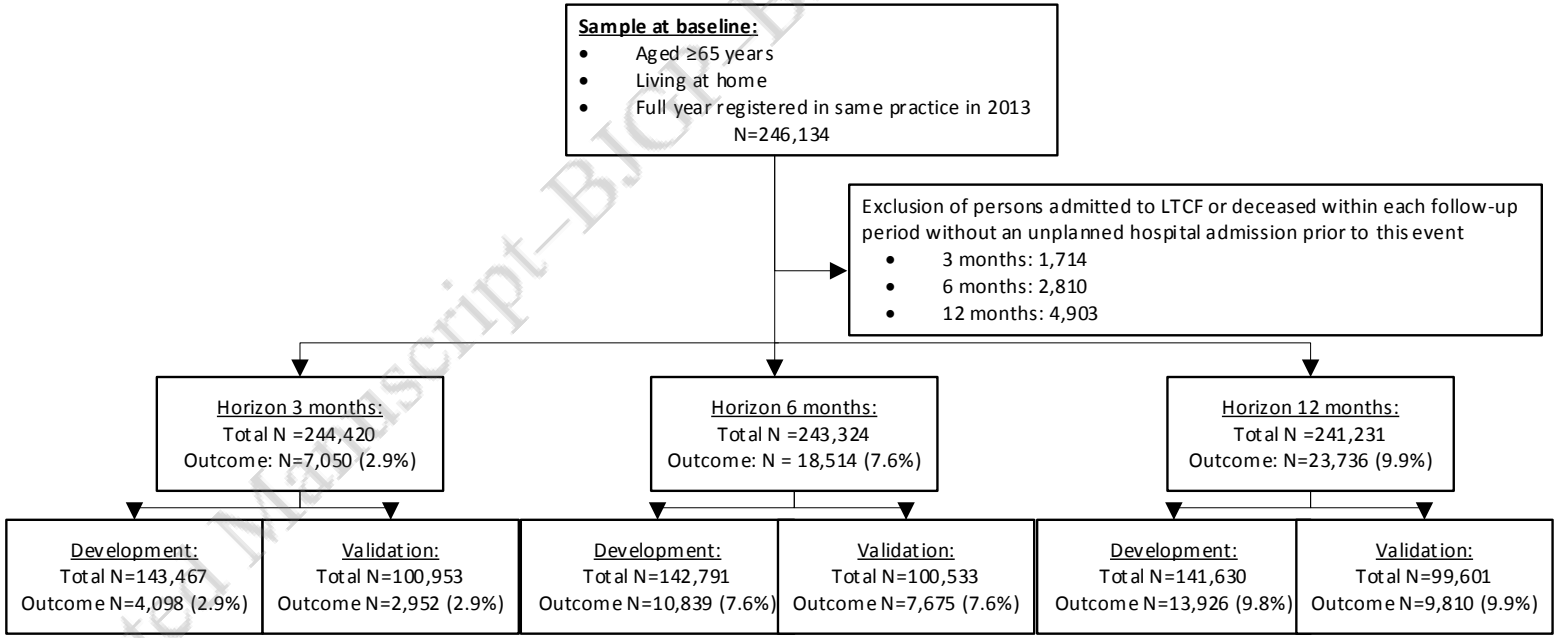

Figure 1 Flow of participants through study. *LTCF = long-term care facility*

Variable	Development (South) Total N = 142,791	Validation (North) Total N = 100,533
Age in years, median (IQR)	72 (10)	72 (10)
Female sex, N (%)	76,892 (53.8)	54,141 (53.9)
Living alone, N (%)	44,104 (30.9)[a]	32,720 (32.5)[a]
Recent widowhood, N (%)	11,073 (7.8)	7,885 (7.8)
Non-western migration background, N (%)	4,065 (2.8)[a]	3,587 (3.6)[a]
Annual household income above €25,500, N (%)*	51,913 (36.4)[a]	34,922 (34.7)[a]
Number of admissions in past year, N (%)		
- One	13,034 (9.1)[a]	9,548 (9.5)[a]
- Two or more	5,258 (3.7)[a]	3,926 (3.9)[a]
Cancer, N (%)	24,637 (17.3)[a]	18,341 (18.2)[a]
Coronary artery disease, N (%)	19,231 (13.5)	13,592 (13.5)
Heart failure, N (%)	5,796 (4.1)[a]	4,617 (4.6)[a]
Stroke (ischemic or haemorrhage) or TIA, N (%)	10,604 (7.4)	7,393 (7.4)
Chronic neck- or back disorder, N (%)	16,459 (11.5)	11,777 (11.7)
Osteoarthritis, N (%)	27,661 (19.4)[a]	20,124 (20.0)[a]
Anxiety disorder, N (%)	2,805 (2.0)	2,081 (2.1)
Depression, N (%)	4,898 (3.4)	3,524 (3.5)
Chronic obstructive pulmonary disease, N (%)	12,908 (9.0)	9,286 (9.2)
Diabetes mellitus, N (%)	23,521 (16.5)[a]	17,423 (17.3)[a]
Dementia, N (%)	1,910 (1.3)	1,365 (1.4)
Multimorbidity, N (%)	59,867 (41.9)[a]	43,534 (43.3)[a]
Polypharmacy, N (%)		
- 5-9 prescription medication	29,951 (21.0)[a]	22,058 (21.9)[a]
- ≥10 prescription medication	4,898 (3.4)[a]	3,471 (3.5)[a]
Blood thinners (VKA, DOAC or antiplatelet), N (%)	41,620 (29.1)[a]	28,643 (28.5)[a]
