

Predicting adverse health outcomes in older emergency department patients: the APOP study

J. de Gelder^{1*}, J.A. Lucke², B. de Groot², A.J. Fogteloo³, S. Anten⁴, K. Mesri⁴, E.W. Steyerberg⁵, C. Heringhaus², G.J. Blauw¹, S.P. Mooijaart^{1,6}

Departments of ¹Gerontology and Geriatrics, ²Emergency Medicine, ³Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands, ⁴Department of Internal Medicine, Alrijne Hospital, Leiderdorp, the Netherlands, ⁵Department of Public Health, Erasmus MC, Rotterdam, the Netherlands, ⁶Institute of Evidence-Based Medicine in Old Age | IEMO, Leiden, the Netherlands, *corresponding author: tel.: +31 (0)71-5266640, fax: +31 (0)71-5248945, email: J.de_Gelder@lumc.nl

ABSTRACT

Background: Older patients experience high rates of adverse outcomes after an emergency department (ED) visit. Early identification of those at high risk could guide preventive interventions and tailored treatment decisions, but available models perform poorly in discriminating those at highest risk. The present study aims to develop and validate a prediction model for functional decline and mortality in older patients presenting to the ED.

Methods: A prospective follow-up study in patients aged ≥ 70 , attending the EDs of the LUMC, the Netherlands (derivation) and Alrijne Hospital, the Netherlands (validation) was conducted. A baseline assessment was performed and the main outcome, a composite of functional decline and mortality, was obtained after 90 days of follow-up.

Results: In total 751 patients were enrolled in the Leiden University Medical Center of whom 230 patients (30.6%) experienced the composite outcome and 71 patients (9.5%) died. The final model for the composite outcome resulted in an area under the curve (AUC) of 0.73 (95% CI 0.67-0.77) and was experienced in 69% of the patients at highest risk. For mortality the AUC was 0.79 (95% CI 0.73-0.85) and 36% of the patients at highest risk died. External validation in 881 patients of Alrijne Hospital showed an AUC of 0.71 (95% CI 0.67-0.75) for the composite outcome and 0.67 (95% CI 0.60-0.73) for mortality.

Conclusion: We successfully developed and validated prediction models for 90-day composite outcome and 90-day mortality in older emergency patients. The benefits

for patient management by implementing these models with preventive interventions have to be investigated.

KEYWORDS

Prognosis, emergency department, adverse outcomes, mortality, older adults

INTRODUCTION

Older patients presenting to the emergency department (ED) experience high rates of adverse outcomes,¹ but they form a heterogeneous group and it is unknown who is at highest risk. The incidence of adverse outcomes is particularly high after three months, with a mortality rate about 10% and increased functional dependence between 10-45%.¹ Early identification of those at highest risk gives an opportunity to guide preventive interventions and informed treatment decisions.²

Current models use either severity of disease or existing geriatric vulnerability for prediction. The Modified Early Warning Score (MEWS) is an indicator of disease severity and showed to be valuable in predicting worse in-hospital outcomes in older patients.³ However, prognostication of MEWS for long-term outcomes in older adults is unknown. The Identification of Seniors At Risk (ISAR)⁴ and Triage Risk Stratification Tool (TRST)⁵ focus on existing geriatric vulnerabilities, such as functional and cognitive impairment, to predict adverse health outcomes. Neither of these tools accurately identify high-risk patients,^{6,7} while

that vulnerable group of patients benefits most from an increased level of attention.

We conducted a prospective follow-up study among all older patients who visited the EDs in the region of Leiden. The aim was to develop and validate a prediction model for adverse health outcomes in older emergency patients. To reflect the condition of the patient, both demographics and severity of disease indicators and existing geriatric vulnerability were taken into account.

METHODS

Study design and setting

We performed a prospective follow-up study in the EDs of two hospitals in the region of Leiden, the Netherlands. The Leiden University Medical Center (LUMC, derivation cohort) is an academic hospital with a level 1 trauma centre and Alrijne Hospital (location Leiderdorp, validation cohort) is a peripheral hospital with a level 2 trauma centre. We considered all patients eligible who fulfilled the inclusion and exclusion criteria. The inclusion criterion was patients aged ≥ 70 presenting for the first time to the ED in the study period. The exclusion criteria were patients who were triaged with highest urgency (code red), who we were not able to approach due to an unstable medical condition, lack of permission of the nurse or physician to enter the room for any reason or due to impaired mental status without an authorised relative to provide informed consent. Also a language barrier and patients who left the waiting room were not eligible. Patients were enrolled 7 days a week for 12 weeks with 24-hour coverage in the LUMC and 12-hour coverage (10.00 am to 10.00 pm) in Alrijne Hospital. Written informed consent was obtained before inclusion. The medical ethics committee of the LUMC and Alrijne waived the necessity for formal approval of the present study, as the study closely follows routine care.

Organisation of emergency care in the Netherlands

Basic health insurance is mandatory in the Netherlands and covers the care from general practitioners (GPs), hospital care and specialist care. Emergency care is provided by GPs and in EDs. Patients in need of immediate care can contact the GP, the GP out-of-hours service, call for an ambulance or go to an ED by themselves. Depending on the urgency, patients are expected to contact or visit the GP first. As with the structure of the LUMC and Alrijne, an increasing number of GP out-of-hours centres are integrated close to the ED to avoid unnecessary ED visits.⁸ In EDs a triage nurse will prioritise patients first based on the severity of their condition; then the patient can either be directed to the emergency room or the waiting room. In the Leiden region, the LUMC and Alrijne Hospital are the

only two EDs, together servicing an unselected catchment area of 400,000 inhabitants of all ages. In both EDs there are no special rooms or trajectories for older patients. Two patient groups bypass the ED and were therefore impossible to include: 1) Older patients with a myocardial infarction in the ambulance who were directly sent to the catheterisation room and 2) older patients with a CVA and eligible for thrombolytic therapy underwent a brief primary assessment in the ED and were then sent to the neurology ward after a CT scan.

Data collection

We included patients in the LUMC from September to November 2014 and in Alrijne Hospital from March to June 2015. In both hospitals teams of medical students were present at the ED from 10.00 am until 10.00 pm to enrol patients, and in the LUMC the ED staff were responsible for inclusion from 10.00 pm until 10.00 am. Before the start of the inclusion period, the medical students and ED staff of the LUMC attended training sessions to guarantee convergence on conducting the questionnaires. The ideal moment for conducting the questionnaires turned out to be 30-45 minutes after arrival of the patient to the ED. At that moment the patient had spoken to the physician and was waiting for lab results or further analysis. The questionnaire took 5-10 minutes to complete. A representative was permitted to answer questions when the patient was unable to provide answers, with the exception of the cognition and self-reported quality of life questions. Questions were collected on a tablet computer and sent directly to a secured database. Additional medical data were extracted automatically from the medical records, verified manually and added to the database.

Baseline

At baseline, data on three domains were assessed: demographics, severity of disease indicators and geriatric measurements. Demographics consisted of age, gender, living arrangements and level of education. A low level of education was defined as elementary school, basic education as community college, middle education as secondary education and high education as higher vocational training or university. Severity of disease indicators consisted of characteristics related to the ED visit: way of arrival, triage category by the Manchester Triage System,⁹ fall-related ED visit, indication to measure vital signs and indication to perform a blood test. Whether the visit was fall related was obtained by asking the question: Is the reason for presentation related to a fall? Indication to measure vital signs or laboratory tests was scored positively when, at the moment of presentation, vital signs needed to be measured or a laboratory test was ordered based on the Manchester Triage System and

local protocols. Geriatric measurements consisted of the number of different medications mentioned by the patient, history of diagnosed dementia reported by patient or proxy, current use of a walking device, the Identification of Seniors at Risk (ISAR)⁴ screening tool, the Six Item Cognitive Impairment test (6CIT)¹⁰ and the Katz Index of Activities of Daily Living (ADL)¹¹ questionnaire. The ISAR was developed for patients aged ≥ 65 and aims to predict the risk of adverse health outcomes six months after the ED visit. The ISAR consists of six dichotomous questions and scores range from 0 to 6 with higher scores denoting higher risk. The 6CIT is a short cognition test and was validated in a Dutch population against the Mini Mental State Examination (MMSE)¹² with a score on the 6CIT of ≥ 11 indicating cognitive impairment (MMSE < 24).¹³ Six questions lead to a score ranging from 0 to 28 with higher scores indicating more cognitive impairment. The Katz ADL indicates functional status two weeks before presentation to the ED to eliminate possible effects of the acute illness and consists of six dichotomous questions on dependence in bathing, dressing, toileting, transfer, eating and incontinence. Scores range from 0 to 6 with higher scores an indication of more dependency.

Outcomes

The main outcome of the study was composite outcome, a composite of functional decline or mortality at 90-day follow-up. Functional decline was defined as at least one point increase in the Katz ADL score or new institutionalisation, defined as a higher level of assisted living at 90 days after ED visit. We analysed 90-day mortality separately. Mortality can be seen as the ultimate decline and might then be taken together with functional decline. On the other hand, the intervention strategy could differ for patients at high risk for mortality. For that reason we developed a separate prediction model for 90-day mortality. A model solely for functional decline is not feasible. Excluding deceased patients would imply that the model is only applicable in patients who will not die within a certain period, which we do not know at the moment of presentation. Three months after the ED visit the patient was contacted by telephone. In case of no response after three attempts on three consecutive days, the GP was contacted to verify the phone number and living status. Finally a letter with the follow-up questions was sent to patients who had not moved to a higher level of assisted living and who were alive according to the information from the GP. Data concerning mortality were derived from the municipal records.

Statistical analysis

Baseline characteristics are presented as mean with standard deviation (SD) in case of normal distribution, median with interquartile range (IQR) in case of skewed

distribution or as numbers with percentages (%). Adequate statistical power for obtaining good predictions requires a minimum of 10 events per candidate predictor.¹⁴ This rule was followed for the composite outcome. To reduce the number of candidate predictors, the most relevant questions from the questionnaires (Katz ADL, ISAR, 6CIT) were pre-selected. Single questions with the highest R-square values on the entire questionnaire score were selected and added to the list of candidate predictors for development of the prediction model. Missing predictors were imputed via single imputation techniques.¹⁵ The prediction model was derived via backward elimination with Akaike's Information Criterion (equivalent to $p < 0.157$ for predictors with 1 df). With this technique the least contributing candidate predictors are deleted until the deterioration in model fit is too large. Discrimination of the models was assessed with the area under the receiver operating characteristic curve (AUC). Internal validation was conducted with a 500 bootstrap sample procedure, where we repeated the backward elimination procedure in each bootstrap sample to estimate the optimism-corrected performance that is expected if the derived prediction model is applied in other datasets. The internal validation procedure also provided a shrinkage factor to adjust the estimated regression coefficients for overfitting.¹⁶ The adjusted regression equation provides predictions for new individuals. It was validated in the Alrijne patients.¹⁷ Calibration of the model, which reflects how well predicted and observed outcomes agree, was examined by using the adjusted regression equation. Calibration was examined graphically with calibration plots, with a goodness of fit test (Hosmer and Lemeshow test¹⁸). The formula $1/(1+e^{-(\text{linear predictor})})$ was applied with the adjusted regression equation to determine the individual risks of experiencing the outcome. Performance of the model for the patients with the highest 30%, 20% and 10% predicted risk was evaluated according to sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio and negative likelihood ratio, with 95% confidence intervals. To compare our model performance with the existing six-item ISAR questionnaire, predictive performance of the ISAR on different cut-off points was also calculated. The level of statistical significance was set at $p < 0.05$. Statistical analyses were performed using IBM SPSS Statistics package (version 20) and R version 3.1.1.