Number of FRIDs, median (IQR)	0 (1)	0 (1)
Taking ≥2 FRIDs and ≥1 blood thinner, N (%)	8,733 (6.1)	6,191 (6.2)
NSAIDs, N (%)	7,666 (5.4)	5,266 (5.2)
Triple whammy **, N (%)	2,473 (1.7)	1,835 (1.8)
Number of CTV contacts with GP or practice nurse past year, median (IQR)	5 (7)	5 (7)
Percentage home visits past year, N (%)		
- No home visits (ref)	116,664 (81.7)[a]	81,542 (81.1)[a]
- Up to 25% home visits	13,695 (9.6)[a]	9,826 (9.8)[a]
- >25% home visits	12,432 (8.7)[a]	9,165 (9.1)[a]
Change in contact rate in past 3 months, N (%)		
- Less than past 3 months (ref)	66,683 (46.7)[a]	46,011 (45.8)[a]
- As much as past 3 months	41,292 (28.9)[a]	29,869 (29.7)[a]
- More than past 3 months	34,816 (24.4)[a]	24,653 (24.5)[a]
Possible care avoider***, N (%)	6,427 (4.5)	4,529 (4.5)
Unplanned hospital admission within 6 months, N (%)	10,839 (7.6)	7,675 (7.6)

Table 1 Characteristics of candidate predictors in 6-month development and validation samples

** Annual household income had 83 missings in the southern sample, and 33 missings in the northern sample.*

*** Concurrent use of an antidiuretic, ACE-inhibitor and NSAID (Supplementary Table 1).*

**** Defined as ≥1 chronic condition and no registered contact with GP or practice nurse in the past year (Supplementary Table 1).*

[a] P<0.05 between groups

COPD = chronic obstructive pulmonary disease, CTV = consults, telephone consults & home visits, DOAC = direct oral anticoagulants, FRIDs = Fall Risk Increasing Drugs, GP = general practitioner, IQR = interquartile range, NSAIDs = non-steroid anti-inflammatory drugs, VKA = vitamin K antagonists.

	Optimal model				Readily-available model				Easy-to-use model			
	Β coefficient (95%-CI)		*OR*	*p-value*	*Β coefficient (95%-CI)*		*OR*	*p-value*	*Β coefficient (95%-CI)*		*OR*	*P-value*
Intercept	-6.036	(-6.26 – -5.81)			-5.627	(-5.85 – -5.40)			-6.784	(-7.00 - -6.57)		
Age (years)	0.040	(0.04 – 0.04)	1.04	<0.001	0.035	(0.03 – 0.04)	1.04	<0.001	0.053	(0.05 – 0.06)	1.05	<0.001
Female sex	-0.279	(-0.32 – -0.24)	0.76	<0.001	-0.330	(-0.37 – -0.29)	0.72	<0.001	-0.23	(-0.27 - -0.19)	0.79	<0.001
Admissions in past year				<0.001								<0.001
None (ref)	0		1						0		1	
1	0.499	(0.44 – 0.56)	1.65						0.660	(0.60 – 0.72)	1.93	
≥2	0.972	(0.90 – 1.05)	2.64						1.23	(1.16 – 1.30)	3.43	
Heart failure									0.378	(0.30 – 0.45)	1.46	<0.001
COPD	0.420	(0.36 – 0.48)	1.52	<0.001	0.430	(0.37 – 0.49)	1.54	<0.001	0.436	(0.38 – 0.49)	1.55	<0.001
Polypharmacy				<0.001				<0.001				<0.001
0-4 medications (ref)	0		1		0		1		0		1	
5-9 medications	0.380	(0.33 – 0.43)	1.46		0.424	(0.37 – 0.47)	1.53		0.574	(0.53 – 0.62)	1.78	
≥10 medications	0.613	(0.53 – 0.70)	1.85		0.715	(0.63 – 0.80)	2.0		1.012	(0.93 – 1.09)	2.75	
Blood thinners	0.214	(0.17 – 0.26)	1.24	<0.001	0.246	(0.20 – 0.29)	1.28	<0.001				
Number of CTV contacts	0.022	(0.02 – 0.03)	1.02	<0.001	0.028	(0.03 – 0.03)	1.03	<0.001				
Percentage home visits				<0.001				<0.001				
None (ref)	0		1		0		1					
≤25%	0.309	(0.25 – 0.37)	1.36		0.498	(0.44 – 0.56)	1.64					
>25%	0.470	(0.41 – 0.53)	1.60		0.641	(0.58 – 0.70)	1.90					
	Development		**Validation**		**Development**		**Validation**		**Development**		**Validation**	
AUC (95%-CI)	0.73 (0.72 – 0.73)		0.73 (0.72-0.73)		0.72 (0.71 – 0.72)		0.72 (0.71-0.72)		0.72 (0.71-0.72)		0.72 (0.71-0.72)	

Table 2 The final prediction models from the multivariable logistic regression based on the development sample together with odds ratios (OR), 95%-confidence intervals (CI), and area under the curve (AUC) in development and validation samples. *CTV = consults, telephone consults & home visits, COPD = chronic obstructive pulmonary disease.*

Probability threshold	**Development**				**Validation**			
	Sensitivity, %	*Specificity, %*	*PPV, %*	*NPV, %*	*Sensitivity, %*	*Specificity, %*	*PPV, %*	*NPV, %*
0.05	79.7	51.2	11.8	96.8	80.4	50.6	11.9	96.9
0.07	64.5	69.8	14.9	96.0	65.7	68.5	14.7	96.0
0.1	48.3	82.6	18.6	95.1	49.6	81.5	18.1	95.1
0.15	31.1	91.5	23.2	94.2	33.3	90.8	23.1	94.3
0.20	20.6	95.4	26.9	93.6	22.9	94.9	27.0	93.7
0.25	14.2	97.4	31.0	93.3	15.7	97.9	29.9	93.3

Table 3 Measures of predictive performance of the optimal model in the development and validation sample for multiple probability thresholds.
The 0.07 threshold is defined by Youden's index as optimal probability threshold. PPV = positive predictive value, NPV = negative predictive value