RESULTS

In the three-month inclusion period a total of 995 older patients presented to the ED of the LUMC. Of these, 19 patients were excluded due to a language barrier or leaving the waiting room. Another 92 patients could not

be approached due to their medical condition, resulting in 884 eligible patients. Of these, 65 patients were missed for inclusion and 68 patients refused informed consent, which led to a study population of 751 patients (85% of eligible patients). Similarly, 881 patients were included in Alrijne Hospital (*figure 1*).

The median age of LUMC patients was 78 years (IQR 74-83) and 80 years (IQR 75-84) in Alrijne patients (*table 1*); 405 (53.9%) patients of the LUMC arrived by ambulance and 201 (26.1%) were triaged as ‘very urgent’. In Alrijne 432 (49.0%) arrived by ambulance and 58 (6.6%) were triaged as ‘very urgent’. In both hospitals the majority of the older patients were independent, with a median Katz ADL score of 0 (IQR 0-1).

In total 230 LUMC patients (30.6%) experienced the composite outcome within 90 days of follow-up and 71 patients (9.5%) died. In Alrijne Hospital 247 patients (28.0%) had the composite outcome and 84 (9.5%) died.

Details on the univariate and multivariable analyses on both outcomes can be found in *supplemental tables 1 and 2*.

The final model for the composite outcome combined age, arrival by ambulance, number of different medications, help needed with bathing or showering, hospital admission in the past six months, help needed at home on a regular base and history of dementia (*table 2*). Ninety-day mortality could be best predicted by combining information of age, gender, living arrangements, a fall prior to ED visit, indication for

blood tests and needing help in dressing. Accuracy of the final models was fair to good, with in the derivation cohort an area under the curve (AUC) of 0.73 (95% CI 0.69-0.77) for the composite outcome and of 0.79 (95% CI 0.73-0.85) for mortality. External validation in Alrijne patients showed an AUC of 0.71 (95% CI 0.67-0.75) for the composite outcome and 0.67 (95% CI 0.60-0.73) for mortality. The formula of the original final models to calculate the individual risk can be found in the legend of *table 2*.

Calibration of the predicted probabilities was satisfactory (*figure 2*), with all Hosmer and Lemeshow goodness-of-fit p-values above 0.05. A stricter limit to assign patients at high risk increased specificity, PPV and the positive likelihood ratio (*table 3*). The PPV ranged from 0.55 (95% CI 0.48-0.61) to 0.69 (95% CI 0.57-0.79) for the composite outcome and from 0.21 (95% CI 0.16-0.27) to 0.36 (95% CI 0.26-0.49) for mortality, depending on the threshold chosen. This implies that in the highest risk group 69% of the patients experienced the composite outcome and 36% died.

The conventional cut-off of 2 points or higher for the ISAR resulted in a PPV of 0.40 (95% CI 0.36-0.45) for composite outcome and 0.13 (95% CI 0.10-0.16) for mortality (*supplemental table 3*). Raising the ISAR cut-off to a score of ≥ 4 points to define patients as high risk yielded a PPV of 0.50 (95% CI 0.40-0.60) for the composite outcome and 0.18 (95% CI 0.12-0.26) for mortality.