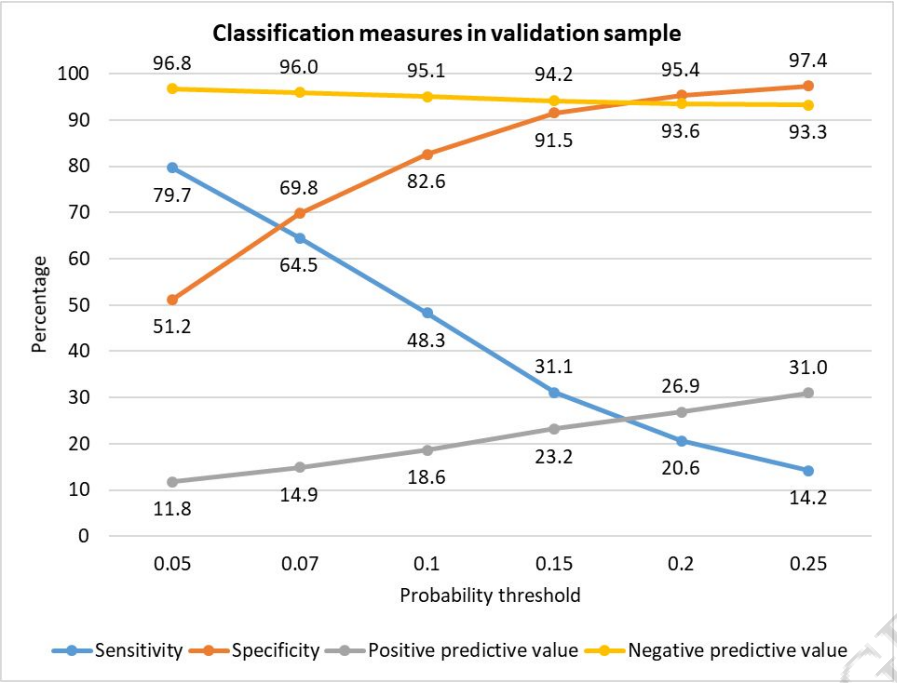

Figure 2 Graphical presentation of performance measures of the optimal model in the validation sample

Clinical implications of choice of cut-off value

Choosing a cut-off value provides the opportunity to stratify patients into low and high-risk groups. This facilitates clinical decision-making. To illustrate this, we consider a practice consisting of 500 community-dwelling patients aged 65 and over. We compare the consequences of two different cut-off values (or probability thresholds): 0.07 and 0.15. We assume a prior probability of 7.6% for each patient (prevalence). The 2x2 contingency tables for both cut-offs are shown below.

Using a cut-off of 0.07 stratifies approximately one third of the practice's older population as high risk, requiring screening or intervention. However, this choice results in a high number of false positives, where individuals are identified as high risk, but do not experience the predicted outcome. At the individual level, using a threshold of 0.07 increases the probability of a patient being classified as high risk for unplanned hospitalisation by a factor of 2, from 7.6% to 14.9%. This means that out of 100 high-risk patients, 15 will have an unplanned hospital admission within 6 months.

Alternatively, using a cut-off value of 0.15, one in ten older patients will be classified as high risk, resulting in a substantially lower number of false positives. However, the number of false negatives doubles, indicating that some potential cases are missed. For a high-risk patient at the 0.15 threshold, the probability of unplanned hospital admissions increases by a factor of 3 to 23.2%. Consequently, out of 100 high risk patients, 23 would experience an unplanned hospitalisation within 6 months.

It depends on the clinician how much risk they are willing to take to avoid missing an unplanned admission. Opting for a lower threshold results in a low number of false negatives, but raises the probability of false positives, requiring a more extensive and labor-intensive screening process.