Figure 1. Flowchart of LUMC and Alrijne patients

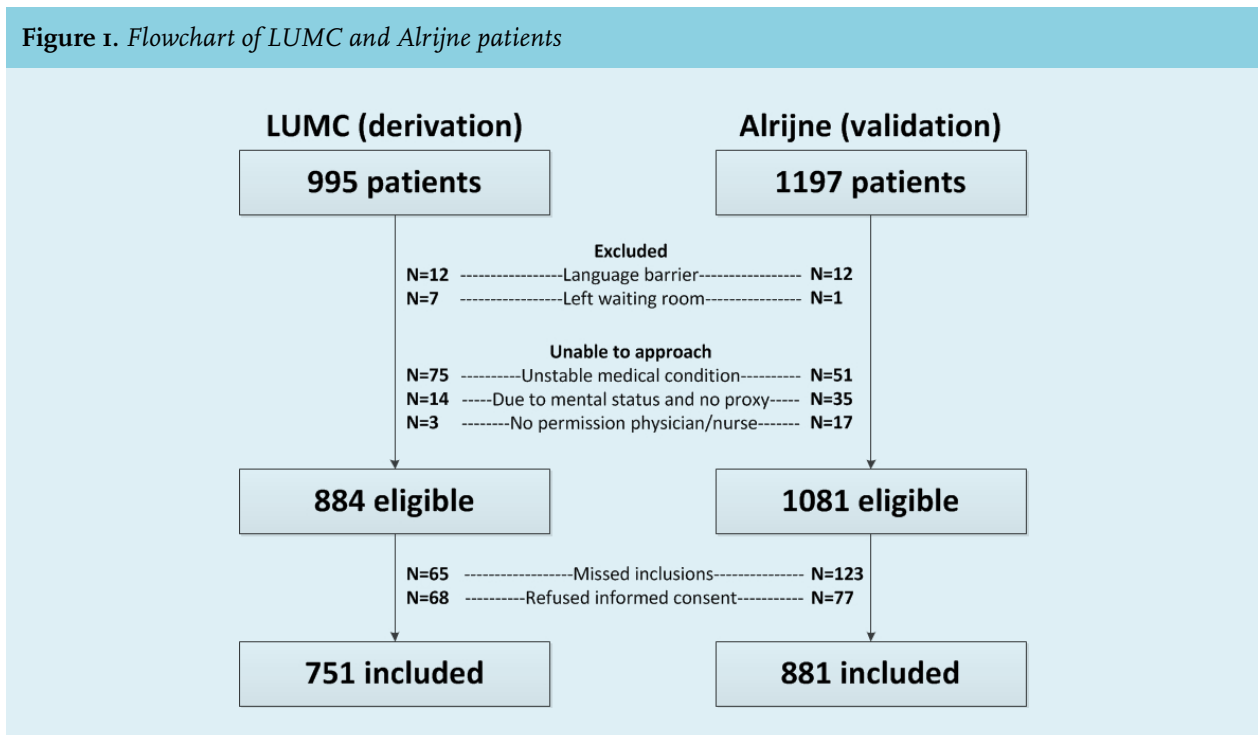


Table 1. Baseline characteristics of older emergency patients at LUMC and Alrijne

Characteristics	LUMC (n = 751)*	Alrijne (n = 881)#
Demographics		
Age (years), median (IQR)	78 (74-83)	80 (75-84)
Female	389 (51.8%)	454 (51.5%)
Living in residential care or nursing home, n (%)	63 (8.4%)	69 (7.8%)
High education, n (%)	155 (20.8%)	164 (18.6%)
Severity of disease indicators		
Arrival by ambulance, n(%)	405 (53.9%)	432 (49.0%)
Triage category, n(%)		
- Standard (Green)	159 (21.2%)	353 (40.1%)
- Urgent (Yellow)	391 (52.1%)	470 (53.3%)
- Very urgent (Orange)	201 (26.8%)	58 (6.6%)
Fall-related visit, n(%)	211 (28.1%)	192 (21.8%)
Indication to perform vital sign measurement(s), n(%)	661 (88.0%)	776 (88.1%)
Indication to perform blood test, n(%)	603 (80.3%)	749 (85.0%)
Geriatric measurements		
Number of different medications, median (IQR)	5 (3-8)	5 (3-8)
Use of walking device, n(%)	302 (40.4%)	378 (42.9%)
Katz ADL score, median (IQR) ¹	0 (0-1)	0 (0-1)
ISAR score, median (IQR) ²	2 (1-3)	2 (1-3)
History of dementia, n(%)	34 (4.5%)	42 (4.8%)
6CIT score, median (IQR) ³	4 (2-8.5)	4 (0-8)
<p>N = number; IQR= Interquartile range. *LUMC data incomplete for education (n = 746), use of walking device (n = 747), Katz ADL score (n = 745), ISAR score (n = 748) and 6CIT score (n = 697). #Alrijne data incomplete for education (n = 878), use of walking device (= 878), Katz ADL score (n = 859), ISAR score (n = 872), 6CIT (n = 791). ¹Higher scores indicating higher dependency (0-6). ²Higher scores indicating higher risk of functional decline (0-6). ³Higher scores indicating worse cognition (0-28).</p>		

DISCUSSION

New externally validated prediction models were presented for older emergency patients by using a combination of demographics, severity of disease indicators and geriatric vulnerability. Performance of the models was satisfactory, with good accuracy and high PPVs.

The predictors used in our models have previously been shown to be predictive of negative health outcomes in other models. The Identification Seniors at Risk (ISAR)⁴ tool and Triage Risk Screening Tool (TRST)⁵ were developed for older patients at the ED. The ISAR is suitable for all older patients, whereas the TRST was developed for those discharged home. Both tools include predominately geriatric vulnerabilities, such as functional and cognitive impairment, and are validated for prediction of negative health outcomes, including functional decline and mortality.^{4,19,20} Scoring systems for disease severity are

also used to predict negative health outcomes of which the Acute Physiology and Chronic Health Evaluation (APACHE II)²¹ and Early Warning Score (MEWS)³ are well known. APACHE II is available online and predicts mortality in intensive care unit patients by using an algorithm consisting of 12 physiological and two disease-related variables. The MEWS weighs the severity of five physiological parameters to identify patients at risk of clinical deterioration and can be used as a bedside evaluation instrument to predict mortality and admission in ED patients.²² The MEWS and APACHE II scores were developed for prediction of worse in-hospital outcomes, whereas the prognostic capabilities in the longer term are unknown, especially in the older population. Recently we showed that directly available clinical data describing disease severity and geriatric vulnerability can be used for prediction in hospitalised older patients.²³ The present study also selected predictors reflecting the acute

Table 2. Final model for 90-day composite outcome and 90-day mortality in older patients

	Composite adverse outcome		Mortality	
	beta	OR (95% CI)	beta	OR (95% CI)
Demographics				
Age (per 5 years increase)	0.293	1.34 (1.16-1.54)	0.462	1.59 (1.28-1.97)
Female			-1.397	0.25 (0.14-0.45)
Living in residential care or nursing home			0.730	2.08 (0.94-4.58)
High education				
Severity of disease indicators				
Arrival by ambulance	0.477	1.61 (1.14-2.28)		
Triage category				
- Standard (Green)				
- Urgent (Yellow)				
- Very urgent (Orange)				
Fall-related ED visit			-0.627	0.53 (0.26-1.10)
Indication for vital measurement(s)				
Indication for blood test(s)			1.254	3.50 (1.24-9.93)
Geriatric measurements				
Number of different medications	0.044	1.05 (1.00-1.09)		
Use of walking device				
Needs help bathing/showering	0.665	1.94 (1.24-3.05)		
Needs help dressing			1.281	3.60 (1.97-6.58)
Hospital admission in past 6 months	0.404	1.50 (1.03-2.17)		
Needed help prior to ED visit	0.716	2.05 (1.38-3.03)		
History of dementia	-0.794	0.45 (0.20-1.01)		
Disorientated in time				
Intercept	-6.557		-10.538	
AUC (95%CI) of final model	0.73 (0.69-0.77)		0.79 (0.73-0.85)	
AUC (95%CI) in Alrijne patients	0.71 (0.67-0.75)		0.67 (0.60-0.73)	
ED = emergency department; AUC = area under the receiver operator curve; CI= confidence interval. Final model equations: 1: 90-day composite outcome: $1/(1+\exp(-(-6.557 + 0.293 \times 'age/5' + 0.477 \times 'arrival\ by\ ambulance' + 0.044 \times 'number\ of\ medications' + 0.665 \times 'need\ help\ bathing\ or\ showering' + 0.404 \times 'hospital\ admission\ in\ the\ past\ six\ months' + 0.716 \times 'need\ help\ prior\ to\ ED\ visit' + -0.794 \times 'history\ of\ dementia'))))$ 2: 90-day mortality: $1/(1+\exp(-(-10.538 + 0.462 \times 'age/5' + -1.397 \times 'female\ gender' + 0.730 \times 'living\ in\ residential\ care/nursing\ home' + -0.627 \times 'fall\ related\ ED\ visit' + 1.254 \times 'indication\ of\ blood\ test' + 1.281 \times 'need\ help\ dressing'))))$				

condition of the older patient visiting the ED and developed prediction models with high specificity and high PPVs. As shown in *table 2*, demographics and severity of disease indicators are important for predicting mortality and geriatric measurements for predicting the composite outcome. It is arguable that those patients with functional impairment at baseline have a higher risk of further decline, and this stresses the importance of obtaining

these measurements of functional capacity in combination with the other parameters for accurate prediction. We showed that history of dementia decreased the risk of the composite outcome. It was an unexpected finding and it may be caused by a larger proportion of patients with dementia living in an institution, and thus better protected from poor outcome, or the group of patients with dementia were less severely ill but referred to the

Figure 2. A: Calibration plot at internal validation of 90-day composite outcomes with a Hosmer and Lemeshow goodness-of-fit p-value of 0.77. The vertical lines represent the relative frequency distribution of predicted probabilities

B: Calibration plot at internal validation of 90-day mortality with a Hosmer and Lemeshow goodness-of-fit p-value of 0.19. The vertical lines represent the relative frequency distribution of predicted probabilities

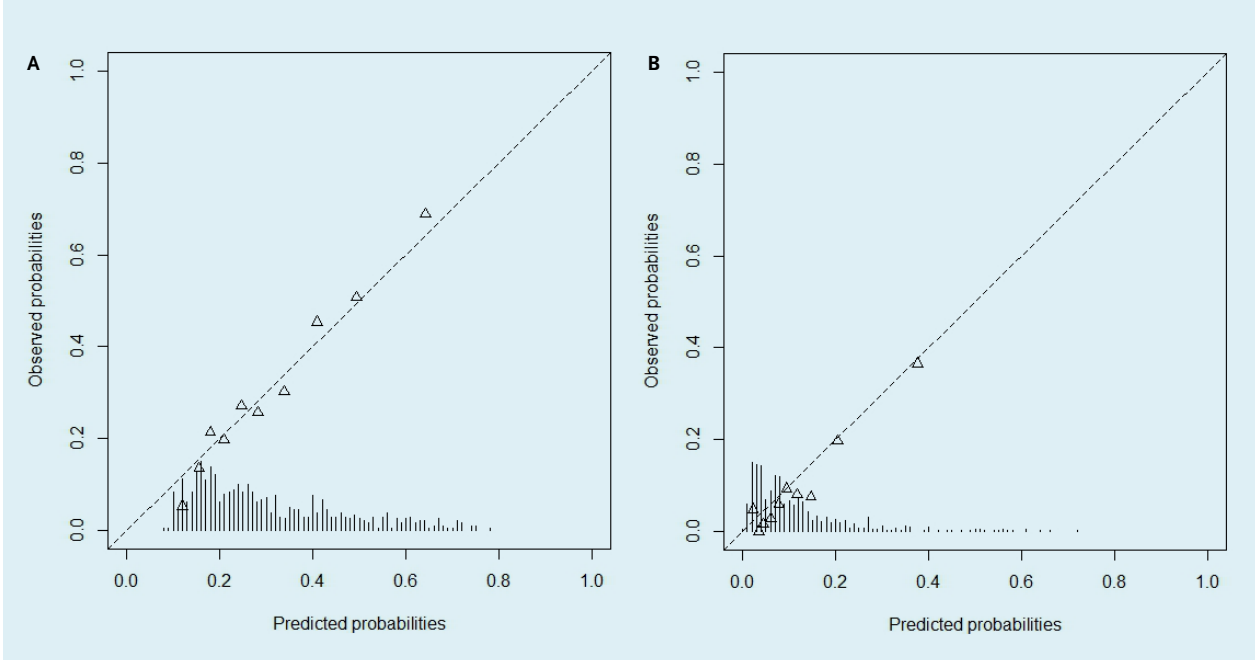


Table 3. Predictive performance of final prediction model for 90-day composite outcome and 90-day mortality