Threshold = 0.07	Admission +	Admission -	**Total**
High risk	25 (TP)	140 (FP)	165
Low risk	13 (FN)	322 (TN)	335
Total	38	462	500

Threshold = 0.15	Admission +	Admission -	**Total**
High risk	12 (TP)	39 (FP)	51
Low risk	26 (FN)	423 (TN)	449
Total	38	462	500

2x2 Contingency tables for threshold 0.07 (left) and 0.15 (right)

Box 1 Practical guidance on choosing the appropriate cut-off value

FN: false negative, FP: false positive, TN: true negative, TP: true positive

Discussion

Summary

In this study, we used routinely recorded and linked health and census data to develop and validate an easy-to-use prediction model for unplanned hospitalisations in community-dwelling older adults. Predictors associated with unplanned hospital admission included age, sex, hospitalisation in the past year, polypharmacy, the use of blood thinners, COPD, heart failure, number of CTV contacts and percentage of home visits. The optimal model showed satisfactory discrimination and good calibration. Moreover, geographic validation, reducing the number of predictors, changing the prediction horizon, and including individuals who died or were admitted to LTCF within the prediction period, all resulted in a negligible decrease in discriminative ability, demonstrating robustness of the model. This model lost discriminatory power in a subsample of individuals with dementia or cognitive decline. Our results enable GPs to identify patients who may benefit from targeted admission prevention strategies. To improve predictions, we emphasise the importance of routine recording or incorporation of hospitalisation data in the EHR.

Comparison with existing literature

We identified 19 existing prediction models to predict hospitalisations in older adults(12). Our model showed similar performance and overlap in the most commonly included variables. However, while many existing models used a 12-month prediction horizon, we chose a 6-month horizon from a clinical perspective, because a high predicted probability of hospitalisation within 6 months is more likely to trigger timely clinical action than the same probability of hospitalisation within 12 months. However, model validation over 12 months demonstrated equivalent discriminatory ability, albeit with systematic underestimation, requiring adjustment of the intercept.

Two previous studies developed a model for people with dementia in primary care, and both demonstrated good predictive performance. These found (changes in) psychotropic medication, psychiatric diagnoses, and hypertension to be important predictors next to previous hospitalisations and polypharmacy(29, 30). Other studies found duration and severity of dementia, caregiver burden and continuity of care associated with hospitalisation(31-33). To improve predictive accuracy in individuals with cognitive decline, these predictors may need to be considered for inclusion in the model.

Strengths and limitations

A strength of this study is the use of multiple approaches for model development, providing valuable insights into the relative effectiveness and practical utility. By considering the advantages and limitations of each approach, healthcare providers and policymakers can make informed decisions about which model is suitable for their specific needs and resources. The use of EHR data enriched with national administrative data resulted in the best predictive model, i.e. the optimal model. Using structured EHR data allows the readily-available model to be implemented nationwide. However, it includes more time-consuming variables compared to the easy-to-use model. By facilitating rapid bedside assessment, the easy-to-use model is more accessible to GPs, while incorporating the most predictive variable: prior hospitalisations. Furthermore, the large longitudinal sample and its nationwide representativeness suggests these findings could be generalised across the Netherlands. This study also has limitations. As advocated in literature(34, 35), updating an existing prediction model is preferred over simply developing a new model, so information from the previous models is not neglected. However, model updating is only valuable provided the original model's development is appropriately performed and variables and outcomes are determined in a similar way(34). For this study however, the low quality of reporting in the previous studies(12), and the lack of several variables in our dataset made updating infeasible. Moreover, differences in care systems between countries complicate the transportability of existing models to other geographical populations(36), and no model had yet been developed in the Netherlands. Altogether, this large sample called for the development of a new model rather than updating an existing one. Nevertheless, to incorporate data from previous models as much as possible, we assessed the variables most frequently included in previous models as candidate predictors for inclusion in our model. Furthermore, although our data are approximately 10 years old, the relevance remains. Reviews have shown the long-term trends and relative stability over time of included predictors of unplanned hospital admissions, such as prior health care use, chronic conditions and polypharmacy(12, 13, 37).