	Number of patients	Sens (95% CI)	Spec (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR- (95% CI)
30% at highest risk							
Composite outcome	225	0.53 (0.47-0.60)	0.80 (0.77-0.84)	0.55 (0.48-0.61)	0.80 (0.76-0.83)	2.73 (2.21-3.38)	0.58 (0.50-0.67)
Mortality	225	0.66 (0.54-0.77)	0.74 (0.70-0.77)	0.21 (0.16-0.27)	0.95 (0.93-0.97)	2.53 (2.05-3.12)	0.46 (0.33-0.63)
20% at highest risk							
Composite outcome	150	0.39 (0.32-0.45)	0.88 (0.85-0.91)	0.59 (0.51-0.67)	0.77 (0.73-0.80)	3.30 (2.48-4.40)	0.69 (0.62-0.77)
Mortality	152	0.59 (0.47-0.70)	0.84 (0.81-0.86)	0.28 (0.21-0.36)	0.95 (0.93-0.97)	3.66 (2.82-4.73)	0.49 (0.37-0.65)
10% at highest risk							
Composite outcome	75	0.23 (0.17-0.29)	0.96 (0.93-0.97)	0.69 (0.57-0.79)	0.74 (0.70-0.77)	5.12 (3.21-8.16)	0.81 (0.75-0.87)
Mortality	74	0.38 (0.27-0.50)	0.93 (0.91-0.95)	0.36 (0.26-0.49)	0.94 (0.91-0.95)	5.50 (3.67-8.25)	0.67 (0.55-0.80)

Sens = sensitivity, Spec = specificity, PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio, CI = confidence intervals. Predictive performance was calculated by applying the results of internal validation procedure.

ED sooner. Alternatively, it could be a chance finding in a small group of patients. Another notable finding is that a comparable percentage of patients in both hospitals arrived by ambulance, while the patients in LUMC are triaged more urgently (*table 1*). We do not have an explanation for this finding, since there are many reasons to arrive by ambulance and both hospitals use the Manchester Triage System. However, we showed that patients who arrive by ambulance are at increased risk to experience the composite outcome. Expected or unexpected, the final prediction models have to be tested in a different population or setting to support general applicability. External validation of both models in the Alrijne patients resulted in a comparable discrimination for the composite outcome and a decrease in AUC of 0.12 for mortality. It is difficult to explain the reason for this decrease in mortality. The fact that the inclusion timeframe was different between hospitals (24 hours in the LUMC vs. 10 am to 10 pm in Alrijne Hospital) is unlikely to have influenced the results substantially, as there were only a very limited number of patients included during the night, and endpoints did not differ between those included during the 'daytime' vs. those included at 'night'. More likely, it could be minor differences in the study population, in ED protocols or parameters which we did not or cannot measure.

Predictive performance of a comparable model, the ISAR,⁴ was analysed in the same study population (*supplemental table 3*). The performance was characterised with high sensitivities and low specificities, resulting in relatively low PPVs and high NPVs. As a consequence ISAR is more useful to 'rule out' patients at high risk, where our models target patients at highest risk. Prediction of individual risk scores on multiple outcomes, as shown with the composite outcome and mortality enable emergency physicians to guide preventive interventions and tailored treatment decisions. As an example, for patients with a predicted risk for the composite adverse outcome of 50% to 65%, safety procedures could be applied, whereas a predicted risk of 65% or higher can lead to more intensive interventions. On one hand standardised interventions should be administered, such as nursing these patients in a comfortable bed and informing the general practitioner. On the other hand, the predicted risk could support the physician in deciding to start physiotherapy or in making an outpatient appointment to prevent deterioration. If the risk of 90-day mortality is also high, this could be an argument to spend more time on diagnostic and therapeutic shared decision making and advanced care planning.

To date, there is no standard screening program for older adults in the ED. This could be due to the low proportion of evidence-based studies designed for the elderly,^{24,25} specifically due to the low number of clinical

impact studies.²⁶ The ultimate goal is to introduce a new generalised prediction tool, suitable for all older emergency patients and to design and test effectiveness of different interventions. Such a model should consist of patient-related parameters rather than organisation-dependent factors, such as the indication to perform measurements. The present model is accurate for older patients in Western Europe. We are planning such external validation studies, which will show whether the model needs to be updated to specific settings. The algorithm can simply be integrated in the electronic patient record to incorporate screening into routine care or be used as an application as developed on the website: <http://screener.apop.eu>.

One of the limitations in the current study is the lack of baseline data on potentially important determinants such as malnutrition, depression and instrumental ADL functioning. Since time is scarce in the acute setting we had to limit the number of questions, instead of performing a comprehensive geriatric assessment. A second limitation is the low proportion of deceased patients within 90 days of follow-up. As a consequence, power for prediction of 90-day mortality was low. The major strength is the unselected representative study population. We included 85% of the eligible older patients 24/7 during 12 weeks. A second strength is the fact that demographics, severity of disease and geriatric vulnerability of the patient were taken into account as a reflection of the condition of the patient.