Implication for research and practice

The implementation of ageing-in-place policies in the Netherlands in 2015, which included a reduction of the residential care capacity, has increased the number of older adults living in the community(38). Since this study used data from before the implementation of this policy, a different case-mix of community-dwelling older adults may be expected. Therefore, validation in more recent data is recommended. Additionally, we emphasise the importance of systematic recording of

hospitalisations in the EHR to enable practical implementation and provide the most accurate risk estimates, as prior hospitalisations is the strongest predictor of future hospitalisations.

Our models may support timely identification and proactive intervention of older patients at risk of unplanned hospitalisation. When selecting the appropriate cut-off value for targeting interventions, clinicians should prioritise factors as patient preference, intervention time burden, and the trade-off between intervention benefits and potential missed cases (Box 1). Finally, our model could assist policymakers estimate the required number of hospital beds in the region to anticipate on the needed capacity.

Funding:

The work was supported by the Netherlands Organization for Health Research and Development (ZonMw) (Grant No. 733050403), Stichting PvE fonds and the I-CARE4OLD project (Grant Agreement No. 96534). More information on the I-CARE4OLD project can be found at http://www.icare4old.eu and https://cordis.europa.eu/project/id/965341. All sponsors had no role in the design, methods, subject recruitment, data collection, analysis or preparation of the study.

Ethical approval

This study has been approved according to the governance code of Nivel Primary Care Database. This can be found under number NZR-00315.063. As pseudonymised data were used that were collected for routine administrative registration purposes, the informed consent of the participants was not necessary. Patients were informed by their GP about the use of their pseudonymised health data and were given the opportunity to object.

Competing interests

The authors declare that they have no competing interests.

Reference list

1. Long SJ, Brown KF, Ames D, Vincent C. What is known about adverse events in older medical hospital inpatients? A systematic review of the literature. Int J Qual Health Care. 2013;25(5):542-54.
2. Shepherd H, Livingston G, Chan J, Sommerlad A. Hospitalisation rates and predictors in people with dementia: a systematic review and meta-analysis. BMC Med. 2019;17(1):130.
3. Fogg C, Griffiths P, Meredith P, Bridges J. Hospital outcomes of older people with cognitive impairment: An integrative review. Int J Geriatr Psychiatry. 2018;33(9):1177-97.
4. Lehmann J, Michalowsky B, Kaczynski A, Thyrian JR, Schenk NS, Esser A, et al. The Impact of Hospitalization on Readmission, Institutionalization, and Mortality of People with Dementia: A Systematic Review and Meta-Analysis. Journal of Alzheimer's Disease. 2018;64:735-49.
5. Araujo de Carvalho I, Epping-Jordan J, Pot AM, Kelley E, Toro N, Thiyagarajan JA, Beard JR. Organizing integrated health-care services to meet older people's needs. Bull World Health Organ. 2017;95(11):756-63.
6. Ministry of Health WaS. Integral Care Agreement "Working together on healthy care". 2022.
7. Baker A, Leak P, Ritchie LD, Lee AJ, Fielding S. Anticipatory care planning and integration: a primary care pilot study aimed at reducing unplanned hospitalisation. Br J Gen Pract. 2012;62(595):e113-20.
8. Pritchard C, Ness A, Symonds N, Siarkowski M, Broadfoot M, McBrien KA, et al. Effectiveness of hospital avoidance interventions among elderly patients: A systematic review. Cjem. 2020;22(4):504-13.
9. Costa AP, Haughton D, Heckman G, Bronskill S, Sinha S, McKelvie R. THE DIVERT-CARE CATALYST TRIAL: TARGETED CHRONIC-DISEASE MANAGEMENT FOR HOME CARE CLIENTS. Innovation in Aging. 2017;1(suppl_1):322-3.
10. Nord M, Lyth J, Alwin J, Marcusson J. Costs and effects of comprehensive geriatric assessment in primary care for older adults with high risk for hospitalisation. BMC Geriatr. 2021;21(1):263.
11. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q. 2005;83(3):457-502.
12. Klunder JH, Panneman SL, Wallace E, de Vries R, Joling KJ, Maarsingh OR, van Hout HPJ. Prediction models for the prediction of unplanned hospital admissions in community-dwelling older adults: A systematic review. PLoS One. 2022;17(9):e0275116.
13. Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. Med Care. 2014;52(8):751-65.
14. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-73.
15. Donker GA. Netherlands Institute for Health Services Research Care Registrations - sentinel stations 2013. In: Research NIfHS, editor. 2014.
16. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018;20(5):e185.
17. Nielen M, Hek K, Weesie Y, Davids R, Korevaar J. How often do Dutch people contact GP practices? Care use in general practice in 2019. Nivel: Utrecht, The Netherlands; 2020.
18. Statistics/Netherlands. Report on Dutch hospital discharge register (LBZBASISTAB). 2023.
19. Lamberts H, Wood M. International Classification of Primary Care. Oxford University Press. 1987.
20. van Oostrom SH, Picavet HSJ, van Gelder BM, Lemmens LC, Hoeymans N, van Dijk CE, et al. Multimorbidity and comorbidity in the Dutch population – data from general practices. BMC Public Health. 2012;12(1):715.
21. van Marum RJ VM, de Vries-Moeselaar AC, et al. . Multidisciplinary guideline for polypharmacy in Older Adults. Richtlijnendatabase.nl; 2012.

22. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology. 2017;17(1):162.
23. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.
24. Marrie RA, Dawson NV, Garland A. Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. J Clin Epidemiol. 2009;62(5):511-7.e1.
25. Fox J. Applied regression analysis and generalized linear models, 2nd ed. Thousand Oaks, CA, US: Sage Publications, Inc; 2008. xxi, 665-xxi, p.
26. Steyerberg EW. Validation of Prediction Models. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing; 2019. p. 329-44.
27. Steyerberg EW. Selection of Main Effects. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing; 2019. p. 207-25.
28. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb). 2016;26(3):297-307.
29. Maust DT, Kim HM, Chiang C, Langa KM, Kales HC. Predicting Risk of Potentially Preventable Hospitalization in Older Adults with Dementia. J Am Geriatr Soc. 2019;67(10):2077-84.
30. Tsang G, Zhou SM, Xie X. Modeling Large Sparse Data for Feature Selection: Hospital Admission Predictions of the Dementia Patients Using Primary Care Electronic Health Records. IEEE J Transl Eng Health Med. 2021;9:3000113.
31. Gungabissoon U, Perera G, Galwey NW, Stewart R. Potentially avoidable causes of hospitalisation in people with dementia: contemporaneous associations by stage of dementia in a South London clinical cohort. BMJ Open. 2022;12(4):e055447.
32. Afonso-Argilés FJ, Meyer G, Stephan A, Comas M, Wübker A, Leino-Kilpi H, et al. Emergency department and hospital admissions among people with dementia living at home or in nursing homes: results of the European RightTimePlaceCare project on their frequency, associated factors and costs. BMC Geriatr. 2020;20(1):453.
33. Rudolph JL, Zanin NM, Jones RN, Marcantonio ER, Fong TG, Yang FM, et al. Hospitalization in community-dwelling persons with Alzheimer's disease: frequency and causes. J Am Geriatr Soc. 2010;58(8):1542-8.
34. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. Journal of Clinical Epidemiology. 2008;61(1):76-86.
35. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Med Res Methodol. 2022;22(1):316.
36. Klunder JH, Bordonis V, Heymans MW, van der Roest HG, Declercq A, Smit JH, et al. Predicting unplanned hospital visits in older home care recipients: a cross-country external validation study. BMC Geriatrics. 2021;21(1):551.
37. O'Caoimh R, Cornally N, Weathers E, O'Sullivan R, Fitzgerald C, Orfila F, et al. Risk prediction in the community: A systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults. Maturitas. 2015;82(1):3-21.
38. Maarse JAM, Jeurissen PP. The policy and politics of the 2015 long-term care reform in the Netherlands. Health Policy. 2016;120(3):241-5.