In conclusion, we successfully developed and validated prediction models for 90-day composite outcome and 90-day mortality in older emergency patients. The benefits for patients by implementing these models with preventive interventions have to be further investigated.

DISCLOSURES

Conflict of Interest: The authors declare no conflict of interest.

Funding: The Institute for Evidence-Based Medicine in Old Age (IEMO) is funded by the Dutch Ministry of Health and Welfare and supported by ZonMW (project number 62700.3002).

REFERENCES

1. Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann Emerg Med.* 2002;39:238-47.
2. Samaras N, Chevalley T, Samaras D, Gold G. Older patients in the emergency department: a review. *Ann Emerg Med.* 2010;56:261-9.
3. Cei M, Bartolomei C, Mumoli N. In-hospital mortality and morbidity of elderly medical patients can be predicted at admission by the Modified Early Warning Score: a prospective study. *Int J Clin Pract.* 2009;63:591-5.

4. McCusker J, Bellavance F, Cardin S, Trepanier S, Verdon J, Ardman O. Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. *J Am Geriatr Soc*. 1999;47:1229-37.
5. Meldon SW, Mion LC, Palmer RM, et al. A brief risk-stratification tool to predict repeat emergency department visits and hospitalizations in older patients discharged from the emergency department. *Acad Emerg Med*. 2003;10:224-32.
6. Carpenter CR, Shelton E, Fowler S, et al. Risk factors and screening instruments to predict adverse outcomes for undifferentiated older emergency department patients: a systematic review and meta-analysis. *Acad Emerg Med*. 2015;22:1-21.
7. Yao JL, Fang J, Lou QQ, Anderson RM. A systematic review of the identification of seniors at risk (ISAR) tool for the prediction of adverse outcome in elderly patients seen in the emergency department. *Int J Clin Exp Med*. 2015;8:4778-86.
8. Schafer W, Kroneman M, Boerma W, et al. The Netherlands: health system review. *Health Syst Transit*. 2010;12:v-xxvii, 1-228.
9. Mackway-Jones K. Manchester Triage Group. *Emergency Triage*. 1997.
10. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. *Am J Psychiatry*. 1983;140:734-9.
11. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged: The index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914-9.
12. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res*. 1975;12:189-98.
13. Tuijl JP, Scholte EM, de Craen AJ, van der Mast RC. Screening for cognitive impairment in older general hospital patients: comparison of the Six-Item Cognitive Impairment Test with the Mini-Mental State Examination. *Int J Geriatr Psychiatry*. 2012;27:755-62.
14. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer; 2009.
15. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59:1087-91.
16. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med*. 1990;9:1303-25.
17. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98:683-90.
18. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York: Wiley; 2000.
19. Buurman BM, van den Berg W, Korevaar JC, Milisen K, de Haan RJ, de Rooij SE. Risk for poor outcomes in older patients discharged from an emergency department: feasibility of four screening instruments. *Eur J Emerg Med*. 2011;18:215-20.
20. Hustey FM, Mion LC, Connor JT, Emerman CL, Campbell J, Palmer RM. A brief risk stratification tool to predict functional decline in older adults discharged from emergency departments. *J Am Geriatr Soc*. 2007;55:1269-74.
21. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818-29.
22. Bulut M, Cebicci H, Sigirli D, et al. The comparison of modified early warning score with rapid emergency medicine score: a prospective multicentre observational cohort study on medical and surgical patients presenting to emergency department. *Emerg Med J*. 2014;31:476-81.
23. De Gelder J, Lucke JA, Heim N, et al. Predicting mortality in acutely hospitalized older patients: a retrospective cohort study. *Intern Emerg Med*. 2016;11:587-94.
24. Broekhuizen K, Pothof A, de Craen AJ, Mooijaart SP. Characteristics of randomized controlled trials designed for elderly: a systematic review. *PLoS One*. 2015;10:e0126709.
25. Mooijaart SP, Broekhuizen K, Trompet S, et al. Evidence-based medicine in older patients: how can we do better? *Neth J Med*. 2015;73:211-8.
26. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10:e1001381.

Supplemental table 1. Univariate and multivariable associations of candidate predictors for 90-day composite outcome in older emergency patients of the Leiden University Medical Center

	Univariate		Multivariate	
	OR (95% CI)	p-value	OR (95% CI)	p-value
Demographics				
Age (per 5 years increase)	1.51 (1.33-1.71)	< 0.001	1.29 (1.10-1.50)	0.002
Female	1.16 (0.85-1.58)	0.353	1.00 (0.70-1.44)	0.985
High education	0.81 (0.54-1.20)	0.285	1.06 (0.68-1.66)	0.784
Living in residential care or nursing home	3.93 (2.31-6.68)	< 0.001	1.66 (0.89-3.11)	0.114
Severity of disease indicators				
Arrival by ambulance	2.01 (1.46-2.77)	< 0.001	1.75 (1.17-2.61)	0.006
<i>Triage category</i>				
Standard (Green)				
Urgent (Yellow)	1.09 (0.73-1.62)	0.684	0.93 (0.58-1.50)	0.772
Very urgent (Orange)	0.92 (0.58-1.45)	0.704	0.60 (0.33-1.11)	0.081
Fall related ED visit	0.95 (0.67-1.34)	0.775	0.77 (0.51-1.19)	0.239
Indication of vital measurement(s)	1.63 (0.97-2.75)	0.067	1.22 (0.56-2.65)	0.616
Indication of blood test(s)	1.48 (0.98-2.23)	0.065	1.06 (0.57-1.96)	0.865
Geriatric measurements				
Number of different medications	1.08 (1.04-1.12)	< 0.001	1.04 (0.99-1.09)	0.087
Use of walking device	3.14 (2.28-4.34)	< 0.001	1.33 (0.86-2.05)	0.194
Needs help bathing/showering	3.94 (2.74-5.66)	< 0.001	2.19 (1.17-4.12)	0.015
Needs help dressing	3.07 (2.09-4.50)	< 0.001	0.66 (0.34-1.27)	0.213
Hospital admission in the past 6 months	1.68 (1.20-2.34)	0.002	1.40 (0.96-2.04)	0.085
Needed help prior to ED visit	3.34 (2.63-5.04)	< 0.001	1.97 (1.29-3.00)	0.002
History of dementia	1.25 (0.61-2.57)	0.546	0.45 (0.19-1.08)	0.073
Disorientated in time	1.78 (1.16-2.73)	0.008	0.93 (0.55-1.57)	0.777

Supplemental table 2. Univariate and multivariable associations of candidate predictors for 90-day mortality in older emergency patients of the Leiden University Medical Center

	Univariate		Multivariate	
	OR (95% CI)	p-value	OR (95% CI)	p-value
Demographics				
Age (per 5 years increase)	1.62 (1.34-1.95)	< 0.001	1.59 (1.25-2.02)	< 0.001
Female	0.33 (0.19-0.57)	< 0.001	0.21 (0.11-0.40)	< 0.001
High education	0.76 (0.40-1.46)	0.415	0.65 (0.31-1.36)	0.251
Living in residential care or nursing home	4.79 (2.59-8.86)	< 0.001	1.96 (0.86-4.44)	0.109
Severity of disease indicators				
Arrival by ambulance	2.03 (1.20-3.43)	0.008	1.53 (0.79-2.97)	0.210
Triage category				
Standard (Green)				
Urgent (Yellow)	1.06 (0.57-1.99)	0.847	0.65 (0.31-1.35)	0.248
Very urgent (Orange)	0.89 (0.43-1.84)	0.747	0.46 (0.19-1.13)	0.091
Fall related ED visit	0.49 (0.26-0.93)	0.030	0.51 (0.24-1.05)	0.076
Indication of vital measurement(s)	5.13 (1.24-21.29)	0.024	2.47 (0.42-14.70)	0.320
Indication of blood test(s)	3.12 (1.39-8.89)	0.008	2.20 (0.64-7.63)	0.213
Geriatric measurements				
Number of different medications	1.05 (0.99-1.11)	0.123	0.97 (0.91-1.05)	0.470
Use of walking device	2.25 (1.37-3.72)	0.001	0.71 (0.34-1.48)	0.359
Needs help bathing/showering	4.88 (2.95-8.08)	0.001	2.05 (0.80-5.25)	0.133
Needs help dressing	4.92 (2.95-8.21)	< 0.001	2.20 (0.87-5.56)	0.097
Hospital admission in the past 6 months	1.81 (1.09-2.99)	0.021	1.46 (0.81-2.60)	0.206
Needed help prior to ED visit	2.87 (1.74-4.73)	0.001	1.46 (0.71-3.00)	0.310
History of dementia	1.29 (0.44-3.78)	0.638	0.50 (0.13-1.85)	0.296
Disorientated in time	2.22 (1.20-4.10)	0.011	0.79 (0.36-1.71)	0.542

Supplemental table 3. External validation of the Identification of Seniors At Risk (ISAR) with different cut-off points in LUMC study population

	Sens (95%CI)	Spec (95%CI)	PPV (95%CI)	NPV (95%CI)	LR+ (95%CI)	LR- (95%CI)
ISAR ≥ 2 (61%)						
Composite outcome	0.80 (0.74-0.85)	0.48 (0.43-0.52)	0.40 (0.36-0.45)	0.84 (0.80-0.88)	1.53 (1.38-1.70)	0.42 (0.32-0.54)
Mortality	0.83 (0.72-0.91)	0.42 (0.38-0.45)	0.13 (0.10-0.16)	0.96 (0.93-0.98)	1.42 (1.26-1.61)	0.41 (0.24-0.68)
ISAR ≥ 3 (35%)						
Composite outcome	0.55 (0.48-0.61)	0.74 (0.70-0.77)	0.48 (0.42-0.55)	0.79 (0.75-0.82)	2.13 (1.77-2.57)	0.61 (0.53-0.70)
Mortality	0.61 (0.48-0.72)	0.68 (0.64-0.72)	0.17 (0.12-0.22)	0.94 (0.92-0.96)	1.90 (1.53-2.36)	0.58 (0.43-0.77)
ISAR ≥ 4 (15%)						
Composite outcome	0.24 (0.19-0.31)	0.89 (0.86-0.92)	0.50 (0.40-0.60)	0.73 (0.69-0.76)	2.27 (1.62-3.17)	0.85 (0.79-0.91)
Mortality	0.28 (0.18-0.40)	0.86 (0.84-0.89)	0.18 (0.12-0.26)	0.92 (0.90-0.94)	2.08 (1.37-3.16)	0.83 (0.72-0.96)

Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; CI = confidence interval.